

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XXIV · NUMBER 3

April 2023

ARTICLES

- 315 Defensive Killing by Police: Analyzing
Uncertain Threat Scenarios
Jennifer M. Page
- 352 Betrayed Expectations: Misdirected Anger and
the Preservation of Ideology
Barrett Emerick and Audrey Yap
- 371 The Right to Emigrate: Exit and Equality in a
World of States
Daniel Sharp
- 409 Institutional Conservatism and the Right to
Exclude
Hallvard Sandven
- 434 Freedom, Desire, and Necessity: Autonomous
Activity as Activity for Its Own Sake
Pascal Brixel
- 466 Toward a Perceptual Solution to Epistemological
Objections to Nonnaturalism
Preston Werner

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

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

Executive Editor

Mark Schroeder

Associate Editors

| | |
|-----------------------|--------------------|
| Saba Bazargan-Forward | Hallie Liberto |
| Stephanie Collins | Errol Lord |
| Dale Dorsey | Tristram McPherson |
| James Dreier | Colleen Murphy |
| Julia Driver | Hille Paakkunainen |
| Anca Gheaus | David Plunkett |

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

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

Managing Editor

Rachel Keith

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

DEFENSIVE KILLING BY POLICE

ANALYZING UNCERTAIN THREAT SCENARIOS

Jennifer M. Page

ALL SELF-DEFENSE is undertaken under uncertain circumstances. If amid a violent encounter, someone pulls out and unlocks a gun, takes aim, and begins to squeeze the trigger, it is always possible for the gun to jam or be out of bullets.¹ However, some self-defense scenarios are far more uncertain, where a person has not revealed a clear intent to use deadly force but makes a movement indicating that they could be about to draw a gun. From the standpoint of a would-be defender, waiting to see what is in the prospective attacker's hands increases the odds of being killed. By the time a gun is visible, the attacker is already in a position to fire, and it may be too late to retrieve one's own weapon to use in self-defense.

In the United States in most jurisdictions, police officers are permitted—and sometimes trained—to use lethal force in scenarios where there is suspicious movement but no visible weapon.² Urey Patrick and John Hall, two career FBI agents and experts on the police use of force, write in a prominent practitioner handbook on police self-defense:

- 1 Ripstein, *Equality, Responsibility, and the Law*, 193; Alexander, "A Unified Theory of Preemptive Self-Defense," 1476–79; Ferzan, "Justifying Self-Defense," 714–15.
- 2 The ~18,000 police departments in the US operate with considerable autonomy. Individual departments may appear to prohibit lethal force if a civilian is not visibly displaying a weapon. For example, a recently revised use-of-force policy for the Cleveland Division of Police states that "officers shall use force only as necessary, meaning only when no reasonably effective alternative to the use of force appears to exist." Deadly force "may be used only if a subject, through their own actions, poses an imminent threat of death or serious physical harm to an officer or another" (Cleveland Division of Police, "Use of Force," 1–4). It would not seem necessary, and the threat posed by a civilian would not seem to be imminent, if an assumed lethal weapon is not visible. However, as the passages from use-of-force trainers quoted above show, requiring that officers use lethal force only when necessary against imminent threats does not necessarily preclude firing at a civilian who makes a suspicious movement, depending on how "necessity" and "imminence" are interpreted.

It is the reasonable belief in the “imminent danger” that creates the “necessity” for deadly force because only deadly force promises to be effective enough within the crucial time constraints needed to protect against imminent danger. “Imminent” means simply that the danger could happen at any moment—it need not have happened, or be happening yet, but *could* happen at any moment. . . . The best use of justified deadly force is preemptive. That means that it is timely enough, and effective enough, to prevent an *imminent* risk of serious injury (about to happen) from becoming a *definite* attempt to cause serious injury (in fact happening). . . . One aspect of deadly force training involves educating police officers that “imminent risk” is reasonable and real much sooner in a confrontation than they may realize.³

Federal Law Enforcement Training Centers (FLETC) materials similarly emphasize the need for police officers to act quickly in response to movements indicative of threats. “Some may remember the old television westerns where the good guy always let the bad guy go for his gun first,” Tim Miller, FLETC use of force expert, writes. “The fact is, action is faster than reaction. Letting someone reach for a gun may be too late for the officer. . . . [Constitutional law] allows officers to react to the threat of violence rather than violence itself.”⁴ Indeed, as the Eighth Circuit appellate court pronounced in *Thompson v. Hubbard*, “An officer is not constitutionally required to wait until he sets eyes upon the weapon before employing deadly force to protect himself against a fleeing suspect who turns and moves as though to draw a gun.”⁵

The present policy norm of permitting police officers to respond to suspicious hand gestures with lethal force can be questioned, however. This is a circumstance where police officers sometimes erroneously kill unarmed persons. But this does not mean that such killings are automatically unjustified. Experts who defend police use of lethal force in what I call “uncertain threat scenarios” have a point. In a country like the US, where guns are widely available, it is true that certain kinds of hand gestures and movements could be indicative of a lethal attack on the officer or other people. Reaction-time studies have moreover illustrated the truth of “action beats reaction.” Even highly trained police officers with their guns drawn take longer to perceive that a civilian is firing on them and fire their own gun than it takes for a civilian to fire.⁶

3 Patrick and Hall, *In Defense of Self and Others*, 100–3.

4 T. Miller, “Introduction.”

5 *Thompson v. Hubbard*, 257 F.3d 896, 899 (8th. Cir. 2001).

6 Blair et al., “Reasonableness and Reaction Time.”

This article evaluates police defensive-force policy norms in uncertain threat scenarios, assessing the justifiability of present norms from a moral perspective.⁷ This is a novel undertaking in at least two respects. Much of the contemporary philosophical literature on the ethics of defensive force examines the context of war, with very little said about the policing context.⁸ This is surprising, since self-defense is one of the primary reasons for the police use of force.⁹ Further, though the philosophical self-defense literature has burgeoned in recent years, uncertain threat scenarios have not been specifically examined.¹⁰

The inquiry begins by examining differences between the police and civilian defensive force contexts, motivating the present focus on police self-defense against uncertain threats. From there, uncertain threat scenarios are introduced, followed by a discussion of what we know from existing (inadequate) data about how often US police mistakenly kill unarmed persons. Finally, police self-defense policy norms for uncertain threats are explored in three contexts: (1) known in-progress violent crimes, (2) interactions with civilians behaving non-aggressively, and (3) interactions with civilians behaving aggressively. As I argue, a norm permitting police officers to use lethal force in uncertain threat scenarios is morally justifiable in context 1. In contexts 2 and 3, the case for such a norm is extremely weak.

1. POLICE VS. CIVILIAN SELF-DEFENSE: SOME DIFFERENCES

Are uncertain threat scenarios unique to the policing context? In one respect, the answer is clearly no. If Michael is in an argument with his neighbor Charlie and suspects that she is about to attack him with a concealed firearm, Charlie's moving her hand toward her waistband might be the last piece of evidence Michael uses to determine that an attack is imminent before drawing and firing

- 7 In focusing on use-of-force policy norms rather than general considerations of permissibility, I take inspiration from Jorgensen's work on self-defense norms (which was published under the name Renée Jorgensen Bolinger). See Bolinger, "Reasonable Mistakes and Regulative Norms" and "The Moral Grounds of Reasonably Mistaken Self-Defense."
- 8 Among the exceptions are Fabre, "War, Policing, and Killing"; and S. Miller, *Shooting to Kill*.
- 9 Harmon, "When Is Police Violence Justified?"
- 10 However, philosophical work on mistaken self-defense (e.g., Bolinger, "Reasonable Mistakes and Regulative Norms" and "The Moral Grounds of Reasonably Mistaken Self-Defense") and merely apparent threats (e.g., Ferzan, "The Bluff") is relevant. Lazar addresses questions of uncertainty in the defensive force context; his analysis implicitly assumes a continuum between non-threats, uncertain threats, and near-certain threats ("In Dubious Battle"). I agree wholeheartedly that there is a continuum, but nevertheless will try to show that it is meaningful in the policing context to conceptualize uncertain threats as distinct defensive force phenomena.

his weapon. However, police shootings of civilians who make suspicious movements are particularly salient, occupying a place in the popular imagination and in media commentary.¹¹ What is more, some features of police self-defense make it likelier that police officers, compared to their civilian counterparts, encounter situations where they discern an uncertain threat and defensively use lethal force. These features have to do with (1) the nature of the police role, (2) the kinds of signals police officers are trained to pick up on, and (3) the kinds of behaviors that the exercise of police authority sometimes provokes.

Police officers, by virtue of their role, face distinctive kinds of unpredictability and danger. They encounter violent criminal suspects in their capacity as violent criminal suspects and are assigned the duty of standing their ground on the state's behalf.¹² The role of an officer making an arrest is thus directly and immediately coercive, and it is natural that some criminal suspects do what they can to avoid apprehension, including using a lethal weapon against an officer.¹³ Even a non-criminal suspect questioned by an officer might be concealing something that could be grounds for arrest. Attacking an officer might disable them (either temporarily or permanently) and allow for an escape, so a life-threatening fight response could come seemingly out of nowhere. Outside the law enforcement sphere, individuals typically do not face the prospect of violent resistance in performing their standard employment-related duties.

Police officers learn how to deal with potentially life-threatening situations in defensive force training programs.¹⁴ Part of this training is intended to improve upon officers' native threat-detection abilities, so that nuances like a slight sag of the clothing are noticed.¹⁵ However, this also means that police

- 11 For example, the popular 2018 movie *The Hate U Give* centers around the fictional police shooting of a young Black man, Khalil, who is killed after reaching into his car for a hairbrush. The relationship between uncertain threat scenarios and mistaken object scenarios is discussed in section 2; the description of an uncertain threat scenario fits because Khalil is shot so quickly. See also Balko, "When Unarmed Men Reach for Their Waistbands"; Delaney and Jeanty, "Police Shootings of Unarmed Men Often Have Something in Common"; Chung, "U.S. Top Court Won't Review Houston Police Shooting 'Waistband' Defense." Irvin identifies "reaching for the waistband" as a type of post-facto rationalization for shootings of unarmed Black men by police ("Policing, Racialization, and Resistance").
- 12 Donnelly, "Police Authority and Practices"; Harmon, "When Is Police Violence Justified?"
- 13 Bittner, "The Functions of the Police in Modern Society"; Donnelly, "Police Authority and Practices"; Alpert and Dunham, *Understanding Police Use of Force*. In this article, I set aside issues related to citizen's arrests and stand-your-ground laws.
- 14 Ho, "Individual and Situational Determinants of the Use of Deadly Force." This being said, in the vast majority of US police-civilian encounters—99 percent—no force whatsoever is used (Bureau of Justice Statistics, "Contacts between Police and the Public").
- 15 Dorn and Dorn, "Seven Signs a Weapon Is Being Concealed." However, though there is not extensive research on the subject, one experimental study found that novice and

officers are in an epistemic position to perceive innocuous behavior as threatening if it lines up with the danger signs they are trained to pick up on. Nick Jacobellis recounts almost firing on an innocent couple:

While looking for smugglers on Key Biscayne one night, I identified myself and asked a male passenger in a car that was stopped near a boat ramp if there were any firearms in their vehicle. Instead of saying, “Yes, officer,” and telling me where his pistol was located, the passenger said absolutely nothing as he leaned forward very quickly, enough to startle me, and opened the glove compartment of the car. I raised my service pistol with my right hand and yelled something like, “US Customs! Don’t move . . . ! Don’t move!” Adding to the chaos, my partner yelled commands at the woman behind the steering wheel who was, of course, screaming at her boyfriend. Seconds passed like hours, as I prepared to shoot. Fortunately, the young male passenger froze just as he started to reach inside the glove compartment. I reached in and recovered a Walther PP from the glove compartment, and disassembled it on the hood of the car. Once we cleared the couple of any wrongdoing, my partner and I left the area after learning a very important lesson. . . . Some law abiding people simply don’t realize that their actions can appear threatening to the police.¹⁶

In spite of having had this experience and recognizing the complexities of knowing who the “good guys” and “bad guys” are in a country where concealed carry is legally permitted, Jacobellis—who is writing to his fellow law enforcement officials as a *Police Magazine* contributor—makes it a point to say multiple times that anyone could be armed: “Write this in bold block letters somewhere across your mind: You cannot assume that someone is unarmed. . . . Just remember that you can’t assume that even a jaywalker is unarmed. So as you approach a subject, suspect, or violator you must be prepared to go tactical at a moment’s notice.”¹⁷ Though police officers are trained to distinguish threats from non-threats, the upshot of this training is the possibility of construing a threat when there is none.

The way police officers exercise their authority can also produce the appearance of a threat. Since police officers are trained to always be on guard against

experienced police officers were no better than psychology students at detecting a concealed weapon carried by a male walking into a courthouse. Moreover, experienced officers were likelier to say that the subject was in fact concealing a weapon, regardless of whether or not this was true (Sweet, Meissner, and Atkinson, “Assessing Law Enforcement Performance in Behavior-Based Threat Detection Tasks Involving a Concealed Weapon or Device”).

16 Jacobellis, “How to Spot a Concealed Firearm.”

17 Jacobellis, “How to Spot a Concealed Firearm.”

an attack, this itself might trigger behaviors indicative of the very kind of threat the officer is aiming to avoid.¹⁸ Again, Jacobellis: “Your job is to ensure that you always remain in a position to exert complete control at all times,” he writes, explaining how to visually monitor the situation at hand. “Quickly scan a subject’s eyes and hands then scan the area around you. . . . Repeat this process until you complete the stop or field interview.”¹⁹ It can be highly unnerving to interact with an armed authority figure who avoids eye contact and treats you as if you might pose a threat to their life. This may provoke nervous verbal responses, fidgeting, or other irregular behavior on the part of civilians who pose no threat, which may be indistinguishable from behaviors typical of civilians who do pose a threat. These behaviors may alert the officer to the possibility that the civilian is concealing a weapon, and a subsequent hand gesture toward the pocket or waistband may prompt the officer to defensively use lethal force. Indeed, social psychologists have found this to be especially true of encounters between police officers and Black people due to “stereotype threat.” The racial stereotype of Black men as violent is longstanding in US society.²⁰ Stereotype threat refers to changes in person’s cognition and behavior when their identity and associated stereotypes become salient.²¹ Researchers have found that Black men asked about their awareness of the Black male stereotype are particularly likely to exhibit nervousness and ostensibly suspicious behavior in interactions with police officers.²² By treating Black men as potentially threatening, police officers can activate threatening-seeming behavior, if the very fact of the encounter has not activated this kind of behavior already.

These last two points about the signals police officers are trained to pay attention to and how police authority might induce threatening-seeming behavior give some context for why a police officer may end up mistaking a harmless movement for the beginning of an attack. However, this itself does not

18 Obviously, there are also cases where officers antagonize civilians by yelling, swearing, hurling insults or racial slurs, being unnecessarily physically aggressive, adopting a macho demeanor, and so on—the stereotypical bad cops of Hollywood films are sometimes found on the street. See, e.g., many of the police encounters described by Chicago youth in Futterman, Hunt, and Kalven, “Youth/Police Encounters.” “Officer-created jeopardy” situations are when an officer’s bad tactical decisions put them at heightened risk of harm (Lee, “Officer-Created Jeopardy”). It is clear that forms of police misconduct such as these might lead civilians to act in ways that police officers perceive as threatening. But appropriate and routine police behavior can do this too.

19 Jacobellis, “How to Spot a Concealed Firearm.”

20 Muhammad, *The Condemnation of Blackness*.

21 Steele, *Whistling Vivaldi*.

22 Najdowski, Bottoms, and Goff, “Stereotype Threat and Racial Differences in Citizens’ Experiences of Police Encounters.”

invalidate a policy of permitting police officers to use lethal force in uncertain threat scenarios. In using defensive force, mistakes are always a possibility. If we wished to eliminate all mistakes in self-defense, we would have to eliminate all self-defense.

2. CONCEPTUALIZING UNCERTAIN THREAT SCENARIOS

The terminology of “uncertain threats” is useful for thinking about police policy norms. Whereas terms like “waistband shootings” and “cell phone shootings” refer to the police shooting of an unarmed person based on the mistaken perception of a threat, the concept of an “uncertain threat” encompasses both eventual mistakes and non-mistakes.²³ What are uncertain threat scenarios? The clearest explanation involves a contrast with “ordinary threat” scenarios. Here, there is an agent who is *liable* to defensive harm, and thus is “not wronged by its infliction—she has no justified complaint against being harmed—and she may not, ordinarily, harmfully defend herself against its infliction.”²⁴ Authors disagree about whether it is culpability; moral responsibility for posing a threat to another’s life; moral responsibility for failing to avail oneself of a reasonable opportunity to avoid posing a threat; treating others as if they lack moral rights against harm that they do in fact possess; or something else that renders an individual liable.²⁵ Nevertheless, liability means that an individual may not justifiably claim a right against harm when defensive force is used against her, and is not entitled to compensation from a self-defender.

In uncertain threat scenarios, however, from the standpoint of the evidence the self-defender has access to, the liability of a potentially threat-imposing civilian is uncertain.²⁶ If the civilian is in fact reaching for a gun with the inten-

23 For “waistband shootings” and “cell phone shootings,” see Aveni, “The MMRMA Deadly Force Project”; Bobb, “The Los Angeles County Sheriff’s Department 30th Semiannual Report”; and Taylor, “Dispatch Priming and the Police Decision to Use Deadly Force” and “Engineering Resilience.”

24 Frowe, *Defensive Killing*, 72.

25 Ferzan, “Justifying Self-Defense” and “Culpable Aggression”; Burri, “Morally Permissible Risk Imposition and Liability to Defensive Harm”; McMahan, “The Basis of Moral Liability to Defensive Killing”; Gordon-Solmon, “What Makes a Person Liable”; Frowe, *Defensive Killing*, ch. 3; and Quong, *The Morality of Defensive Force*.

26 Maybe “uncertain threat scenarios” is not the best term, because it could connote cases where, e.g., Innocent Threat is launched through the air against their will by Villain, who loudly announces this. The probability that they will land on Person Stuck in Well, crushing them to death within seconds, is below the probability typical in ordinary Innocent Threat scenarios, but is not so low that we can automatically rule out the permissibility of Person Stuck in Well pushing a vaporizer button to obliterate Innocent Threat midair—if

tion to use it to kill a police officer, they are the prototypical aggressor who is liable to being killed in self-defense. However, they may also be non-labile. If a police officer making an arrest commands a civilian to stop and put their hands in the air, they may be attempting to comply, but arm fatigue may cause them to drop their hands, or they may have an involuntary reflex to scratch an itch. If the officer inflicts harm and acts permissibly in doing so, this must be in virtue of something other than the civilian's liability.

Accordingly, in an ordinary threat scenario, there is an *A* who ϕ s, and ϕ -ing is the type of action that, on the evidence, indicates that *A* has the intent and means to seriously harm or kill *B* in the next moment. In an uncertain threat scenario, on the evidence, *A*'s ϕ -ing *could* indicate that *A* has the intent and means to seriously harm or kill *B* in the next moment—the probability of this being the case falls short of the probability typical in ordinary threat scenarios, but it is not so low that we can automatically rule out the permissibility of *A* defensively using lethal force.²⁷

In both ordinary and uncertain threat scenarios, liability to being killed is not the only justification available for the use of defensive force. There are also lesser evil justifications, where a person not known to be liable is killed for the sake of a greater good—say, saving an appropriately high number of other persons.²⁸ Moreover, as I have phrased things, *B* could be the individual deciding whether to use lethal force in self-defense, or there could be another person, *C*, who faces an other-defense decision about saving one or more persons in *B*'s position.²⁹

Again, one reason why it is controversial for police officers to use lethal force in uncertain threat scenarios is the risk of mistakenly shooting an unarmed, non-labile person. In “cell phone shooting” cases, an unarmed person is shot while reaching for an object like a phone or wallet.³⁰ In some cases, a police officer might see the non-weapon object and still shoot. Is this an uncertain

it is permissible to defensively kill Innocent Threat, that is. This would not qualify as an uncertain threat scenario on the definition I give, since it is known that Innocent Threat has no intent to crush Person Stuck in Well.

- 27 The terminology I am using is a simplification of the wordier terms “ordinary *lethal* threat scenarios” and “uncertain *lethal* threat scenarios”: I am not thinking of cases that potentially involve a hard push or shove, but rather, a threat to a person's life.
- 28 McMahan, *Killing in War*; Bazargan, “Killing Minimally Responsible Threats”; Frowe, *Defensive Killing*; Lazar, “In Dubious Battle.”
- 29 Zimring finds that, in over 95 percent of cases of police killings, police officers used lethal force to protect the lives of police officers. He does not disaggregate this statistic into officers protecting themselves versus officers protecting other officers (*When Police Kill*, 63).
- 30 Taylor, “Dispatch Priming and the Police Decision to Use Deadly Force” and “Engineering Resilience.”

threat scenario? The answer depends on the details. In the police academy, simulation-based training exercises teach cadets to recognize cues that a person is armed, form the right judgment about whether a civilian is reaching for a weapon, and shoot persons reaching for a weapon before they fire on the officer.³¹ All this happens extremely fast. In a slow-motion version of events, we may be able to see a chronology where a police officer makes the decision to shoot, activating a sequence where they unholster their gun, unlock the safety, aim, and pull the trigger. If sometime during this sequence the civilian makes the object they are holding visible and it is a phone, it may be too late for the officer to hold their fire. Once the sequence is in motion, training-based muscle memory kicks in, and it can be very hard to halt the brakes.³² It may even be that the decision to fire is made a split second after a phone is made visible, because the officer processes the object so quickly that he misperceives what it is. An error of this kind fits the parameters of an uncertain threat scenario: the civilian's ϕ -ing is their quickly taking an object out of their pocket that could be a gun. However, this is different from mistaken threat scenarios where a non-weapon object is visible all along and a police officer fires because they misperceive it as a weapon. The latter scenario—an evidence- and fact-relative non-threat paired with a belief-relative perception of a threat—does not count as an uncertain threat scenario.³³

In the policing context, ϕ -ing often refers to dropping one's hands toward one's waistband, touching one's waistband, putting one's hand in one's pocket, reaching into a car window or a compartment of one's car, and so on. People use guns with their hands, so police officers are trained to be alert to what a person's hands are doing.³⁴ Uncertain threat scenarios are broader than cases of suspicious hand gestures, however. Every year individuals are killed by US

31 Ho, "Individual and Situational Determinants of the Use of Deadly Force"; *Economist*, "Simulators Teach Police."

32 Aveni, "The MMRMA Deadly Force Project," 18–19.

33 For more on the language of "belief-relative," "evidence-relative," and "fact-relative," see Parfit, *On What Matters*, ch. 7.

34 Siegfried, "How to Watch the Hands." It does not have to be unknown whether a person is armed with a gun for the situation to count as an uncertain threat scenario. In the US, it is legal for most people to carry concealed weapons and it cannot be assumed that every armed person intends to harm others. Situations where a civilian interacting with a police officer visibly has a weapon, informs the officer that they have a weapon, is reported to have a weapon by someone who has phoned this into a police dispatcher, or is licensed to carry a concealed weapon according to a database consulted by the officer, all have the potential to be uncertain threat scenarios. However, because section 3 is already lengthy, I do not specifically consider a scenario where, on the evidence, someone being questioned or resisting arrest has a gun.

police because their vehicles are perceived as lethal weapons.³⁵ If an aggressor with lethal intent is accelerating in an officer's direction and the latter cannot move away, this is an ordinary threat scenario where the driver is liable to being killed. However, in some cases a driver moves their car in a way that could indicate their intent to seriously harm or kill the officer—running the officer over or dragging them—but it is nevertheless ambiguous as to whether this is the case. This is an uncertain threat scenario, even though this kind of ϕ -ing is quite different from ϕ -ing in waistband cases.

We do not have data—from the laboratory or the real world—on uncertain threat scenarios exactly fitting the parameters described above. Still, there has been enough interest in police shootings where unarmed civilians make suspicious hand gestures or are holding non-weapon objects that we can get a rough empirical sense of the phenomenon. In a study conducted by Aveni, actors were videoed turning and reaching into their clothing. They then displayed either a weapon, an object like a wallet or phone, or were empty handed. The lighting conditions were poor, simulating nighttime conditions typical of many police shootings, and the broader context of the civilian's movement was a burglary, mugging, or robbery. Collectively, 307 officers participating in the study shot 38 percent of unarmed persons.³⁶

A 2011 report about the Los Angeles County Sheriff's Department (LASD) examined the prevalence of "state-of-mind" shootings, where civilians "were perceived to be reaching for or holding a firearm, but were not confirmed to be holding a weapon at the time the shooting occurred."³⁷ The report found that 21 percent of all shootings from 2005 to 2010 fit this description, disproportionately of Black and Latino men. The majority (61 percent) were confirmed as unarmed immediately after they were shot; a small minority (4 percent) were confirmed as armed.³⁸ Though the 61 percent unarmed statistic suggests a much higher error rate than in the lab, if a person makes a sudden hand gesture and it turns out that they are in fact pulling out a gun, officers are unlikely to say that they fired their weapons based on a suspicious movement. Instead, they will probably say that seeing a weapon was the reason they fired, and the case will be classified as a standard police shooting of an armed suspect, not a

35 Lowery, Bever, and Mettler, "Police Have Killed Nearly 200 People Who Were in Moving Vehicles since 2015, including 15-Year-Old Jordan Edwards."

36 Aveni, "The MMRMA Deadly Force Project."

37 Bobb, "The Los Angeles County Sheriff's Department 30th Semiannual Report," 51.

38 The rest escaped apprehension and were unarmed later upon arrest (13 percent), discovered to have discarded a weapon nearby (9 percent) or earlier (4 percent), or had an unknown weapon status (8 percent) (Bobb, "The Los Angeles County Sheriff's Department 30th Semiannual Report," 58–59).

state-of-mind shooting. Despite this, the LASD data gives a picture of the overall rate at which LASD officers use lethal force against unarmed, likely non-liable persons—at minimum, 12.8 percent of persons in the dataset.³⁹ This is higher, however, than a Philadelphia dataset that considers “threat perception failure” shootings, where an officer wounds or kills an unarmed civilian because of a movement or misperceiving a non-weapon object—7.3 percent of police shootings from 2007 to 2013 fit this description, accounting for around half of all shootings of unarmed people, disproportionately of Black civilians.⁴⁰

Unfortunately, there is no national-level equivalent of the Los Angeles County and Philadelphia reports, since the US federal government does not have a mandatory reporting system for tracking the police use of lethal force. Several crowdsourced databases have stepped in.⁴¹ According to six years of data collected by the *Washington Post*'s Fatal Force project, 6.4 percent of all fatal police shootings are of unarmed persons.⁴² This means that the national rate at which unarmed people are killed by police is lower than the rate of threat perception failure shootings in Philadelphia. What accounts for this? Geography may be playing a role: the highest rates of killings by police are in the African American neighborhoods of large cities and in rural regions of the West and Midwest.⁴³ Perhaps more crucially, however, unlike Los Angeles County and Philadelphia, the *Washington Post* database does not consider nonfatal police shootings—no national-level data exists, crowdsourced or otherwise, on people wounded by police firearms. As Zimring discusses, when individuals

39 This calculation is based on the statistic that 61 percent of state-of-mind shootings were of persons immediately confirmed as unarmed who had not discarded a weapon.

40 Fachner and Carter, “Assessment of Deadly Force,” 2–3, 30–33.

41 The *Washington Post*'s Fatal Force project, the *Guardian*'s The Counted project (2015–2016), and Fatal Encounters, run by D. Brian Burghart and team, are the most prominent. The *Washington Post* collects data on police shooting deaths, whereas the *Guardian* collected data on all non-self-inflicted police killings, e.g., deaths by shooting, Taser, chokehold, and police vehicle. The Fatal Encounters database is the broadest and includes self-inflicted deaths in police custody or during a police pursuit. Here I reference the *Washington Post*'s statistics because of its exclusive focus on police shooting deaths. When someone moves their hands toward their waistband from the surrender position, police officers potentially respond by firing their weapons, not by putting the civilian in an asphyxiating chokehold.

42 This statistic is from early July 2021. At this point in time, 6,419 people had been recorded as fatally shot by US police since 2015: besides those counted as unarmed, 58.7 percent were armed with a gun, 17.1 percent had a knife, 3.3 percent were counted as being armed with a vehicle, 3.6 percent had a toy weapon, and the weapon status of the rest—8.2 percent and 2.7 percent—were classified as “other” and “unknown,” respectively (*Washington Post*, “Fatal Force”). US police deaths are overwhelmingly caused by firearms: of the 58.4 officers killed annually between 2008 and 2013, 92 percent were fatally shot (Zimring, *When Police Kill*, 95).

43 O'Flaherty and Sethi, *Shadows of Doubt*, ch. 8.

are shot, there is a positive correlation between the number of bullets that hit and the death rate.⁴⁴ Though this is speculative, perhaps when police officers are in uncertain threat scenarios, they fire fewer bullets. This would result in a higher rate of overall use of lethal force in response to suspicious movements compared to fatal uses alone.

3. POLICE SELF-DEFENSE AGAINST UNCERTAIN THREATS ACROSS CONTEXTS

We are looking for the right police policy norm for uncertain threat scenarios. Different kinds of uncertain threat scenarios raise different kinds of considerations about what police officers should do. Let us examine three contexts: (1) known in-progress violent crimes; (2) interactions with civilians behaving non-aggressively; and (3) interactions with civilians behaving aggressively.

3.1. *Known in-Progress Violent Crimes*

Consider a case, *Bank Robbery*, where police officers dispatched to a bank robbery are told that the suspect has a gun and has already shot a teller. They arrive and a male fitting the suspect's description is yelling orders. One officer commands the suspect to freeze and put his hands in the air. The suspect, who is not visibly displaying a weapon, does not freeze and moves his hands downward in the direction of his waistband. When the context of an uncertain threat scenario is a known in-progress violent crime, what policy norms should govern the police use of lethal force?

To begin with, the police clearly have probable cause to arrest the robbery suspect. It is important to distinguish between force used to make arrests and defensive force. The suspect's hand gesture could indicate other crimes that he may have to answer for in a court of law—e.g., (attempted) assault on an officer—but apprehending him for this and for alleged crimes committed beforehand is a separate matter from police officers defending themselves and bystanders.

Despite this, the robbery suspect's immediately prior alleged crimes are not irrelevant to the police officers' decision whether to use force. They factor into the threat probability- and fairness-based considerations that should be taken into account. From a threat probability standpoint, the police know that the suspect is reportedly armed and violent; that he has already shot a teller says something about his willingness to use lethal force against innocent persons. Compared to, say, a compliant driver at a traffic stop who makes an equivalent hand gesture, there is greater evidence of the robbery suspect's gesture indicating his liability to being defensively killed.

44 Zimring, *When Police Kill*, 63–69.

From a fairness perspective, though a firearm is not visible, the robber's prior actions create a situation where the police are not unreasonable to perceive him as having a gun. There is a fine line to be walked here. On one hand, as a criminal suspect, the robber is presumptively innocent. The police may not make an affirmative judgment on his guilt on matters that extend beyond the immediate threat he may or may not pose—there is truth to the adage that we do not want the police to be the judge, jury, and executioner when it comes to criminal conduct.⁴⁵ On the other hand, his prior violent actions, which also happen to be criminal actions, are the basis of the evidence the police have for gauging a higher-than-usual probability of his suspicious hand gesture indicating a lethal threat. These actions matter not only from a descriptive threat probability standpoint, but also have further normative meaning. Say the robber, after shooting the teller, ditches his gun, which the police do not know when they spot him moving his hands downward. Though he no longer poses a threat, he still seems to have changed the normative landscape through his prior actions.

A case discussed by Ferzan is relevant here: a robber points an unloaded gun at a 7-Eleven clerk, demands money, and says he is going to murder him. Though the threat is insincere, Ferzan points out that we often consider insincere actions as altering our rights and duties. An insincere promise is still a promise. The insincere abandonment of property is still the abandonment of property. It is thus plausible to say that the 7-Eleven robber forfeits his rights against being defensively killed by the clerk.⁴⁶ On Jorgensen's account, the 7-Eleven robber's actions similarly entail a forfeiture of certain rights. She rejects the language of liability, but explains that such an individual makes themselves "vulnerable" to defensive harm.⁴⁷ From a risk distribution standpoint, it would be unfair for the 7-Eleven robber to retain his rights against harm, so the clerk does not wrong him by killing him.⁴⁸

45 It is widely thought—though not universally accepted (see Tadros, *The Ends of Harm*)—that defensive harm and punitive harm are separate matters. If Thief starts attacking Victim to steal their wallet, and Victim averts the attack by punching Thief once, if Victim delivered further blows, it is intuitive to say that this is punitive and thus unjustified (Frowe, *Defensive Killing*, 108–9). Fletcher goes further and argues for the defensive–punitive harm distinction on the grounds that private self-defense is instrumentally necessary for maintaining a system of social cooperation, whereas state punishment serves the cause of justice ("Punishment and Self-Defense"). Of course, here we are talking about police self-defense rather than private self-defense, but the underlying logic of the defensive–punitive harm distinction would seem to still stand. See also McMahan, *Killing in War*, 67.

46 Ferzan, "The Bluff."

47 Bolinger, "The Moral Grounds of Reasonably Mistaken Self-Defense," 4.

48 For Jorgensen, reasonable mistakes can sometimes wrong a victim, but I take Ferzan's 7-Eleven robber case to be equivalent to Jorgensen's *Stalker* case where the mistakenly

There are normatively important differences between the 7-Eleven robber and the suspect in Bank Robbery, however. Whereas all evidence points to the 7-Eleven robber posing an ordinary threat in the moment of being defensively killed, the meaning of the bank robber's hand gesture is uncertain from an evidentiary standpoint. The 7-Eleven robber's actions (pointing a gun, stating that he will murder the clerk) have the widely recognized function of communicating the presence of a lethal threat. Moving one's hands downward, even in the context of an armed bank robbery where the suspect is not following police orders, fall short of this standard. Again, the movements police officers pick up on as signals that a person is going for a gun are often so subtle that agents cannot be expected to avoid making them; it cannot be assumed that they are made intentionally. It could be that the robber is about to put his hands in a surrender position, but through some neuromuscular fluke, his hands slightly drop downward on the way to going upward. The bank robber thus would not seem to have forfeited his rights against harm in the same way the 7-Eleven robber has.

Nevertheless, the bank robber would seem to have forfeited *something* due to his actions prior to making the suspicious hand gesture. This something—here invoking a formulation more in line with Jorgensen than Ferzan—is his claim-right to being treated as having a moral status equal to that of an innocent person in determining a fair allocation of risk.⁴⁹ Forfeiting such a right would not seem

defended-against party is not wronged. See Bolinger, "The Moral Grounds of Reasonably Mistaken Self-Defense," 2.

- 49 With such a formulation, would I consider the Bank Robbery suspect *liable* to being defensively killed? As I have conceptualized uncertain threat scenarios, the uncertain element is precisely a civilian's liability: a driver at a traffic stop who in fact reaches for a gun to use against a police officer is liable to being killed; a driver reaching for his wallet is not (Page, "Reparations for Police Killings," 960–61). At first glance, this would also seem to be the case for the suspect in Bank Robbery: his posing a fact-relative threat is a necessary condition for him to be liable. But maybe things are not so simple. Accounts by Ferzan ("The Bluff") as well as McMahan ("Who Is Morally Liable to Be Killed in War," 555–56), and Quong (*The Morality of Defensive Force*, 42–45)—but not Bolinger, "The Moral Grounds of Reasonably Mistaken Self-Defense," 4), or Frowe (*Defensive Killing*, 85–86)—endorse the language of liability to describe the moral situation of the bluffing 7-Eleven bank robber even though he poses no fact-relative threat. Is the Bank Robbery suspect liable to being defensively killed based on his actions leading up to his hand gesture, despite his hand gesture's meaning being uncertain?

For my own part, I am reluctant to say that an unarmed suspect who drops his hands as a result of a neuromuscular fluke is liable to being killed for reason of his prior actions; I think that this objectionably collapses the distinction between defensive harm and punishment. However, suppose the suspect's hands are in the surrender position and he intentionally drops them a few times, repeating to the police officers, "Wanna fight? Wanna fight?" If he is unarmed, he is in the same position as the bluffing 7-Eleven robber; here I *am* fine saying that he is liable to being killed. The difference seems to be the degree of

to be an all-or-nothing affair. There are possible versions of Bank Robbery where the robber's wrongful actions vary in their severity, e.g., shooting five bank tellers versus pistol-whipping one teller. In all cases it seems appropriate that the robber bear a level of risk that is greater than the innocent persons who are present at the crime scene. However, if the pistol-whip does not do much harm, this would not seem to reduce the robber's moral status very much, whereas an individual who commits an extremely morally serious wrong (like shooting any number of bank tellers) may effectively be in a morally indistinguishable position from the individual who intentionally presents himself as posing, or who in fact poses, an ordinary threat. The rights against harm that this individual retains are so minimal as to be practically nonexistent. The benefit of this overall formulation is that it allows for sensitivity to circumstantial differences in different kinds of violent crime-based uncertain threat scenarios.

The threat probability- and fairness-based considerations that factor into police officers' decisions whether to shoot a violent criminal suspect who poses an uncertain threat are not the only considerations in play. Police officers have an extremely difficult job in a situation like Bank Robbery: not only are they required to instantaneously recognize the nature of the suspect's hand gesture, but they must also have enough situational awareness to factor in at least two other considerations. First, how many bystanders' lives are potentially at risk? Police officers with good tactical skills will have chosen a position from which to confront the suspect where, if the suspect begins shooting in their direction, the bullets are less likely to hit innocent bystanders. Standing in front of twenty bank customers is a bad idea; standing in front of no one, taking cover if

agency exercised by the suspect in making the movement that the police consider decisive in determining that he poses an imminent lethal threat. A willful movement seems to render the suspect liable. A non-willful movement is different—his movement is equivalent to the bodily movements made by the non-liable Innocent Threats of the self-defense literature (e.g., Otsuka, "Killing the Innocent in Self-Defense"; McMahan, *Killing in War*, ch. 4; Quong, *The Morality of Defensive Force*, ch. 3; Burri, "The Toss-Up between a Profiting, Innocent Threat and His Victim"; Frowe, *Defensive Killing*, ch. 2)—except that if he unarmed, he is not even a threat, so why should we consider him liable? Nevertheless, his right to be treated as having a moral status equal to that of an innocent person in determining a fair allocation of risk has been so diminished that in practice he is indistinguishable from his liable counterpart. With such a formulation, I am happy to follow Jorgensen and label such an individual "vulnerable" to being defensively killed rather than using the language of liability (Bolinger, "The Moral Grounds of Reasonably Mistaken Self-Defense"). This leaves it open for debate as to whether, e.g., he may claim compensation for his injuries if he survives the police shooting. See also Frowe (*Defensive Killing*, 85–86), who emphasizes that only individuals who pose a fact-relative threat are potentially liable to *defensive* harm; nevertheless, there are cases where non-threatening individuals acting unjustly are liable to other kinds of harm besides the defensive kind.

possible, is much better. However, tactical positioning may not always be possible, making it necessary to consider how many innocent lives are potentially in danger if the police wait to verify that the suspect is in fact drawing a gun before they fire. Second, what is the distance between the suspect and the police officers, and what are the officers' expected hit rates based on this distance?⁵⁰ How many bystanders could be shot by stray bullets if the police miss their target? Is the robber (perhaps strategically) standing in front of twenty bank customers?

Questions like these matter because there may be a lesser evil justification for using lethal force against the suspect in Bank Robbery. Lazar has stressed the importance of considering lesser evil justifications, and not just liability justifications, when it comes to defensive force decision-making under uncertainty.⁵¹ He argues against a fixed threshold view where, once it is sufficiently likely—say, a probability of 90 percent—that an individual is liable to being killed, this activates a permission to defensively kill them. Rather, just as deontologists admit that an innocent person may be sacrificed in order to save the lives of a sufficiently high number of innocent others, high stakes may justify lowering the threshold for the probability that an individual is liable to being defensively killed. If the robbery suspect is potentially able to harm many innocent people if given the time to draw a gun, depending on the numbers at stake, it may be justifiable for the police to use lethal force against the robbery suspect based on this alone, even if we are considering the pistol-whipping variation on Bank Robbery.

In sum, when we start thinking about the likelihood of the robbery suspect's hand gesture indicating a genuine threat, the degree to which he has forfeited his right to be treated as an innocent person from a risk allocation standpoint, the risks of waiting to verify a weapon and potentially allowing the suspect to fire, and the risk of accidentally shooting bystanders, clearly many factors should influence a police officer's decision to shoot. Nevertheless, in the context of a known in-progress violent crime like Bank Robbery, the threat probability-based, fairness-based, and lesser evil reasons for a police officer to respond to uncertain threats with lethal force form an overall justification for a policy norm permitting this. This does not mean that every police shooting of a violent criminal suspect who poses an uncertain threat is automatically justified—on the contrary—but police officers should be given the latitude to make their best judgment as to whether the context warrants the use of lethal

50 Empirical studies on US policing have long shown that officers in the field do not hit their targets with great accuracy. A recent study of Dallas data shows that only 123 of 354 bullets fired by police between 2003 and 2017 hit their intended target—a 35 percent hit rate (Donner and Popovich, "Hitting (or Missing) the Mark").

51 Lazar, "In Dubious Battle."

force. Without a norm permitting lethal force, police officers would be prohibited from, e.g., firing on an active school shooter who makes a movement to retrieve the weapon he has briefly concealed.

Two caveats. First, what is communicated in police firearms training matters normatively. Instructors ought to emphasize that neither the use of lethal force itself nor a lower evidentiary threshold for using lethal force are permissible as punishment for the suspect's alleged crimes, or based on the idea that the suspect is a "bad guy" whose life has less inherent worth than anyone else's. Second, the specification that the scenario is a *known* in-progress violent crime is extremely important. It puts the evidentiary threshold for a police officer to judge that an in-progress violent crime is occurring much higher than, say, in the case of a "suspected" in-progress violent crime. That the meaning of a suspect's hand gesture is uncertain does not imply that the context can be uncertain as well.

3.2. *Civilians Behaving Non-aggressively*

Consider *Loud Music*. A shop owner calls the police to complain about a man lingering outside her store; there have been a number of recent car break-ins on the street. Two officers respond, see the man outside the shop, and decide to strike up a conversation. The man, unbeknownst to them, is listening to loud music in the wireless headphones hidden by his winter hat. He cannot hear what they are saying and reaches into his jacket pocket for his phone to turn down the music. They yell at him to freeze and slowly show his hands; his jacket pocket is weighted down in a way that could mean he has a gun. But the man cannot hear them and he is not making eye contact. He quickly moves his hand out of his pocket and one officer shoots him.

Loud Music fits in with Jorgensen's work on mistaken self-defense.⁵² A self-defender forms an erroneous belief that another person is about to mount an attack. According to what norms should we judge mistaken self-defenders? Jorgensen points to a major issue with evidentialist norms, which require that self-defenders form reasonable judgments about when to use defensive force on the basis of available evidence.⁵³ Running through a large body of empirical literature on implicit racial bias in the self-defense context, she argues that being Black might be a "perverse signal" that society uses as a heuristic for assessing threateningness. Black people, however, have "a justice-based claim against being put in a position where they appear threatening by default."⁵⁴

52 Here I refer to Bolinger, "Reasonable Mistakes and Regulative Norms."

53 Jorgensen is primarily concerned with the civilian self-defense context but briefly discusses police self-defenders as susceptible to implicit racial biases. Bolinger, "Reasonable Mistakes and Regulative Norms," 206–7.

54 Bolinger, "Reasonable Mistakes and Regulative Norms," 209.

Ideally, we would work as a society to reduce racial biases, but if reform efforts fall short, the next-best alternative is a fact-relative norm. As normatively attractive as it is to give self-defenders space to make reasonable mistakes, because society might be systematically biased in terms of why a mistake counts as reasonable, our remaining option is to hold strictly accountable all self-defenders who harm or kill a person who, like the police-shooting victim in *Loud Music*, objectively posed no threat.

From a law enforcement perspective, applying Jorgensen's proposal to the policing context would be highly controversial. In *Graham v. Connor*, the Supreme Court made eminently clear that the legal standard to be used in assessing the police use of force—the “reasonable officer” standard—is evidence relative, not fact relative.⁵⁵ Aside from this, comparing the police and civilian self-defense contexts, civilians may be more prone to making uninformed defense decisions based on gut feelings: there is little that ensures that civilian self-defenders are knowledgeable about when defensive force is legally and morally justifiable.⁵⁶ For civilians who make a defensive force decision based on an erroneous instinct of feeling threatened, a fact-relative norm could at least promote accountability after wrongful harm has been inflicted. Accountability, however, is a poor substitute for avoiding a serious injury or death. Since police are trained on when to use lethal force, if we think there is a problem with the officer's decision to fire in *Loud Music*, policy changes could prevent, or at least discourage, mistakes like this from taking place.

Relevant to *Loud Music* is the fact that the victim was not being violent or aggressive toward the officer. Though he appeared to be ignoring the officer's commands, this was only because he could not hear what the officer was saying. He is a completely innocent party, in other words, but the police officer, also an innocent party, does not know that. As the officer interpreted the situation, the man's failure to heed orders, his sagging jacket pocket, and the way he moved his hands were evidence of an imminent attack. Philosophical authors have long been concerned with the ethics of defensive force between innocent parties. In one commonly discussed case, residents of a town where a serial killer

55 Police officers making a split-second decision do not have the luxury of “the 20/20 vision of hindsight,” the ruling emphasized. *Graham v. Connor*, 490 U.S. 386, 397 (1989), 396–97. “Monday morning quarterbacking is not allowed,” as a Third Circuit ruling puts the point. *Lamont v. New Jersey*, 637 F.3d 177 (3rd Cir. 2011), 183.

56 Simons, “Self-Defense,” 53–55, 88–89. Though many US states require that individuals take a course on gun safety and the laws of defensive firearm use as part of the gun-permit process, 20 states allow permitless carry. Buchholz, “Which States Allow the Permitless Carry of Guns?” Civilian self-defenders who use forms of force other than firearms would not necessarily have any knowledge of the law and ethics of self-defense.

is known to be hiding are shown the killer's picture and told that he will immediately kill anyone upon sight. In the coincidence of all coincidences, the serial killer's identical twin happens to be driving through town and has car trouble. He knocks at a resident's door for help, she answers, and immediately attacks him in self-defense.⁵⁷ The case of *Resident* tests at least two moral questions that arise in thinking about mistaken self-defense between innocent parties. May the twin fight back against the resident in counter-defense, given that her belief that he poses a threat is arguably well-founded? Also, if a third-party observer with a sniper rifle (a police officer, say) grasps the resident's mistake but is too far away to shout a warning, should they shoot the resident to save the twin? Has the resident made herself liable to harm, in other words, even though her mistake is nonculpable?

Several prominent accounts view the resident as liable, though there is disagreement why—for McMahan, the resident is morally responsible for engaging in a foreseeably risk-imposing activity; for Quong, the resident erroneously treats the twin as if he lacks rights that people normally possess.⁵⁸ The resident's liability means that the twin is morally entitled to counter-defend against her use of defensive force, and a third-party observer may intervene and choose the twin's life over the resident's. Importantly, though the goal of analyzing a case like *Resident* is to work out the nature of liability to defensive harm (and the implications for persons who are only minimally responsible for the threats they pose in war), the possibility of counter-defense and third-party intervention also serve as built-in checks against fact-relative wrongful defensive harm.

However, these checks do not straightforwardly carry over into the policing context. If a civilian is able to counter-defend against a police officer making a fact-relative mistake, this is likely to be perceived by the officer as a threat, plain and simple, and things are not likely to go well for the civilian.⁵⁹ Moreover, consider a variant on *Loud Music* where the partner of the officer who shoots believes the victim to be unarmed. Even if the partner were extremely

57 McMahan, "The Basis of Moral Liability to Defensive Killing" and *Killing in War*; Quong, *The Morality of Defensive Force*; Ferzan, "Culpable Aggression"; Bazargan, "Killing Minimally Responsible Threats."

58 McMahan, "The Basis of Moral Liability to Defensive Killing," 402, and *Killing in War*, 176–78; Quong, *The Morality of Defensive Force*, 34–39. Bazargan argues that the resident may be permissibly killed by a third party in order to save the twin, but she is not liable to this fate since it is disproportionate to her level of moral responsibility. (The twin's life is to be preferred because it is a greater injustice to be killed when you are both non-liable and your threatener is more morally responsible than you for the situation.) Bazargan, "Killing Minimally Responsible Threats."

59 A civilian defending themselves against a police officer making a reasonable mistake is a case discussed by Draper, "Defense," 74.

confident that his colleague was making a mistake and could intervene in time, he is unlikely to act like the police officer who chooses the innocent twin over the innocently mistaken resident. It is hard to imagine a police officer seriously harming or killing a fellow officer to prevent an act of self-defense against a civilian misconstrued as posing a threat.

The lack of checks available to innocent civilians subject to mistaken defensive force by police officers puts the former in a vulnerable position. It is possible that this is nevertheless defensible. However, there are two conditions not currently met that would seem to be requisite: a *Justification Condition* and a *Valuing Civilian Lives Condition*. To meet the Justification Condition, the government would have to show that the trade-offs made in a policy that allows the police use of lethal force against non-aggressive civilians who pose uncertain threats is justifiable overall. To meet the Valuing Civilian Lives Condition, police cultural norms would have to mirror democratic norms, and at the very least recognize the lives of police officers and civilians as equally valuable. (It would also be permissible to prioritize the protection of civilian lives.)

Let us turn to the Justification Condition first. It is a basic democratic idea that policies that involve the government's exercise of power, particularly its coercive power, over citizens must be justified.⁶⁰ Forst puts citizens' "right to justification" at the center of political legitimacy.⁶¹ For Waldron, it is a matter of respect for individual agency that "all aspects of the social order should be either made acceptable or be capable of being made acceptable to every last individual."⁶² As Gaus writes, "Unlike private citizens, public officials are under a standing obligation to justify themselves. . . . Public officials must be able to provide publicly accessible reasons justifying what they do."⁶³ There are few instances of state coercion so weighty as a public officer taking a citizen's life, as "the right to life and physical security" is "the most basic claim of every human being."⁶⁴ For a policy that allows police officers to use lethal force against persons whose liability to defensive harm is unknown, the bar of justification is thus high.

60 There is a large public justification literature devoted to unpacking this idea. Must government policies be justifiable only to an idealized reasonable citizenry, or to actual embodied individuals who may not always be reasonable? Further, not all authors agree that specific policies must be justifiable, so long as the basic structure of society and constitutional essentials are. Admittedly, my point here hinges on the idea that individual policies must be justifiable.

61 Forst, *The Right to Justification*.

62 Waldron, "Theoretical Foundations of Liberalism," 128.

63 Gaus, *Justificatory Liberalism*, 251, 199.

64 Ashworth, "Self-Defence and the Right to Life," 282.

This high bar is not met at present. Even though some evidence suggests that the shooting victim in Loud Music may have been armed, there is only uncertain evidence—nothing like the much higher probability of an armed attack continuing as in Bank Robbery. It is moreover only from a police officer's perspective that it is possibly reasonable to construe this uncertain evidence as involving a threat. If a random pedestrian went to talk to the man in Loud Music and was met with non-responsiveness as he fumbled with his jacket pocket, it would be completely unreasonable for her to perceive an attack and immediately kill the man in self-defense. It is only because police authority will sometimes bring out suddenly violent behavior, and because the man's behavior fit some of the signs police officers are trained to pick up on, that the officer who fired shots interpreted the evidence as he did.

Our central question is not, however, whether it is evidence-relative permissible for police officers to use lethal force in a given uncertain threat scenario. It is about what police policy norms should be. The evidence relevant to setting justifiable policy norms is not only what a given officer sees in the moment, but also how this evidence maps onto aggregate data about similar cases.

Recall the abysmal state of official data collection on police shootings in the US, however. "We still live in a society in which the best data on police use of force come to us not from the government or from scholars, but from the *Washington Post*," as James Fyfe, one of the twentieth century's leading police use of force researchers, lamented in 2002.⁶⁵ US federal, state, and local governments have made no sincere attempt to provide evidence that a policy permitting the police use of lethal force in uncertain threat scenarios is justifiable. Granted, it is hard to collect objective data on "waistband shootings" and the like—again, when civilians do not have a lethal weapon, police officers will explain that they formed a reasonable belief that a civilian was armed and fired their gun because of the hand gesture the civilian made, but when civilians are in fact reaching for a gun, the shooting is likely to be framed as an ordinary threat posed by an armed attacker. Nevertheless, with body and dashboard cameras being increasingly used, it is possible for a slow-motion video analysis to pinpoint the exact moment a police officer decided to use lethal force, and determine what kind of threat the officer faced at that point in time.⁶⁶ From there, calculations would need to be made about the likelihood of suspicious hand gestures indicating an armed attack in different kinds of contexts.⁶⁷ How

65 Fyfe, "Too Many Missing Cases," 99.

66 For an example of this kind of analysis, see Forensic Architecture, "The Killing of Harith Augustus."

67 Though I discuss three contexts in this article, there are others I do not specifically consider, including "fleeing felon" cases, cases involving known owners of concealed handguns, and

often are officers mistakenly shooting and injuring or killing unarmed persons (or armed persons misconstrued as reaching for their gun)? How often are officers injured or killed by armed attackers? Is the risk of serious injury or death so great for police officers that a sufficiently high number of lives are saved by a lesser evil justification-based policy that permits shooting persons who pose uncertain threats? Police rhetoric often suggests that this is the reality, and such a claim seems credible enough in a case like Bank Robbery without there being a strong need for further justification. But in other kinds of cases, the US public is not obligated to take law enforcement officials at their word. If the data showed that a comparatively large number of non-liable civilians are seriously injured or killed to protect a comparatively small number of police officers in situations where non-aggressive civilians make suspicious hand gestures, it would not be justifiable to have a policy permitting the use of lethal force in such cases.

This takes us to the Valuing Civilian Lives Condition. Police use of force policies sometimes incorporate sanctity of life provisions, e.g., “It is the policy of the Philadelphia Police Department that officers hold the highest regard for the sanctity of human life, dignity, and liberty of all persons.”⁶⁸ At the same time, a common mentality about officer safety is encapsulated in law enforcement aphorisms like, “Better to be judged by twelve than carried out by six” and “shoot first.”⁶⁹ Seth Stoughton, a former police officer and use of force researcher, describes how in police training, officer safety “is so heavily emphasized that it takes on almost religious significance”:

Rookie officers are taught what is widely known as the “first rule of law enforcement”: An officer’s overriding goal every day is to go home at the end of their shift. But cops live in a hostile world. They learn that every encounter, every individual is a potential threat. They always have to be on their guard because, as cops often say, “complacency kills.”... Hesitation can be fatal. So officers are trained to shoot before a threat is fully realized, to not wait until the last minute because the last minute may be too late.⁷⁰

cases of individuals having mental health crises and/or attempts at “suicide by cop.” Interestingly, in suicide-by-cop cases, data from Los Angeles suggests that police officers are not at a great risk of harm. In 419 fatal and nonfatal cases examined by researchers, only one LAPD officer suffered an injury and, 98 percent of the time, LAPD officers were able to resolve incidents with no force or less lethal force. Jordan, Panza, and Dempsey, “Suicide by Cop.”

68 Philadelphia Police Department, “Directive 10.1.”

69 Waegel, “How Police Justify the Use of Deadly Force,” 147.

70 Stoughton, “How Police Training Contributes to Avoidable Deaths.”

Along similar lines, a *New York Magazine* journalist narrates the experience of Officer Richard Haste, a White New York City police officer who killed Ramarley Graham, an unarmed Black teenager, in 2012:

He remembered the video-game-like simulator he'd trained on at the Academy. Often the targets were guys reaching into their waistbands. When Haste was slow to shoot, a sergeant was always there to yell, "You're dead now!" "I had no more time left," Haste later told investigators. "I felt that if I waited one more second that this person was going to draw a firearm and shoot me."⁷¹

By their very nature, such training methods do not value police officers and civilians equally, but heavily weight the interests of the officer in police/civilian encounters. By teaching novice police officers that, once their adversary begins to reach for a gun, there is no time for an officer to do anything but fire before being fired upon, this all but says that police officers are not required to bear any risk of being harmed in this situation—better that a civilian who is not actually an adversary be wounded or killed than an officer's hesitation put him at a heightened risk. This amounts to a norm of police culture where the lives of civilians are valued less than the lives of police officers.⁷²

From a democratic standpoint, such a norm is dubious: the idea that "no one person should have his interests counted more than those of any other person" is a fundamental democratic value.⁷³ For the interests of state representatives to be given more weight than those of political subjects is particularly indefensible, given democracy's longstanding vigilance toward state encroachments on individual freedoms. The state's very claim to political authority derives from its status as a guarantor of basic democratic rights; of these, an individual's right to his or her continued existence is surely the most important and inviolable.⁷⁴ Democracy thus demands that police and civilian lives are valued equally. Novice police officers should not enter into the job believing that its dangers are mitigated by special self-defense protections.

Some have argued, however, not that police and civilian lives should be valued equally, but that the lives of civilians should be valued significantly

71 Walsh, "Can Ramarley Graham's Family Get Justice for His Death?"

72 Not every law enforcement organization actively promotes this outlook. See the Police Executive Research Forum (PERF) publication *Guiding Principles on Use of Force*. Granted, PERF's original recommendations received so much law enforcement pushback that significant revisions were made to be more aligned with officer safety-focused views. Ranalli, "Adding Perspective to the PERF Guiding Principles on Use of Force."

73 Brettschneider, *Democratic Rights*, 23.

74 Valentini, "Human Rights, Freedom, and Political Authority," 580–83.

more. This argument is closely linked to the traditional (albeit often idealized) British view of the police role. The “beloved” British bobby is depicted in the figure of PC Dixon, the fictional police officer at the center of the 1950 movie *The Blue Lamp*. Dixon is a heroic figure committed to always doing the right thing; he dies after being shot by a gun-brandishing criminal with whom he tries to reason instead of using force.⁷⁵ In this spirit, Gardner emphasizes the positive duty of police officers to protect members of the public, seeing the police–civilian relationship as analogous to the parent–child relationship.⁷⁶ Filicide evokes a very particular moral horror because the very person who is supposed to protect and care for a child has perverted this duty in murdering them; similarly, “killings by police officers are among the worst there can be.”⁷⁷ For Gardner, even when it is clearly necessary to kill a person who will otherwise kill their fellow citizens, this is a moral event whose significance should not be understated, a tragedy he likens to *Sophie’s Choice*.⁷⁸ Civilians’ interests should thus be given more weight in defensive force encounters, and heavily so.

Arguably, having a police culture that sees the protectorate role as possibly requiring self-sacrifice is more feasible when there is a low likelihood of police officers actually facing this outcome: in the period between 2012 and 2020, no on-duty British police officers were fatally shot.⁷⁹ But this observation does not refute Gardner. The protection of civilian lives is recognizable as a democratic public good. The contingencies of individual personality make some people more attracted to careers where they assume higher levels of personal risk for public-spirited reasons. Individuals are surely permitted to take on these risks. A democratic government would seem to also be permitted to fashion the police role such that it is incumbent upon police officers, based on how they are trained and police cultural norms, to be willing to sacrifice themselves to advance the project of democratic states as guarantors of rights.⁸⁰ On this logic, there is no issue with England recruiting individuals to serve in the tradition of bobbies like PC Dixon who would die before killing a civilian who threatens his life, so long as they are fully informed. And this would be permissible even in democratic contexts where gun violence is much more prevalent than in England.

75 Reiner, *The Politics of the Police*, 68.

76 Gardner, “Criminals in Uniform.” I thank Tarek Yusari Khaliliyeh for bringing Gardner’s essay to my attention.

77 Gardner, “Criminals in Uniform,” 106.

78 Gardner, “Criminals in Uniform,” 107–8, 113.

79 Police Roll of Honour Trust, “United Kingdom Annual Roll.”

80 See also Harmon, “When Is Police Violence Justified?” 1157.

However, I do not think that Gardner's argument shows that democratic governments are required to fashion the policing profession such that civilians' lives are favored heavily over police officers'. Beneath their roles, police officers are individuals who are the moral equals of the civilians they encounter. At most, the Valuing Civilian Lives Condition seems to show that the state is prohibited from treating the rights of police officers as significantly weightier than those of civilians. There is space for the state to construct the police role in a range of ways consonant with this prohibition—from giving the lives of civilians and police officers the same weight to heavily weighting the lives of civilians.

A final observation before moving on. I have discussed the Valuing Civilian Lives Condition mostly in terms of police culture, but the condition may also require that certain kinds of policy changes be implemented. It is an empirical matter as to what policy measures would be effective in reducing the number of non-labile civilians who are killed by police, but suppose that there are such measures—e.g., requiring that police officers retreat or use de-escalation tactics when appropriate, spending more resources on conflict resolution training, maintaining crisis intervention teams staffed by mental health professionals to respond to certain kinds of calls, and so on. It seems plausible to say that if police culture were to change but no further concrete measures were taken to reduce civilian deaths, the Valuing Civilian Lives Condition would not be satisfied.

By way of summary: if (1) data released to the US public showed that suspicious hand gestures by non-aggressive civilians are, in the aggregate, so overwhelmingly indicative of a deadly threat that a sufficiently high number of lives are saved by a policy that permits police officers to defensively use lethal force, and (2) US law enforcement demonstrated a strong commitment to safeguarding civilian lives, a policy permitting killing non-aggressive civilians in cases like Loud Music would be defensible. As things stand currently, neither condition is met.

3.3. *Civilians Behaving Aggressively*

Despite what has been said about the failure of US law enforcement to meet the Valuing Civilian Lives Condition, if police officers used lethal force every single time a detainee moved their hands, the number of people killed by US police each year in scenarios like Loud Music would be much higher than it currently is. Police officers are trained to develop good judgment about when suspicious hand gestures mean actual danger. Though non-aggression typically indicates a lower likelihood that a civilian poses a threat, aggressive noncompliance indicates the opposite. "Police training universally recognizes noncompliance as a danger signal," Patrick and Hall write.⁸¹

81 Patrick and Hall, *In Defense of Self and Others*, 105.

A noncompliant individual is not liable to being killed by the police because of being noncompliant.⁸² At the same time, one might think that if a civilian has made a threatening movement or gesture while acting aggressively, if there is one party who should bear the greater risk of harm, it should be the civilian. Let us examine the strongest argument for a policy permitting police officers to use lethal force by supposing that the Valuing Civilian Lives Condition is met. In *Resisting Arrest*, a non-self-favoring police officer is in the process of arresting a civilian for a crime he is suspected of. The civilian refuses to follow the officer's commands. He struggles and twists as the officer is trying to grab ahold of him and kicks him in the shins, asking if he wants to fight and saying that the officer has it coming.

Philosophical work on "provocateurs" is relevant to a case like this. Clearly, it is not permissible for the police officer to respond by harming the provocateur-civilian simply because he gives in to the provocation.⁸³ Ferzan argues, however, that if a respondent impermissibly inflicts harm due to being provoked, the provocateur who "started the fight" may not subsequently defend themselves and is not owed compensation for their injuries.⁸⁴ Like the 7-Eleven robber who threatens a clerk with an unloaded gun, the provocateur has forfeited certain rights through their culpable actions.

If this correct, it seems plausible that if a provocateur-civilian makes a gesture indicative of an uncertain threat, it is only fair that they are no longer entitled to be treated as an innocent party and the officer's moral equal. Through their culpable actions, they have created a situation where it is reasonable for the police officer to feel threatened, and it seems plausible for the latter to be permitted

82 I accept that a noncompliant individual might be subject to harm of some kind during a legal arrest or subject to proportionate force so that a police officer can protect themselves from nonlethal harm; my only claim here is about the noncompliant person's non-liability to being defensively *killed* due to being noncompliant. As Bank Robbery showed, however, an actively violent criminal suspect can be noncompliant. This section's focus is cases where a noncompliant civilian has not shown signs of being *harmfully violent*. In *Resisting Arrest*, below, I assume that it hurts to get kicked in the shins, but this does not cause injury to the officer. But what if the civilian punches the officer, giving them a black eye? What if the civilian is a "fleeing felon"? Some authors have argued that fleeing felons are liable to being defensively killed under certain circumstances (e.g., S. Miller, *Shooting to Kill*, 129–37). Alas, I set such questions aside in the present analysis.

83 Ferzan, "Provocateurs"; Hecht, "Provocateurs and Their Rights." For Ferzan, provocateurs are individuals who do not pose a fact-relative threat, but, for our purposes, it is more useful to construe *Resisting Arrest* as a case where the provocateur's underlying intents are unknown, which is closer to Hecht's understanding.

84 Ferzan, "Provocateurs," 599, 614–16. Hecht argues that it matters how much provocateurs contribute to the wrongful harm against themselves; they may counter-defend against harm that exceeds their contribution ("Provocateurs and Their Rights," 176–80).

to defend themselves against the risk of the provocateur-civilian's movement indicating a lethal threat. Moreover, threat probability-based considerations may come into play here: compared to a non-aggressive person, the provocateur-civilian would seem likelier to escalate to genuine physical violence.⁸⁵

This conclusion might seem to be supported by arguments made by Jorgensen, who has defended the idea that mistakes in self-defense are reasonable when they conform to an "assumptive signaling" norm.⁸⁶ According to this norm, self-defenders may act on signals of aggression that can easily and reasonably be avoided by their performer and are a matter of public knowledge. Consider a man who trails a woman's exact path in a parking garage. He may be lost in thought and not thinking about her perception of him as he heads to his car, but it is fair to say that he should pay closer attention in this situation. The cost of avoiding this behavior would be minimal and he can be expected to know that parking lots are places where women face a heightened sexual assault risk. He has thus transgressed the relevant assumptive signaling norm. While a woman who sprays him with Mace is reasonable even though she is mistaken, a White person in West Oakland who is getting money out of an ATM, sees a Black man approach, assumes he is about to rob her, and uses Mace acts unreasonably. Being Black and being an ATM customer are not things that a just society should require that he avoid, thus the features of the situation leading her to decide that he is a threat do not pass the test for being valid assumptive signals.

Following Jorgensen, it would be extremely unfair to require that individuals avoid listening to loud music lest they are unable to hear the commands of a police officer who happens to decide to interview them. But the situation of noncompliant civilians seems different. Individuals surely have a fair opportunity to avoid noncompliant behavior, and should also be able to predict that failing to heed police commands and fighting off an officer might lead them to think that they are looking to mount a lethal attack.

However, this conclusion is too hasty. Aggressively resisting police authority is not necessarily valid as an assumptive signal of danger and as a basis for a civilian's being subject to lethal harm if they make a suspicious movement. This is because there are a range of circumstances where aggressive conduct might be excused or even justified.

First, noncompliant behavior may be a response to being repeatedly subject to unjust police practices. In low-income, majority Black communities in the US, police officers frequently engage in forms of public order policing, including the practice of stop-and-frisk, that violate individuals' rights of free movement

85 Hecht, "Provocateurs and Their Rights," 174.

86 Bolinger, "The Moral Grounds of Reasonably Mistaken Self-Defense."

and privacy. At a systemic level, these policing practices can perpetuate race- and class-based second-class citizenship and deprive residents of the freedoms that democracies are supposed to safeguard.⁸⁷ In his discussion of inner-city ghettos, Shelby defends individuals' right to dissent under such conditions: "It is crucial, given the duty of justice and on grounds of self-respect, that the ghetto poor make manifest their principled dissatisfaction with the existing social order."⁸⁸ It may be permissible for an individual who objects to being recurrently detained as part of a public-order policing project to engage in non-compliant behavior—as well as it being overly burdensome to ask that this individual *not* engage in noncompliant behavior.⁸⁹

Second and relatedly, because of the stereotype of Black people as violent, in Black communities, there is a common view that one should unconditionally submit to police authority since it is unsafe to do otherwise. "The Talk" is a rite of passage where parents explain the importance of behaving obediently in interactions with police officers, even if an officer is rude or uses unwarranted force. In Paul Butler's words:

It is best not to assert too many rights. If you are not sure whether you actually are being detained, politely ask, "Officer, am I free to go?" If they say, "no," don't ask them what their reasonable suspicion is. Do not, at this point, ask to see an attorney (you don't have a right to one during a stop anyway). Do not ask if their body camera is on. Don't ask why they are touching your private parts or going into your pockets. Never tell cops, "You can't do this." It sets them off, and, under the law of the streets, yes, they can.⁹⁰

Facing a heightened risk of being harmed or killed by law enforcement officials because of one's race is deeply unjust.⁹¹ In the thick of an interaction with police,

87 Roberts, "Race, Vagueness, and the Social Meaning of Order-Maintenance Policing"; Natapoff, *Punishment without Crime*, ch. 6; Page and King, "Truth and Reparation for the U.S. Imprisonment and Policing Regime"; and King and Page, "Towards Transitional Justice?"

88 Shelby, *Dark Ghettos*, 227, emphasis removed. See also Valentini, "Human Rights, Freedom, and Political Authority," 587–88.

89 Bolinger, "The Moral Grounds of Reasonably Mistaken Self-Defense," 10.

90 Butler, *Chokehold*, 206.

91 Why are Black people killed by police at higher rates than White people in the US? O'Flaherty and Sethi distinguish between the fear hypothesis, where Black people are shot more often because police officers implicitly see Black civilians as more threatening, and the contact hypothesis, which says that because Black people have more police encounters, a higher rate of police shootings is statistically logical. See O'Flaherty and Sethi, *Shadows of Doubt*, ch. 8. Importantly, though some critics of the Black Lives Matter movement point to contact hypothesis research as evidence that there is not racism in US law enforcement,

an individual may be overcome by a sense of cynicism toward an institutional arrangement whose the logic *The Talk* encapsulates, and protest by doing the opposite of what *The Talk* counsels.

Third, noncompliance could be a way of protesting the enforcement of an unjust law. It might be unjust for the state to criminalize what it criminalizes, and there may be racial disparities both in what is criminalized and how criminal statutes are enforced.⁹² Noncompliant behavior might be a legitimate form of civil disobedience.

Fourth, a police officer might arrest a civilian illegally or wrongly suspect them of a crime they did not commit. Some argue that the court is the right place to fight an illegal or wrongful arrest, since noncompliant behavior will neither prove the civilian's innocence nor change the officer's mind about making the arrest.⁹³ However, a civilian may reasonably believe that the bail amount will be so high that he will be stuck in pretrial detention and will lose his job, neglect his dependents, etc.; and/or that he will be forced into an unfair plea bargain; and/or that the ordeal will get him into "the system" from which it will be impossible to extricate himself. Noncompliance might be a form of self-defense against the harm of an arrest to which he is not liable. Of course, given that the police officer acts with the power of the state, self-defense is likely to be futile. A success condition is often incorporated into theories of self-defense. However, as some have argued, the success condition need not be met if there are other important interests at stake, like the victim's registering their protest of the wrong. If a rape victim knows that she cannot overpower her rapist, she is still permitted to harm him to stand up for her moral worth.⁹⁴

In sum, we have considered a case, *Resisting Arrest*, where a provocateur-civilian acts in ways that make it reasonable for a police officer to perceive them as intending violence. If they make a movement indicative of an uncertain threat, they arguably forfeit their right to be treated as an innocent party and the officer's equal in determining a fair allocation of risk. This would make it justifiable to have a norm permitting the police officer to defend themselves against the risk of a lethal attack by using lethal force. However, considerations around over-policing, race, and the injustice of the US carceral system undermine the normative validity of noncompliant, aggressive behavior as a signal

the two hypotheses roughly map onto the distinction between individual (implicit) racism and structural racism. Structural racism may describe why Black communities are both disproportionately overpoliced and subject to higher rates of police shootings.

92 Bobo and Thompson, "Unfair by Design."

93 Chevigny, "The Right to Resist an Unlawful Arrest," 1136.

94 Statman, "On the Success Condition for Legitimate Self-Defense"; Frowe, *Defensive Killing*, 109–18.

of threateningness. Moreover, it is plausible to think that an individual who engages in aggressive behavior to protest an unjust criminal justice system does not actually intend to harm a police officer; they just want to register their protest and perhaps create an opportunity to flee. This suggestion is in principle empirically verifiable, and if it is right, it would mean that suspicious movements in certain kinds of arrest scenarios indicate a lower risk of a fact-relative threat to the officer than is typically assumed. Hence fairness-based and hypothesized threat probability-based considerations—paired with the non-satisfaction of the Justification and Valuing Civilian Lives Conditions—argue against a norm permitting the police use of lethal force against aggressive civilians who pose uncertain threats.

And if this is not convincing, consider that the details of Resisting Arrest were deliberately fashioned to make the civilian fit the description of a provocateur who is intentionally trying to pick a fight with the officer. This was done in an effort to lay out the best possible argument for aggressive civilians forfeiting their right to be treated as a police officer's equal. But not all real-world scenarios where individuals resist arrest, fail to comply with an officer's commands during a stop, etc., are like this—often individuals are simply reacting to the circumstances at hand in a heated way and do not aim to incite a police officer into using force.⁹⁵ Thus it should not be assumed that a typical case of noncompliance involves a civilian making incendiary threats. If a civilian is not a provocateur, it is much less clear that their right to be treated as the police officer's equal is forfeited to such a radical extent that it becomes permissible for the officer to use lethal force. This further undermines the argument for a police policy norm permitting lethal force to be used against aggressive civilians in uncertain threat scenarios.

4. CONCLUSION

As I have argued in this paper, focusing on the United States, the context of a known in-progress violent crime is a special case where there may be threat probability-based, fairness-based, and lesser evil reasons for a police officer to use lethal force against a suspect who poses an uncertain threat. Taken together, these reasons form a justification for a policy norm permitting the police use of lethal force in this context. This does not mean that police officers are automatically justified in shooting a violent criminal suspect who poses an uncertain threat, but it is justifiable to have a norm allowing officers to make

95 Chevigny, "The Right to Resist an Unlawful Arrest," 1134.

a judgment-based determination that there is an all-things-considered reason to use lethal force.

However, circumstances are markedly different in the context of a police interaction with a non-aggressive civilian, where a policy norm permitting the police use of lethal force is not justified. In order to be justified, state officials would have to show that the norm saves lives overall (satisfying a “Justification Condition”) and that police cultural norms meet democratic standards, seeing police and civilian lives as, at the very least, equally valuable (satisfying a “Valuing Civilian Lives Condition”). Neither condition is currently satisfied, as I have argued.

Finally, this paper has considered a scenario where a civilian is a “provocateur” who makes a suspicious hand gesture while behaving aggressively toward a police officer. That a civilian is engaged in culpable, avoidable behavior would seem to argue for a policy norm permitting the police use of lethal force in this kind of situation. However, unjust background conditions in the US may mean that certain forms of aggressive behavior are excused or justified, and may also decrease the likelihood that aggressive behavior indicates a fact-relative lethal threat to an officer. These considerations, paired with the non-fulfillment of the Justification and Valuing Civilian Lives Conditions, cast doubt upon the defensibility of a policy norm permitting police officers to use lethal force against aggressive persons who pose uncertain threats.

Of course, a range of other types of cases were not examined in the paper, e.g., mental health crises, attempts at “suicide by cop,” “fleeing felon” cases, and situations where non-aggressive civilians are armed with a concealed handgun. Future work is needed to assess norms permitting the police use of lethal force against persons who pose uncertain threats in such contexts.⁹⁶

University of Zurich
jennifer.page@uzh.ch

96 The author wishes to thank the participants of the Moral Philosophy Colloquium at the University of Zurich for reading and giving comments on multiple drafts, and also the participants of the Rethinking Risk Workshop at the University of Zurich, the Jurisprudence Discussion Group at the Oxford Faculty of Law, and the attendees of the October 2018 Public Ethics Talk at Leiden University. Special thanks go to Friedemann Bieber, Susanne Burri, Lisa Hecht, Tarek Yusari Khaliliyeh, Markus Kneer, Felix Koch, Daniel Messelken, Jan Alejandro Garcia Olier, Andrei Poama, Philipp Reichling, Stefan Riedener, and Peter Schaber for their contributions, as well as to two anonymous referees for feedback that significantly improved the paper.

REFERENCES

- Alexander, Larry. "A Unified Theory of Preemptive Self-Defense." *Notre Dame Law Review* 74, no. 5 (1999): 1475–1505.
- Alpert, Geoffrey P., and Roger G. Dunham. *Understanding Police Use of Force: Officers, Suspects, and Reciprocity*. Cambridge: Cambridge University Press, 2004.
- Ashworth, A. J. "Self-Defence and the Right to Life." *Cambridge Law Journal* 34, no. 2 (November 1975): 282–307.
- Aveni, Thomas J. "The MMRMA Deadly Force Project: A Critical Analysis of Police Shootings Under Ambiguous Circumstances." Police Policy Studies Council, February 9, 2008.
- Balko, Radley. "When Unarmed Men Reach for Their Waistbands." *Washington Post*, August 29, 2014. <https://www.washingtonpost.com/news/the-watch/wp/2014/08/29/when-unarmed-men-reach-for-their-waistbands/>.
- Bazargan, Saba. "Killing Minimally Responsible Threats." *Ethics* 125, no. 1 (October 2014): 114–36.
- Bittner, Egon. "The Functions of the Police in Modern Society." Chevy Chase, MD: Center for Studies of Crime and Delinquency, National Institute of Mental Health, 1970.
- Blair, J. Pete, Joycelyn Pollock, Don Montague, Terry Nichols, John Curnutt, and David Burns. "Reasonableness and Reaction Time." *Police Quarterly* 14, no. 4 (December 2011): 323–43.
- Bobb, Merrick J. "The Los Angeles County Sheriff's Department 30th Semiannual Report." Los Angeles: Police Assessment Resource Center, 2011.
- Bobo, Lawrence D., and Victor Thompson. "Unfair by Design: The War on Drugs, Race, and the Legitimacy of the Criminal Justice System." *Social Research* 73, no. 2 (Summer 2006): 445–72.
- Bolinger, Renée Jorgensen. "The Moral Grounds of Reasonably Mistaken Self-Defense." *Philosophy and Phenomenological Research* 103, no. 1 (July 2021): 140–56.
- . "Reasonable Mistakes and Regulative Norms: Racial Bias in Defensive Harm." *Journal of Political Philosophy* 25, no. 2 (June 2017): 196–217.
- Brettschneider, Corey. *Democratic Rights: The Substance of Self-Government*. Princeton, NJ: Princeton University Press, 2007.
- Buchholz, Katharina. "Which States Allow the Permitless Carry of Guns?" *Stastica*, June 18, 2021. <https://www.statista.com/chart/20047/gun-carry-laws-in-us-states>.
- Bureau of Justice Statistics. "Contacts between Police and the Public." Washington, DC: US Department of Justice, February 2001. <https://bjs.ojp.gov/>

- content/pub/pdf/cpp99.pdf.
- Burri, Susanne. "Morally Permissible Risk Imposition and Liability to Defensive Harm." *Law and Philosophy* 39, no. 4 (August 2020): 381–408.
- . "The Toss-Up between a Profiting, Innocent Threat and His Victim." *Journal of Political Philosophy* 23, no. 2 (June 2015): 146–65.
- Butler, Paul. *Chokehold: Policing Black Men*. New York: The New Press, 2017.
- Chevigny, Paul G. "The Right to Resist an Unlawful Arrest." *Yale Law Journal* 78, no. 7 (June 1969): 1128–50.
- Chung, Andrew. "us Top Court Won't Review Houston Police Shooting 'Waistband' Defense." Reuters, April 24, 2017. <https://www.reuters.com/article/us-usa-court-police-idUSKBN17Q1F5>.
- Cleveland Division of Police. "Use of Force." January 1, 2018. https://www.clevelandohio.gov/sites/default/files/forms_publications/01.10.2018General.pdf?id=12398.
- Delaney, Arthur, and Diane Jeanty. "Police Shootings of Unarmed Men Often Have Something in Common: The Waistband Defense." *Huffington Post*, December 10, 2014. https://www.huffpost.com/entry/police-shootings_n_6303846.
- Donnelly, Richard C. "Police Authority and Practices." *Annals of the American Academy of Political and Social Science* 339, no. 1 (January 1962): 90–110.
- Donner, Christopher M., and Nicole Popovich. "Hitting (or Missing) the Mark: An Examination of Police Shooting Accuracy in Officer-Involved Shooting Incidents." *Policing: An International Journal* 42, no. 3 (January 2018): 474–89.
- Dorn, Michael, and Chris Dorn. "Seven Signs a Weapon Is Being Concealed." Crisis Prevention Institute, 2006. <https://www.crisisprevention.com/CPI/media/Media/Resources/KnowledgeBase/seven-signs-a-weapon-is-being-concealed.pdf>.
- Draper, Kaila. "Defense." *Philosophical Studies* 145, no. 1 (July 2009): 69–88.
- Fabre, Cécile. "War, Policing, and Killing." In *The SAGE Handbook of Global Policing*, edited by Ben Bradford, Beatrice Jauregui, Ian Loader, and Jonny Steinberg, 261–78. London: SAGE Publications, 2016.
- Fachner, George, and Steven Carter. "Collaborative Reform Initiative: An Assessment of Deadly Force in the Philadelphia Police Department." Office of Community Oriented Policing Services, 2015. <https://www.phillypolice.com/assets/directives/cops-w0753-pub.pdf>.
- Economist*. "Simulators Teach Police and Their Critics When to Shoot." July 11, 2019. <https://www.economist.com/united-states/2019/07/11/simulators-teach-police-and-their-critics-when-to-shoot>.
- Ferzan, Kimberly Kessler. "The Bluff: The Power of Insincere Actions." *Legal*

- Theory* 23, no. 3 (September 2017): 168–202.
- . “Culpable Aggression: The Basis for Moral Liability to Defensive Killing.” *Ohio State Journal of Criminal Law* 9 (2012): 669–97.
- . “Justifying Self-Defense.” *Law and Philosophy* 24, no. 6 (November 2005): 711–49.
- . “Provocateurs.” *Criminal Law and Philosophy* 7, no. 2 (October 2013): 597–622.
- Fletcher, George P. “Punishment and Self-Defense.” *Law and Philosophy* 8, no. 2 (August 1989): 201–15.
- Forensic Architecture. “The Killing of Harith Augustus.” September 18, 2019. <https://forensic-architecture.org/investigation/the-killing-of-harith-augustus>.
- Forst, Rainer. *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York: Columbia University Press, 2012.
- Frowe, Helen. *Defensive Killing*. Oxford: Oxford University Press, 2014.
- Futterman, Craig B., Chaclyn Hunt, and Jamie Kalven. “Youth/Police Encounters on Chicago’s South Side: Acknowledging the Realities.” *University of Chicago Legal Forum* (2016): 125–211.
- Fyfe, James J. “Too Many Missing Cases: Holes in Our Knowledge about Police Use of Force.” *Justice Research and Policy* 4, nos. 1–2 (2002): 87–102.
- Gardner, John. “Criminals in Uniform.” In *The Constitution of the Criminal Law*, edited by R. A. Duff, Lindsay Farmer, S. E. Marshall, Massimo Renzo, and Victor Tadros, 97–118. Oxford: Oxford University Press, 2013.
- Gaus, Gerald F. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. Oxford: Oxford University Press, 1996.
- Gordon-Solmon, Kerah. “What Makes a Person Liable to Defensive Harm?” *Philosophy and Phenomenological Research* 97, no. 3 (November 2018): 543–67.
- Harmon, Rachel A. “When Is Police Violence Justified?” *Northwestern University Law Review* 102, no. 3 (2008): 1119–99.
- Hecht, Lisa. “Provocateurs and Their Rights to Self-Defence.” *Criminal Law and Philosophy* 13, no. 1 (March 2019): 165–85.
- Ho, Taiping. “Individual and Situational Determinants of the Use of Deadly Force: A Simulation.” *American Journal of Criminal Justice* 18, no. 1 (March 1994): 41–60.
- Irvin, Sherri. “Policing, Racialization, and Resistance: An Aesthetic Analysis.” Unpublished manuscript, 2021.
- Jacobellis, Nick. “How to Spot a Concealed Firearm.” *Police Magazine*, November 1, 2007. <https://www.policemag.com/339839/how-to-spot-a-concealed-firearm>.

- Jordan, Alejandra, Nancy R. Panza, and Charles Dempsey. "Suicide by Cop: A New Perspective on an Old Phenomenon." *Police Quarterly* 23, no. 1 (March 2020): 82–105.
- King, Desmond S., and Jennifer M. Page. "Towards Transitional Justice? Black Reparations and the End of Mass Incarceration." *Ethnic and Racial Studies* 41, no. 4 (2018): 739–58.
- Lazar, Seth. "In Dubious Battle: Uncertainty and the Ethics of Killing." *Philosophical Studies* 175, no. 4 (April 2018): 859–83.
- Lee, Cynthia. "Officer-Created Jeopardy: Broadening the Time Frame for Assessing a Police Officer's Use of Deadly Force." *George Washington Law Review* 89 (2021).
- Lowery, Wesley, Lindsey Bever, and Katie Mettler. "Police Have Killed Nearly 200 People Who Were in Moving Vehicles since 2015, including 15-Year-Old Jordan Edwards." *Washington Post*, May 3, 2017. <https://www.washingtonpost.com/news/post-nation/wp/2017/05/03/police-have-killed-nearly-200-people-who-were-in-moving-vehicles-since-2015-including-15-year-old-jordan-edwards/>.
- McMahan, Jeff. "The Basis of Moral Liability to Defensive Killing." *Philosophical Issues* 15, no. 1 (October 2005): 386–405.
- . *Killing in War*. Oxford: Oxford University Press, 2009.
- . "Who Is Morally Liable to Be Killed in War." *Analysis* 71, no. 3 (July 2011): 544–59.
- Miller, Seumas. *Shooting to Kill: The Ethics of Police and Military Use of Lethal Force*. Oxford: Oxford University Press, 2016.
- Miller, Tim. "Introduction." *FLETC Legal Aspects of Use of Force Podcast Series*. 2016. <https://www.fletc.gov/sites/default/files/2016%20UOF%20Podcast%20Introduction.pdf>.
- Muhammad, Khalil Gibran. *The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America*. Cambridge, MA: Harvard University Press, 2011.
- Najdowski, Cynthia J., Bette L. Bottoms, and Phillip Atiba Goff. "Stereotype Threat and Racial Differences in Citizens' Experiences of Police Encounters." *Law and Human Behavior* 39, no. 5 (October 2015): 463–77.
- Natapoff, Alexandra. *Punishment without Crime: How Our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal*. New York: Basic Books, 2018.
- O'Flaherty, Brendan, and Rajiv Sethi. *Shadows of Doubt: Stereotypes, Crime, and the Pursuit of Justice*. Cambridge, MA: Harvard University Press, 2019.
- Otsuka, Michael. "Killing the Innocent in Self-Defense." *Philosophy and Public Affairs* 23, no. 1 (January 1994): 74–94.

- Page, Jennifer M. "Reparations for Police Killings." *Perspectives on Politics* 17, no. 4 (December 2019): 958–72.
- Page, Jennifer M., and Desmond King. "Truth and Reparation for the U.S. Imprisonment and Policing Regime: A Transitional Justice Perspective." *Du Bois Review* 19, no. 2 (Fall 2022): 209–31.
- Parfit, Derek. *On What Matters*. Vol. 1. Oxford: Oxford University Press, 2011.
- Patrick, Urey W., and John C. Hall. *In Defense of Self and Others . . . : Issues, Facts, and Fallacies—The Realities of Law Enforcement's Use of Deadly Force*. 3rd ed. Durham, NC: Carolina Academic Press, 2017.
- Police Executive Research Forum. *Guiding Principles on Use of Force*. March 2016. <https://www.policeforum.org/assets/30%20guiding%20principles.pdf>.
- Police Roll of Honour Trust. "United Kingdom Annual Roll." February 7, 2020. <https://policememorial.org.uk>.
- Quong, Jonathan. *The Morality of Defensive Force*. Oxford: Oxford University Press, 2020.
- Ranalli, Michael. "Adding Perspective to the PERF Guiding Principles on Use of Force: What Police Administrators Should Consider." *New York Chief's Chronicle*, June 2016.
- Reiner, Robert. *The Politics of the Police*. 4th ed. Oxford: Oxford University Press, 2010.
- Ripstein, Arthur. *Equality, Responsibility, and the Law*. Cambridge: Cambridge University Press, 1999.
- Roberts, Dorothy E. "Race, Vagueness, and the Social Meaning of Order-Maintenance Policing." *Journal of Criminal Law and Criminology* 89, no. 3 (Spring 1999): 775–836.
- Shelby, Tommie. *Dark Ghettos: Injustice, Dissent, and Reform*. Cambridge: Belknap Press, 2016.
- Siegfried, Mike. "How to Watch the Hands." *Police Magazine*, January 28, 2010. <https://www.policemag.com/340257/how-to-watch-the-hands>.
- Simons, Kenneth W. "Self-Defense: Reasonable Beliefs or Reasonable Self-Control?" *New Criminal Law Review* 11, no. 1 (2008): 51–90.
- Statman, Daniel. "On the Success Condition for Legitimate Self-Defense." *Ethics* 118, no. 4 (July 2008): 659–86.
- Steele, Claude M. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. New York: W. W. Norton and Company, 2010.
- Stoughton, Seth. "How Police Training Contributes to Avoidable Deaths." *Atlantic*, December 12, 2014. <https://www.theatlantic.com/national/archive/2014/12/police-gun-shooting-training-ferguson/383681/>.
- Sweet, Dawn M., Christian A. Meissner, and Dominick J. Atkinson. "Assessing Law Enforcement Performance in Behavior-Based Threat Detection Tasks

- Involving a Concealed Weapon or Device." *Law and Human Behavior* 41, no. 5 (October 2017): 411–20.
- Tadros, Victor. *The Ends of Harm: The Moral Foundations of Criminal Law*. Oxford: Oxford University Press, 2011.
- Taylor, Paul L. "Dispatch Priming and the Police Decision to Use Deadly Force." *Police Quarterly* 23, no. 3 (September 2020): 311–32.
- . "'Engineering Resilience' into Split-Second Shoot/No Shoot Decisions: The Effect of Muzzle-Position." *Police Quarterly* 24, no. 2 (June 2021): 185–204.
- Tillman, George Jr., dir. *The Hate U Give*. Los Angeles: 20th Century Fox, 2018.
- Valentini, Laura. "Human Rights, Freedom, and Political Authority." *Political Theory* 40, no. 5 (October 2012): 573–601.
- Waegel, William B. "How Police Justify the Use of Deadly Force." *Social Problems* 32, no. 2 (December 1984): 144–55.
- Waldron, Jeremy. "Theoretical Foundations of Liberalism." *Philosophical Quarterly* 37, no. 147 (April 1987): 127–50.
- Walsh, James D. "Can Ramarley Graham's Family Get Justice for His Death?" *New York Magazine*, June 14, 2017. <http://nymag.com/daily/intelligencer/2017/06/ramarley-graham-nypd-shooting.html>.
- Washington Post*. "Fatal Force." <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/>.
- Zimring, Franklin E. *When Police Kill*. Cambridge, MA: Harvard University Press, 2017.

BETRAYED EXPECTATIONS

MISDIRECTED ANGER AND THE PRESERVATION OF IDEOLOGY

Barrett Emerick and Audrey Yap

AFTER the 2016 presidential election in the United States it was common to encounter think pieces and hot takes from folks excusing rural, poor, white Americans for having voted for Donald Trump.¹ Although his campaign was grounded in and employed racism and xenophobia, both overtly and covertly, apologists for those voters argued that their anger was legitimate and exculpatory; they had been economically exploited and politically marginalized. The anger of poor rural whites alienated from the concerns of urban elites was seen as an obvious reason why they would find someone like Trump appealing (though that does not answer the question of why the obvious bigotry of the Trump campaign was not a deal breaker for them). Our view in this paper is that this was an instance of a general phenomenon where a group's justified anger is redirected toward an inappropriate source. This will provide us with a way of understanding the causes of many cases of misplaced anger without excusing the harmful actions to which that anger often leads. Though our examples are primarily drawn from a North American cultural context, in which the US political landscape dominates, we believe that the phenomenon we describe is ubiquitous.

The particular kind of anger we unpack in this paper is anger that is partially justified but misdirected.² Fully justified anger is both grounded in an appropriate source and directed toward the appropriate system or agent of that system. Anger can be *partially* justified by being grounded in an appropriate source and directed toward an inappropriate target, or by being directed to an appropriate target and grounded in an inappropriate source. Our paper focuses on the former, in that we are considering anger that is grounded in or

1 Two examples: Kurtzleben, "Rural Voters Played a Big Part in Helping Trump Defeat Clinton"; and Leonard, "Why Rural America Voted for Trump."

2 For more on different senses of the appropriateness of emotion, see D'Arms and Jacobson, "The Moralistic Fallacy." For the aptness of emotion as a "fitting response to the world," see Srinivasan, "The Aptness of Anger."

is the result of unjust and oppressive systems. But it is misdirected, in that the target of the anger is not those oppressive systems or their agents. As such, we will call it justified-but-misdirected anger. The misdirection we will explore is born from the ideologies that sustain those oppressive systems. Following Sally Haslanger, we understand an ideology to be “the background cognitive and affective frame that gives actions and reactions meaning within a social system and contributes to its survival.”³ Ideologies are the social stories we are trained in and in which we train others, often without conscious awareness and in ways that are constrained (as they always are) by whatever conceptual resources are available in the relevant social imaginaries. They provide social scripts for how to act and what outcomes to expect from our actions.⁴ Sometimes those ideological scripts are accurate and just. Often they are neither and they distort our understanding of the world and misdirect what would otherwise be appropriate anger in ways that preserve the ideology itself. Exploring that phenomenon is the primary aim of this paper. In short, we agree that many poor, rural white Americans were right to be angry, but argue that their anger was misdirected away from the economic systems that exploit and marginalize them, and toward immigrants and people of color who are also just trying to survive under capitalism. Our analysis will consider how social location bears on what emotions someone is encouraged to feel and how they are able to interpret those emotions. Specifically, we will consider the ways that gender, race, and class affect those moral-emotional and epistemic phenomena.

In many cases, anger (both appropriately and inappropriately targeted) is born from a sense of *expectation and betrayal* that someone feels when, despite having done “everything right,” things did not turn out the way they had been told or trained to believe they would. Poor, rural whites were trained in the American Dream, which says that if you work hard you can get ahead. Furthermore, US culture tends to be broadly individualistic, with an emphasis on individual rights and freedoms, even when the pursuit of such freedoms is in tension with the greater good.⁵ We contend that when that meritocratic dream

3 Haslanger, *Resisting Reality*, 447.

4 Because scripts are prescriptive, playing the role you have been assigned is incentivized or rewarded, and deviating from the script is disincentivized or penalized. Over time, one might become conditioned to thoroughly be the person that one has been trained to be. For more on this, see Lindemann, *Damaged Identities, Narrative Repair and Holding and Letting Go*; Haslanger, *Resisting Reality*; and Hacking, *The Social Construction of What?*

5 As one example, Americans’ strong resistance to mask wearing and distancing restrictions in response to COVID-19 is often attributed to a mixture of individualism and national exceptionalism; see Andrew, “America’s Response to the Coronavirus Is the Most American Thing Ever.”

crashes against the shores of a capitalist reality, someone who has worked hard throughout their life and still lives in relative poverty is justified in feeling both betrayed and angry that their expectations are not met, even if the actions they take as a result of their anger are morally wrong.

Examples of anger born from betrayal and unmet expectations are legion. Consider three more examples:

1. Members of Generation Y who were told that if they went to college and earned a degree, they would be able to find good jobs and pay off the massive student loans they had taken out to pay for it.
2. The disillusion Black and brown folks might experience upon learning (often at a very young age) that police are a violent extension of a white supremacist state.
3. Men who are trained in masculinity and expect sex if they act like “gentlemen.”

Each of these cases is grounded in the experience of being trained to believe that the world is a certain way and that if you act appropriately good outcomes will follow, only to discover that that is often not true. Such training, whether implicit or explicit, contributes to the ways in which our expectations are built up in the first place. Our use of “expectation” is normative rather than descriptive. A descriptive expectation is simply a kind of probability judgement of what we think is likely to happen. The expectations we consider are normative, in that they also include the belief that what is likely to happen *ought* to happen. Several other important caveats need to be made before moving on.

First, because people occupy different social locations and so are trained to believe different things, they will experience different types of betrayal from different sets of failed expectations. Indeed, the betrayed expectations of poor rural whites are different from those in example 1, who are saddled with serious student loan debt, though they might all be angry about being economically oppressed.

Second, not everyone who occupies the same social location will experience the same moral-emotional response to betrayed expectations; that certain emotional responses generally accompany experiences that tend to track particular social locations does not mean that *everyone* who occupies that social location will feel the same way, nor does it mean that everyone occupying that social location must have the same expectations.

Third, those who occupy some social locations will be trained *not* to feel angry even when their expectations are betrayed. Others will not experience betrayal because they will not ever have had any expectations that things would be otherwise.

Finally, we want to be very clear that the fact that feelings of betrayal and anger might be justified does not yet tell us anything about what the morally appropriate ways to express or act on that anger involve. It is completely consistent with our analysis that a justified emotional response can lead to an action that is deeply morally wrong.

We will proceed as follows. In section 1, we briefly unpack the way that we understand anger and its relationship to expectations and betrayal in general. In section 2, we analyze in depth the particular forms of expectations and betrayal that are bound up with the social promise that if you act one way or another certain outcomes will follow. We will explore some of the ways that one's anger can be misdirected by oppressive ideologies in a way that maintains and perpetuates those ideologies.

1. EXPECTATION AND BETRAYAL

Much has been written analyzing anger. Our aim is not to reinvent the wheel but to build off the good work that has already been done and then apply it in a new way that helps to better make sense of the world and the ideologies and social structures that are appropriate targets of anger. Specifically, we follow P. F. Strawson, who understands moral anger (or resentment) as a reactive attitude that is appropriate or warranted in response to a moral wrong.⁶ Strawson argues that there is an important difference between becoming angry that an event has occurred and becoming angry with another person who I believe has wronged me. So, if a lightning bolt burns down my house I might become angry that my home has been destroyed. If an arsonist burns down my house I might become angry with them for having destroyed my home. We will only focus on *moral anger with* (rather than *nonmoral anger that*) for the remainder of this paper.⁷

The key difference is that in the arsonist case my anger expresses the dual judgment that *the arsonist is a person*—a moral agent—who is the appropriate target of praise or blame, and that *I am a person* who can be wronged by others. Both judgements are forms of respect. In recognizing the arsonist as an agent who could have done otherwise, I regard them as a person rather than as a thing or naturally occurring event like a lightning strike.⁸ In recognizing myself as deserving certain kinds of treatment (or not deserving others) I recognize my own self-worth and value. In this we follow many feminist theorists who have

6 Strawson, "Freedom and Resentment."

7 Assuming, of course, that I do not subscribe to a belief system that attributes agency to natural events.

8 Strawson, "Freedom and Resentment," 6.

argued for similar claims.⁹ For instance, Elizabeth Spelman argues that “[t]o be angry at [someone who has wronged me] is to make myself, at least on this occasion, his judge—to have, and to express, a standard against which I assess his conduct. If he is in other ways regarded as my superior, when I get angry at him I at least on that occasion am regarding him as no more and no less than my equal.”¹⁰ As such, Spelman also recognizes that anger can be an act of insubordination in these latter cases. When I get angry with someone, I act as though I have the right to judge their behavior. I also signal that I will not tolerate future wrongful treatment and either *expect or demand* that the person who wronged me change their ways.¹¹ Note that understanding anger to convey that expectation does not necessarily entail a threat or wish that some harm befall one’s wrongdoer. Martha Nussbaum argues that anger essentially involves a desire for retribution.¹² Though it sometimes might, we see no reason to think that such a desire is universal to all experiences of anger.¹³

When we attend to anger and resentment as personal reactive attitudes, we recognize their role in, as Margaret Urban Walker notes:

The ongoing definition and enforcement of the standards by which we live, our unequal authority to define and enforce them, and the collective task of keeping vital our senses and practices of responsibility. Unexpected, or, in the observer’s view, improper displays of resentment highlight our disputes and misunderstandings about our standards and about the nature and membership of communities.¹⁴

That last point is crucial and bears repeating: my anger conveys either an expectation or a demand that the one who wronged me act differently in the future and is fundamental to existing in the world as a profoundly social being. Walker argues that “we navigate the human world around us by forming and acting on normative expectations of others and of ourselves. . . . A normative expectation anticipates compliance more or less (and sometimes scarcely at all), but always

9 Lorde, “The Uses of Anger”; Frye, *The Politics of Reality*; Spelman, “Anger and Insubordination”; MacLachlan, “Unreasonable Resentments”; and Murphy, “Forgiveness and Resentment.”

10 Spelman, “Anger and Insubordination,” 266.

11 Spelman, “Anger and Insubordination,” 267. A similar point is made in Cherry, “The Errors and Limitations of Our ‘Anger-Evaluating’ Ways.”

12 Nussbaum, *Anger and Forgiveness*.

13 For a helpful critique of Nussbaum’s view, see Srinivasan, “Would Politics Be Better Off Without Anger?”

14 Walker, *Moral Repair*, 136–37.

embodies a *demand* for that form of behavior we think we've a right to."¹⁵ Our account is similar to Walker's but with an important difference in terminology. In our view, the difference between an expectation and a demand is that an expectation includes (but does not solely consist in) a probability judgment about how things are likely to unfold whereas a demand does not. Recall that, as we use the term, expectations are normative: an expectation is a demand that I believe will likely be met and ought to be met. When I say that I expect you to make good on your promise I convey that I am holding you to it—you ought to keep your promise—and that is the future that I believe will and ought to unfold. When I merely demand that you keep your promise, I convey nothing about what future I anticipate will follow.

We will return to this distinction later in explaining several types of reactions one might have in the face of wrongdoing, which are often related both to one's relative privilege and to the particular relationships and contexts in which one is operating. Since humans are social beings, our expectations are grounded in and born from social norms and practices.

Consider this case. One roommate says to another, "I'll do the cooking if you wash the dishes." When the second roommate eats the cooking but fails to wash up (without good reason) the first is likely to become angry. That anger response demonstrates several things about the first roommate:

1. They recognize that the second roommate is an agent who made (bad) choices about how to spend their time and energy.
2. They recognize themselves as someone who is affected negatively and undeservedly by their roommate's failure.
3. They operate on a script about how roommates behave toward one another, derived from a larger social imaginary. Such scripts give rise to the moral-emotional responses described above: the tendency to expect or demand certain kinds of treatment (or not), to become angry (or not), and toward whom or what that anger is directed. In this case, the roommate has an expectation that their arrangement will be honored and becomes justifiably angry (or at least frustrated) when it is not.¹⁶
4. Even if they do not "stay on script" the material circumstances they encounter might cause a certain moral-emotional response.

15 Walker, *Moral Repair*, 24.

16 Here we follow Hilde Lindemann in thinking that narratives help craft our understanding of ourselves and each other within social relationships (*Damaged Identities, Narrative Repair and Holding and Letting Go*). Much more could be said about the role of those narratives or scripts, but would take us beyond the scope of this paper.

If the first roommate has a long history with the second roommate failing to honor such agreements, they might express a demand but hold no expectation that their roommate will make good on their word. They might also fail to either expect or demand if they have been trained to do so, lack proper self-regard, or both. That list of factors is not exhaustive, but those are at least some of the relevant options for making sense of their moral-emotional response. The important takeaway here is that, as profoundly social beings, we not only learn how to act and become who we are in relation to others, but we also learn what to expect or demand from others in light of the social location we occupy. “There is a sense in which any individual’s guilt or anger, joy or triumph, presupposes the existence of a social group capable of feeling guilt, anger, joy, or triumph. This is not to say that group emotions historically precede or are logically prior to the emotions of individuals; it is to say that individual experience is simultaneously social experience.”¹⁷ All of that can be more or less morally appropriate, and more or less in line with what justice requires.

One important implication of that point is that our emotional responses often have epistemic value.¹⁸ Many feminist and anti-racist scholars have worked to reject a view of emotion that regards it as anathema to reason. Far from being a hindrance to thinking and perceiving clearly, our feelings can often help us to more accurately make sense of and understand features of our situation.¹⁹ This is true both for those who experience conventional emotional responses and those who experience what Alison Jaggar calls “outlaw emotions” (those emotions that are unconventional or run counter to dominant ideologies or social imaginaries).²⁰

Jaggar argues that outlaw emotions are most obviously epistemically valuable insofar as they motivate new investigations into the nature and causes of various phenomena. Which problems should be solved? Which issues deserve further analysis or study? Outlaw emotions provide political motivation and help both to select which problems to pursue and the methods by which they are investigated.²¹

Furthermore, outlaw emotions can themselves be useful in perceiving that the world and the story we tell about it in the social imaginary do not match up. Jaggar writes, “Only when we reflect on our initially puzzling irritability,

17 Jaggar, “Love and Knowledge,” 382.

18 For a helpful discussion of the epistemic value of contempt and other moral emotion, see Bell, “A Woman’s Scorn,” 85–88.

19 Jaggar, “Love and Knowledge,” 387.

20 Jaggar, “Love and Knowledge,” 387.

21 Jaggar, “Love and Knowledge,” 387.

revulsion, anger, or fear may we bring to consciousness our 'gut-level' awareness that we are in a situation of coercion, cruelty, injustice, or danger."²²

In short: our emotional responses, both conventional and outlaw, can tell us something about the world. The first roommate's anger might only point to the fact that the second roommate did not keep their promise. But a lack of anger on their part might be more informative—perhaps something about the state of their relationship and their history together, about their self-regard, or both. Millennials with huge student loan debt who are angry about being told to give up soy lattes or avocado toast might correctly link that anger to the oppressiveness of capitalism and the fact that the social promises they had been made have not been kept. It might indicate that they feel as though they have been duped and exploited to the benefit of the extremely wealthy. In both cases, what one feels or does not feel, and whether those feelings are conventional or outlaw, identify or reveal features of the social context in which the agent is operating.

These responses need not be revelatory. Sometimes we know exactly why we are angry, and there is nothing further that needs to be learned about our situation. Furthermore, emotional responses do not *always* reveal something accurate about the world; like our other means of understanding the world they can go awry. As Jaggar says, "Although our emotions are epistemologically indispensable, they are not epistemologically indisputable."²³ At least some of the time when people are angry, the social context in which they are operating encourages them to direct their anger to an inappropriate target.²⁴ We will consider several cases below in which this misdirection is part of the self-preservation mechanism of particular ideologies.

2. BETRAYAL IN ACTION

As Jaggar points out, the presence of anger implies a background of social norms in which that anger is situated. One could not be betrayed, after all, by

22 Jaggar, "Love and Knowledge," 387.

23 Jaggar, "Love and Knowledge," 389.

24 There are other reasons why anger or other emotional responses might be misdirected. Psychotherapists in a broadly Freudian tradition identify a phenomenon called displacement as a kind of defense mechanism in which strong emotional responses are directed away from their genuine targets toward more vulnerable substitutes. For example, a child being bullied at school might take out their anger and frustration on a younger sibling. See, for instance Clark, *Defense Mechanisms in the Counseling Process*. While it is certainly possible that displacement is responsible for some of the misdirection we identify in this paper, we still think it is relevant to talk about the social mechanisms that encourage and support these misdirections. Thanks to Lisa Tessman for this point.

unfaithfulness if there were no social norms about fidelity in the first place.²⁵ This section will connect anger to the particular social norms that shape the ways in which it is (or is not!) felt and the targets to which it is (or is not!) directed.

We can be and feel betrayed in a wide variety of ways, and such betrayals are connected to the trust we have in the various individuals, offices, and institutions to whom and to which we are vulnerable. Your trust could be betrayed by a friend who breaks a promise, by a doctor who violates rights to confidentiality with no good reason, by an academic institution that denies tenure despite having met the stated requirements, or by other drivers who fail to observe the same traffic laws.

Feeling betrayed is different from actually being betrayed. I might feel betrayed if I trust in a kind of cosmic justice that things work out for the best (or that the arc of the moral universe bends toward justice). I might feel betrayed by a celebrity who turns out to be a creep despite cultivating a nice guy image, but since I do not know the celebrity personally, to say that they have betrayed me would be inappropriate. That is not to suggest there is not real pain associated with such revelations, nor is it to discount the significance of feeling betrayed. It is instead to highlight that actually being betrayed relies fundamentally on a breakdown in a relationship or set of relationships that are at least minimally bidirectional or reciprocal. In the celebrity case and the cosmic justice case, the relationships in question are unidirectional or one sided.²⁶

Trust in our friends is generally built up through different means than trust in people in virtue of the social roles they play, or in institutions to which we belong, but that trust, once in place, implies that we hold those people or institutions responsible for living up to certain obligations. While interpersonal trust is certainly relevant to the misdirected senses of betrayal we consider here,

25 Jaggar, "Love and Knowledge," 382.

26 Relationships can be more or less bidirectional. Some public figures with a track record of taking strong political stands could arguably be said to have a responsibility to their fans or followers, which would make their feelings of betrayal appropriate. For instance, it might be the case that J. K. Rowling's very public transphobia does constitute a betrayal of her fans (Wallis, "J. K. Rowling Doubles Down on Transphobic Comments, Reveals She's an Abuse Survivor"). Since some of the positive press around her books has suggested that reading them decreases prejudice and increases empathy toward marginalized groups, that might give her a responsibility to her fans to refrain from attacking marginalized groups herself (McDade-Montez and Dore, "Supporting Diversity & Inclusion"). It might also be the case that in the age of social media, where public figures respond interactively with fans, their relationship is sufficiently bidirectional so as to again make feelings of betrayal appropriate. We will not make either argument here, but just note that, by our definitions, betrayal can only occur in relationships in which there is some minimal reciprocity.

we will spend more time discussing trust in institutions and in particular social roles. This is because our sense of the responsibilities that institutions (and those playing specific roles in those institutions) have to us are largely shaped by our background social beliefs about the way the world is, and the things within it to which we are entitled.

To complicate things further, we do not all share the same background beliefs about the way the world is, the things we can reliably expect from people who occupy certain positions, or the things that others in our social world owe us. We will argue that many instances of misdirected anger and betrayal are the result of distorted expectations stemming from ideologies that encourage us to think of the world in ways that reproduce injustice. This does not mean that all cases of anger and betrayal can be explained in this way, but at least some can, and these are important cases. And this illustrates an important way in which certain ideologies maintain themselves—namely, by directing people's anger or resentment away from the ideology, often toward a more vulnerable target.

One important factor that shapes people's expectations is their social location. Women who have been raised with a normalized sense of their own inequality, or of the entitlement of others to their body, may simply expect sexual harassment as an inevitable annoyance. So although they might not like or welcome the comments and treatment they receive, they might not see that behavior as violating an established social norm or see the harassers as having failed to live up to appropriate standards of behavior. As a result, they might not have a sense that they could hold the harasser responsible for their actions (a point to which we will return). Conversely, many men under patriarchy have also been raised with a certain expectation of what women owe them and how it is appropriate to behave toward women. What follows is that women who fail to live up to those expectations will then frequently become the targets of such men's anger for having failed to satisfy social norms that were imposed on them by a sexist ideology.²⁷

Since our interest here is in describing the ways in which betrayal and its associated feelings of anger might be misdirected, we will distinguish between four different kinds of anger. First, there are cases of unjustified anger.²⁸ In such cases, people might simply be mistaken about the facts of a situation; I might be

27 Manne, *Down Girl*.

28 Our terminology here differs from some who write about anger. Srinivasan, talks about it in terms of aptness and fittingness ("The Aptness of Anger"). We use the language of justification in order to make some finer-grained distinctions that the concept of aptness does not quite capture. More specifically, we aim to distinguish between the appropriateness of anger as a moral response, and the appropriateness of the way in which that anger is targeted. Apt anger is appropriate in both of those ways, but the anger we consider is only appropriate in one of them.

angry with my roommate, thinking they had borrowed one of my things without permission, when in fact I had merely mislaid the item in question. Alternately, I might become angry about an unforeseen rainstorm during a camping trip that resulted in a cold and wet night. I might direct that anger toward my camping buddy, getting mad at them when they had done nothing wrong. (In this case, my anger is inappropriate partly because I treat what should have been an instance of *anger that* as an instance of *anger with*.) In these two examples, the anger is unjustified because no one acted wrongly: my roommate has not taken liberties with my possessions, the world does not owe me a sunny day, and my camping buddy could not have made good on that obligation even if it did. Both of these cases involve wholly unjustified instances of anger—no one did anything wrong even though I experience some misfortune in both cases. We will not consider further cases of wholly unjustified anger in this paper.²⁹

The second category consists of anger that is both justified and appropriately targeted. These are cases in which we are right to be angry about a wrong that has been done, and are also angry at those who are primarily responsible for that wrong. For instance, if my roommate had in fact taken my things without permission or good reason, I would be justified in being angry with them for failing to respect my personal boundaries. Anger does not need to be directed at individual people and can instead be directed at institutions or groups more generally. The Black Lives Matter rallies after the murders of George Floyd and Breonna Taylor in mid-2020 were expressions of a great deal of justified and appropriately targeted anger about police brutality and racism. While people were certainly angry at the individual police officers who committed the murders, the very institution of policing in North America was the target of anger amid calls to defund or abolish the police. Greta Thunberg's address to the United Nations in September 2019 was also an expression of justified anger at world leaders' collective inaction on matters of climate change. In all of these cases, people are angry in the face of a genuine betrayal, and they are angry at those who are responsible for that betrayal. Note that both cases illustrate that one can be betrayed without having ever expected otherwise. People aware of the realities of police brutality do not generally have an *expectation* that police officers will behave ethically or respectfully toward Black people, but can still *demand* that they do so. Thus we can be angry with people for failing to live up to a social norm even when that failure is utterly predictable. Yet while this category of anger is relevant to our considerations here, it is also well-covered territory in the literature. Instead, we will turn next to categories of anger that have received somewhat less attention.

29 Note that if I accuse my roommate of stealing my item but my friend stole it instead, my anger would be justified but misplaced; there is a culprit but I have named the wrong person.

The third and fourth categories of anger we consider are cases in which anger would be justified—when there has been a genuine betrayal or failure on someone's behalf to live up to a social norm. However, we do not always get angry with those who have in fact betrayed us or violated our trust. The third category consists of cases in which we would be justified in feeling angry but do not. Sometimes these are cases in which our sense of what we are owed has been sufficiently eroded by oppressive social norms, such that we do not recognize them as moral violations. For instance, women living in a patriarchal society might lose a sense of entitlement to their own bodies or sense of safety (or might never have such a sense in the first place). In these situations, such women might not get angry about sexual harassment or misconduct that they experience, because it does not occur to them to hold the perpetrators to a different moral standard. One of the authors remembers being in graduate school with a faculty member who often made her (and other women students) feel uncomfortable. Yet in the climate at the time, she and several other women students simply took it for granted that they would just have to take it upon themselves to avoid that person. According to our framework, the author was not angry (though she would have been justified in being angry), because she was not placing the appropriate demands on the faculty member. Her anger, or lack thereof, was not a matter of what she expected him to do, but what kind of conduct she felt she could demand from him. This allows us to see that certain kinds of social contexts can distort the standards to which we hold other people, so that we fail to hold them responsible for their wrongdoing or, as we will soon see, hold the wrong people responsible for wrongdoing.

The fourth and final category will consider cases in which someone's anger is a response to a genuine wrong or injustice, but the targets of their anger are not the ones who have wronged them. These are the cases of misdirected anger that are the main focus of our paper. Before Elliot Rodger killed seven people (including himself) and injured several others, he released a video about his so-called day of retribution. In it, he expressed a great deal of anger over his solitude, rejection, and lack of romantic and sexual success. The stated targets of his anger, as he addressed those he was planning to punish, were the "spoiled, stuck-up, blonde" sorority girls living nearby. While it may be disputable whether Rodger had the right to be angry in the first place over the solitude he was experiencing, (Amia Srinivasan has an interesting discussion of these kinds of questions), one potential diagnosis of misogynist anger is as a reaction against norms of masculinity from someone who clearly did not live up to such norms.³⁰

30 Srinivasan, "Does Anyone Have the Right to Sex?"

Part of the insidious work of patriarchy, then, is to misdirect anger away from the oppressive system that holds people to particular and frequently unattainable gendered standards of appearance and behavior, and toward those who are subordinated within that system. Had Rodger been angry at a system that tells boys from a young age that they are supposed to be a certain kind of man, and that at least some of their success as such a man is measured by their ability to attract women who also look and act in gender-appropriate ways, his anger would have been both justified and appropriately directed. But instead, that same system encourages the belief that women owe men an assortment of feminine-coded goods, like care and physical affection, in virtue of their position within a social hierarchy.³¹ As a result, Rodger, and others who have followed in his footsteps, internalized a system of beliefs under which women as a group can be held collectively responsible for their lack of romantic success. Much as people who believe that law enforcement is there for their protection might be angry at the police's failure to prevent crimes, within patriarchal ideologies, men are encouraged to believe that women are there to provide them with care and can become angry with women when such care is lacking. This category of justified but misdirected anger is intended to chart a middle course between "himpathy," Kate Manne's term for our tendency to provide excessive sympathy to (relatively privileged) men who commit acts of violence, and a carceral logic in which the circumstances of a person's wrongdoing are irrelevant. We think it is possible and necessary to both hold people responsible for their wrongful actions and to understand why such actions might have seemed the correct course of action at the time.³²

Another case in which anger is justified but misdirected are the situations of "poor rural whites" with whom we began, who are angry about certain aspects of their economic situation. They might be facing economic anxiety and the prospect of both job and status loss in the face of industries relying less and less on manual labor. Jared Yates Sexton's autobiographical book describes the misdirected anger of men like those in his family at the economy's shift away from manufacturing and manual labor-intensive jobs like the ones on which they and their fathers before them had come to rely:

While [these economic shifts] were signs of progress, men reacted as if they themselves were threatened instead of the patriarchal order that imprisoned them. They doubled down on misogyny, discriminated

31 Manne, *Down Girl*.

32 Spelman is very helpful for thinking about whose suffering secures recognition and uptake—and ways that suffering can be co-opted by those in positions of power (*Fruits of Sorrow*).

against women in the workplace, blocked the upward mobility of anyone but themselves, opposed civil rights as corrective measures that would have improved the economy, and supported politicians who promised to oppose progress and swore to bring back the former economic order they had languished in their entire lives.³³

The situation that Sexton describes is exactly one of justified but misdirected anger. As we said at the outset, for anger to be *wholly* justified it would have to both be *grounded* appropriately (it would have arisen in response to some actual wrong or injustice) and also be *directed* appropriately (focused on the responsible person or system). In distinguishing the grounds of anger from the extent to which it is appropriately directed, we hope to expose ways in which systems such as white supremacy and patriarchy preserve themselves. We argue that it is part of the self-preservation mechanisms of such ideological structures that they misdirect the anger of those harmed by them toward others (often already oppressed groups). For example, in the case of white American manufacturing workers whose employment has become increasingly insecure, a belief that they are entitled to a job of that kind in that industry can easily lead them to blame others, particularly out-group members such as immigrants of color, for taking jobs away from them. Such workers have good reasons to be angry: the exploitative nature of capitalism, the inflexibility of the economy, the lack of retraining opportunities, or environmental devastation, to name a few. However, the ideology of a North American industrialist capitalism leads them to blame immigrants (often presumed to be here illegally), or other countries generally, for taking the jobs that should rightfully be theirs. In other words, their anger is justified but misdirected away from its appropriate causes.

We argue that such toxic ideologies contribute to this disconnect by producing a sense of entitlement in people in which they wrongfully hold others responsible for producing desirable outcomes. This is perhaps more obvious in the case of many incels, who might be justifiably angry about our society's beauty standards (for people of all genders) that are frequently heteronormative, racist, sizeist, and ableist. As an implicit endorsement of those standards, many incels simply accept that they are naturally less desirable than other, more conventionally attractive men. But instead of questioning the sources and oppressiveness of such body ideals, many incels instead criticize women for failing to see the merits of "nice guys" or, as Rodger called himself, a "Supreme Gentleman." That incels focus on women in general as the source of their problems, or at least view them as appropriate scapegoats for punishment, is not coincidental, but is rather a feature of patriarchal ideology. Manne, for example, argues that

33 Sexton, *The Man They Wanted Me to Be*, 20.

misogyny, and much misogynist violence, is a means for the preservation of a sexist social order, calling misogyny the “law enforcement branch” of patriarchy.³⁴ Much of that is premised on the idea that women owe men their care and affection. Without that sense of entitlement to such “feminine-coded goods,” it would not make sense to frame misogynist violence as punishment or retribution. This means that patriarchy not only tries to hold people to rigid gender norms that can ultimately cause them various kinds of existential harm, but it also encourages men who fail to live up to those masculine standards to blame women if they do not get the romantic or sexual attention they desire.

These kinds of scapegoating strategies have also been crucial to the campaigns of many recent North American conservative politicians, providing the public with racialized bogeymen in order to prompt support for their ultimately destructive policies. As Ian Haney López argues:

Conservative dog whistling made minorities, not concentrated wealth, the pressing enemy of the white middle class. It didn't seem to matter that the actual monetary transfers to nonwhites were trivial. . . . What mattered was the sense that blacks [sic] were getting more than they deserved, at the expense of white taxpayers. The middle class no longer saw itself in opposition to concentrated wealth, but instead it saw itself beset by grasping minorities.³⁵

The impact of racial scapegoating, such as the gendered and racialized images of welfare queens, predatory immigrant men, and the concept of anchor babies, served to channel middle- and working-class white anger away from capitalist policies that ultimately hurt them economically. What this has meant, practically speaking, is that even extremely wealthy politicians like Donald Trump who have arguably no lived experience of the everyday economic struggles of their voter base can portray themselves as friends of ordinary working white people. They can do so by simultaneously telling such people that they are entitled to their fair share of the “American Dream” in the form of a secure job paying a good wage with the promise of further upward mobility, and that their progress toward this dream has been stymied by an array of villains who all happen to have non-white faces.

This misdirected anger is in the service of what Haney López calls *strategic racism*, which cultivates and leverages animus against racialized groups.³⁶ Pledges to “Make America Great Again” often invoke a nostalgic image of

34 Manne, *Down Girl*, 78.

35 López, *Dog Whistle Politics*, 58.

36 López, *Dog Whistle Politics*, 80.

America, in which playing by the rules and working hard would ensure financial stability through to a secure retirement. Rarely do such politicians champion government policies that would do things like guarantee health care for all, or ensure a living wage that would provide actual security for those facing economic precarity, much less make good on such social promises. Instead, the homogeneity of the invoked middle-class utopia in the cultural imagination makes it obvious who can easily be blamed for its absence: the non-white people who are conspicuously absent from this idealized society. In other words, it is, and has long been, in the interest of the wealthy to direct attention away from their own accumulation of wealth; non-white folks, portrayed in a variety of ways (dangerous, lazy, conniving, etc.) are an easy target.

The ongoing success of strategic racism is in the way in which its architects profit from the justified but misdirected anger that it helps to cultivate. As those in positions of power continue to support economic policies that concentrate wealth in the hands of a tiny minority of people, they cultivate and encourage a worldview that allows a white voter base to lay the blame for their hardship elsewhere. In other words, many white people are trained to form an expectation of what they are due, and to lay the blame for their failure to receive their due at the feet of non-white folks, who are much more similarly positioned to them than they might think. Many such people are right, however, to be angry, because, having “done everything right”—having worked hard to take responsibility for their own lives—they failed to receive the rewards that they were assured would be forthcoming if they did so.³⁷

We should not, however, overstate our case. The racial animus that is tapped by strategic racism to misdirect appropriately grounded anger can also result in anger that is entirely unjustified. In our initial discussion of unjustified anger, the cases we considered were accidental—no conspiracy or ideology led me to think that my roommate had taken liberties with my things. Contrast these cases of unjustified and accidental anger with the anger that many felt with quarterback Colin Kaepernick, who protested police violence against people of color by “taking a knee” during the American national anthem at the start of football games. Many fans were angry with Kaepernick, ostensibly for a variety of things (making a football game “political,” failing to just do his job, being disrespectful or unpatriotic to the United States). Donald Trump himself called a person who would take a knee during the national anthem a “son of a bitch.” Alternatively, consider someone who is angry with activists who protested the murder of George Floyd by Minneapolis police officers. Stuck at

37 An episode of *The Daily* from *The New York Times* provides helpful first-person testimony to the experience of feeling betrayed despite having done everything right. Barbaro, “The Epidemic of Unemployment.”

an intersection while the crowd marches by, the driver stewes with resentment about being made late to an appointment. In contrast to the misplaced item or the rained-out camping trip, a fan's anger with Kaepernick or the driver's anger with protesters is unjustified but non-accidental. There is something moral on the scene—Kaepernick and the protesters are both committing an act of protest against profound racial injustice. Perhaps there is some kind of genuine inconvenience to the fan and the driver—the fan's enjoyment of the game is sullied by a political statement or by a reminder that racial injustice exists, while the driver's trip across town takes longer and perhaps they miss an important appointment as a result. But neither Kaepernick nor the protesters acted wrongly, despite the costs borne by the fan or the driver. In short, while these are not cases of justified anger, they do share many of the same racist sources as the cases of justified but misdirected anger we considered earlier.

In this paper, we have argued that when someone's trust is wrongfully violated it can give rise to justified anger, and that such anger can often be misdirected by oppressive ideologies in a way that works to preserve the ideology itself or social institutions they enable. In short, we have been working to recognize the ways that someone might be right to be angry but might be wrong about toward whom or against what their anger should be directed. Note that this is still a case of being *angry with* instead of *angry that*. An unexpected rainstorm is under nobody's control. Social institutions like capitalism or patriarchy are not under the control of any individual person but are certainly perpetuated and supported by individuals. In short: while it is inappropriate to be angry with a rainstorm, it is quite appropriate to be angry with capitalism or patriarchy. Though in some cases one's anger might not be directed against any particular agent of oppression, oppression itself is a social phenomenon and so an appropriate target of moral anger. It is our hope that identifying the appropriate source of that anger is useful in working to tell new and liberatory social stories in a world shot through with injustice.³⁸

St. Mary's College of Maryland
bmemerick@smcm.edu

University of Victoria
ayap@uvic.ca

38 Many thanks to Chris Goto-Jones, Sara Protasi, Katie Stockdale, and Lisa Tessman for their invaluable feedback on an earlier draft of this paper. This paper also benefited from a conversation with Jane Dryden and her class at Mount Allison University.

REFERENCES

- Andrew, Scottie. "America's Response to the Coronavirus Is the Most American Thing Ever." CNN, May 19, 2020. <https://www.cnn.com/2020/05/19/us/american-individualism-coronavirus-trnd/index.html>.
- Barbaro, Michael, "The Epidemic of Unemployment." *The Daily*. June 24, 2020. <https://www.nytimes.com/2020/06/24/podcasts/the-daily/unemployment-coronavirus-pandemic.html>.
- Bell, Macalester. "A Woman's Scorn: Toward a Feminist Defense of Contempt as a Moral Emotion." *Hypatia* 20, no. 4 (November 2005): 80–93.
- Cherry, Myisha. "The Errors and Limitations of Our 'Anger-Evaluating' Ways." In *The Moral Psychology of Anger*, edited by Myisha Cherry and Owen Flanagan, 49–65. Lanham, MD: Rowman and Littlefield, 2018.
- Clark, Arthur J. *Defense Mechanisms in the Counseling Process*. Thousand Oaks, CA: SAGE Publications, 1998.
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (July 2000): 65–90.
- Frye, Marilyn. *The Politics of Reality: Essays in Feminist Theory*. New York: Crossing Press, 1983.
- Hacking, Ian. *The Social Construction of What?* Cambridge, MA: Harvard University Press, 2000.
- Haslanger, Sally. *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press, 2012.
- Jaggar, Alison M. "Love and Knowledge: Emotion in Feminist Epistemology." In *Just Methods: An Interdisciplinary Feminist Reader*, 378–91. New York: Taylor and Francis, 2014.
- Kurtzleben, Danielle. "Rural Voters Played a Big Part in Helping Trump Defeat Clinton." NPR, November 14, 2016. <https://www.npr.org/2016/11/14/501737150/rural-voters-played-a-big-part-in-helping-trump-defeat-clinton>.
- Leonard, Robert. "Why Rural America Voted for Trump." *New York Times*, January 5, 2017. <https://www.nytimes.com/2017/01/05/opinion/why-rural-america-voted-for-trump.html>.
- Lindemann, Hilde. *Damaged Identities, Narrative Repair*. Ithaca, NY: Cornell University Press, 2001.
- . *Holding and Letting Go: The Social Practice of Personal Identities*. New York: Oxford University Press, 2014.
- López, Ian Haney. *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class*. New York: Oxford University Press, 2014.

- Lorde, Audre. "The Uses of Anger." *Women's Studies Quarterly* 9, no. 3 (Fall 1981): 7–10.
- MacLachlan, Alice. "Unreasonable Resentments." *Journal of Social Philosophy* 41, no. 4 (Winter 2010): 422–41.
- Manne, Kate. *Down Girl: The Logic of Misogyny*. New York: Oxford University Press, 2018.
- McDade-Montez, Elizabeth, and Rebecca A. Dore. "Supporting Diversity and Inclusion: Harry Potter's Greatest Spell." *The ETR Blog*, November 13, 2018. <https://www.etr.org/blog/supporting-diversity-inclusion-harry-potters-greatest-spell>.
- Nussbaum, Martha. *Anger and Forgiveness: Resentment, Generosity, Justice*. New York: Oxford University Press, 2016.
- Sexton, Jared Yates. *The Man They Wanted Me to Be: Toxic Masculinity and a Crisis of Our Own Making*. Berkeley: Counterpoint, 2019.
- Spelman, Elizabeth V. "Anger and Insubordination." In *Women, Knowledge, and Reality: Explorations in Feminist Philosophy*, edited by Ann Garry and Marilyn Pearsall, 263–73. New York: Routledge, 1989.
- . *Fruits of Sorrow*. Boston: Beacon Press, 1998.
- Srinivasan, Amia. "The Aptness of Anger." *Journal of Political Philosophy* 26, no. 2 (June 2018): 123–44.
- . "Does Anyone Have the Right to Sex?" *London Review of Books* 40, no. 6 (March 2018): 5–10.
- . "Would Politics Be Better Off Without Anger?" *The Nation*, December 19–26, 2016.
- Strawson, P. F. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–28. New York: Routledge, 1974.
- Walker, Margaret Urban. *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. New York: Cambridge University Press, 2006.
- Wallis, Adam. "J.K. Rowling Doubles Down on Transphobic Comments, Reveals She's an Abuse Survivor." *Global News*, June 11, 2020.

THE RIGHT TO EMIGRATE

EXIT AND EQUALITY IN A WORLD OF STATES

Daniel Sharp

IT IS WIDELY ACCEPTED that there is a right to emigrate: that one has the liberty to leave one's state if one wishes, and one's state may not permissibly prevent one from exiting its territory.¹ But what justifies this right? And what does this justification entail for the duties of states? Here, there is less consensus. This paper explores these issues. It offers a pluralist defense of the right to emigrate, on which this right has three primary grounds: (1) protecting basic rights, (2) realizing core freedoms, and (3) promoting social equality. To make this argument, I first show that existing justifications of the right to emigrate are incomplete. Second, I offer a novel egalitarian defense of the right to emigrate. Third, I demonstrate that this egalitarian position has important consequences for how states may regulate immigration.

I begin, in section 1, by clarifying what the right to emigrate is and what a defense of it must do. In section 2, I consider the limitations of standard defenses of the right to emigrate. This demonstrates that existing theories need to be supplemented by an alternative account. In section 3, I offer such an account, based on relational equality. Relational egalitarians allege that we have a claim against standing in relations of social inferiority with others and that this offers a potent complaint against unequal relations of political rule. The usual answer to this complaint is to *transform* the unequal relations that exist within states into equal ones. I propose that a robust right to emigrate constitutes a complementary response, which mitigates social inequality by making it more escapable. After considering objections in section 4, I argue in section 5 that the egalitarian case for the right to emigrate yields a corresponding case for a right to immigrate of a distinctive sort. In section 6, I conclude by drawing out the implications of this argument.

1 Despite its presence in the UN Declaration of Human Rights, respect for this right is not universal, even on a minimal understanding of what the right to emigrate entails.

1. WHAT IS THE RIGHT TO EMIGRATE?

The right to emigrate is, as used herein, a moral claim right to physically leave one's state's territory. This right is distinct from other, related exit rights. It is distinct from the right to leave a social group, as emigration need not entail breaking social ties.² Similarly, it is distinct from the right to renounce one's citizenship or shed one's political obligations. Leaving one's state does not usually require renouncing one's membership nor does it necessarily imply a dissolution of one's obligations.³ The right to emigrate is, rather, a right to leave one's state's *territory*.⁴ Moreover, the right to emigrate is a moral claim right, not a mere liberty. As such, it entails duties on other agents. Chiefly, it entails a negative duty on one's state not to interfere with one's departure.⁵

The clearest way to justify a right to emigrate is to show that it is necessary to protect some important interest or value. A successful defense along these lines must (1) identify some important interest or value, (2) explain how the right to emigrate helps protect that interest or helps realize that value, and (3) show that that interest or value justifies imposing duties on others.⁶ A theory that does these things fulfills the *primary task* of justifying the right to emigrate. However, a *full* account of the right to emigrate should do more than this. It should explain the scope of the right and whether it is a primary or remedial right.⁷ It should also explain what duties the right to emigrate imposes on one's

2 Okin, "Mistresses of Their Own Destiny"; Kukathas, "Are There Any Cultural Rights?"

3 Stilz, "Is There an Unqualified Right to Leave?"

4 Within the category of rights of territorial exit, one can distinguish a right to emigrate, which is long term or permanent, from a right to leave one's state temporarily for purposes of international tourism, visiting family, conducting business, and the like. The justification of the former right may be distinct from the latter, as the underlying interests at stake in these two forms of mobility may be different. Plausibly, the latter, insofar as it is a right, is grounded in the value of freedom of movement (see section 2.2). I confine my discussion to the right to emigrate.

5 The right to emigrate also comes with other Hohfeldian incidents, such as immunity against being interfered with in certain ways; it also, I argue later, correlates with positive duties on the part of other states. I stress that I am not offering an analysis of the *legal* right to emigrate in international law, nor does my argument directly imply that the right to emigrate is a human right.

6 Note that my formulation has a broader justificatory basis than interest theories, narrowly understood. For a defense of interest theories, see Raz, *The Morality of Freedom*. Compare Oberman, "Immigration as a Human Right."

7 On remedial rights, see Buchanan, "Theories of Secession," 34–36.

state.⁸ Finally, it should explain whether the right to emigrate entails a right to immigrate.⁹ I discuss these issues in sections 4–6.¹⁰

Section 2 considers the standard defenses of the right to emigrate. I argue that they are on their own inadequate. However, one assumption is worth making explicit first. This is that any full justification of the right to emigrate will be pluralistic: as there are several important interests at stake in emigration, the right to emigrate plausibly has multiple grounds. This implies that partial theories, which explain why people have such a right in a certain range of cases, are not thereby defective; they are merely incomplete. When I argue that certain accounts are only partial, I am claiming that these accounts must be *supplemented*, not rejected.

2. STANDARD DEFENSES OF THE RIGHT TO EMIGRATE

This section examines standard justifications of the right to emigrate. I argue that two standard accounts identify important interests that partially ground the right to emigrate. However, I show that these accounts are incomplete and need to be supplemented by some additional ground. I then argue that existing accounts of what these additional grounds might be are inadequate (in sections 2.3 and 2.4). This makes space for my egalitarian argument (in section 3).

2.1. Basic Rights

A first justification of the right to emigrate appeals to its role in securing people's basic rights.¹¹ On this view, the right to emigrate is “essentially protective.” Its function is to provide a “safety valve” for those whose states violate or fail to protect their basic rights.¹² No one should be forced to bear the cost of their basic rights going unprotected. A state that denies a person whose basic rights are in peril the opportunity to leave would be compelling her to bear this cost. This seems especially wrong when the state that denies exit also culpably fails to protect the person's rights.

8 For the standard view, see Whelan, “Citizenship and the Right to Leave,” 658.

9 For discussion of this issue, see Lenard, “Exit and the Duty to Admit”; Wellman and Cole, *Debating the Ethics of Immigration*; Cole, *Philosophies of Exclusion*.

10 I largely set aside a further question: When may the state restrict emigration? This has been the focus of the debate about the right to emigrate. However, this issue can only be adequately approached once one understands what justifies the right to emigrate in the first place (see section 4.2). I acknowledge, however, that the right to emigrate is not absolute.

11 Lenard, “Exit and the Duty to Admit”; Mancilla, *The Right of Necessity*, 113.

12 Lenard, “Exit and the Duty to Admit,” 5.

This line of reasoning suggests that when one's basic rights are not protected, one may leave and seek protection elsewhere. This argument adequately explains why those facing persecution, state breakdown, generalized violence, severe poverty, or other special hardships have an unassailable claim to emigrate. It explains the special claims of refugees and other necessitous migrants. Protecting basic rights is one ground of the right to emigrate.

However, it is not the only ground. The basic rights view only justifies a right to leave for those whose rights are in peril. The basic rights of Norwegians are not in jeopardy. So, if the basic rights view exhausts the justification of the right to emigrate, Norwegians lack a right to emigrate. But it is widely believed that it would be deeply wrong for Norway to prohibit its citizens from emigrating. The basic rights view cannot explain this: where individuals already live in a state that protects their basic rights, they lack, on this view, a claim to exit. The basic rights view thus "traps citizens of liberal democratic states in their home states."¹³ It therefore forms, at most, part of the full justification for the right to emigrate.

One might seek to broaden the core idea that the function of exit is protective to include other interests worth protecting, such as one's interest in adequate options.¹⁴ Such a broadening move has some plausibility. But the prospects for such an extension are limited: it is unclear how such an account can ground the claims to emigrate of the citizens of decently affluent, moderately sized liberal democracies. Basic rights thus cannot alone form the basis of a full theory of the right to emigrate, even if they form a crucial part of any such theory.

2.2. *Freedom of Movement*

The basic rights view must therefore be supplemented by some other normative consideration that better explains the right's scope. According to a second prominent theory, people have a range of general positive interests in freedom of movement: in moving, traveling, and residing where they desire. These include interests in pursuing one's attachments and associations, executing one's projects and plans, and having various life possibilities open: the argument can be developed in different ways based on which positive interests one thinks matter most.¹⁵ Since, it is argued, these interests ground basic liberties in the domestic context, they can, it is alleged, also ground a general right to

13 Lenard, "Exit and the Duty to Admit," 13. Lenard thinks that those whose basic rights are protected may also have positive interests "in living an autonomous life," which yield a more general claim to move. I consider this view momentarily.

14 Miller, "Is There a Human Right to Immigrate?"

15 Oberman defends this position ("Emigration in a Time of Cholera") and appeals to our interest in being free to access the full range of life options when they make important personal decisions; see also Oberman, "Immigration as a Human Right." Stilz bases her

immigrate or move freely across borders. But immigrating requires emigrating. So, the right to freedom of movement entails a right to emigrate.

I believe that our freedom interests play an important role in the justification of the right to emigrate. Nevertheless, I do not believe the value of freedom of movement is the sole ground of the right to emigrate. First, while the freedoms connected to emigration are important, they do not, on their own, ground a particularly strong right to emigrate. One's general interest in liberty may generate a substantial presumption in favor of a right to emigrate.¹⁶ However, the general presumption can, as elsewhere, plausibly be overridden in many cases where countervailing interests are at stake. Moreover, the significance of the positive freedoms protected by the right to emigrate varies widely. Weighty freedoms are not always at stake in emigration, given the diversity of migrants' circumstances.¹⁷ So, insofar as the case for the right to emigrate rests solely on its importance for freedom, it is doubtful that this always grounds a strong right to emigrate.

Second, the standard criticisms of the right to immigrate can also be applied to this defense of the right to emigrate.¹⁸ Consider Miller's response to Oberman's defense of the right to immigrate. Miller argues that positive interests in autonomy cannot ground a right to immigrate: we have no claim to the *optimal* menu of options from which to fashion our lives; we only have a claim to an *adequate* range of options. This claim, if true, undermines Oberman's defense of the right to immigrate.¹⁹ But it also undermines any corresponding defense of the right to emigrate. If Miller's adequate range view is correct, then the freedom of movement defense of the right to emigrate fails to ground such a right when one's state is sufficiently large and well-resourced. My aim is not to defend Miller's criticism here.²⁰ My point is rather that *if* Miller's critique undermines the right to immigrate, it *also* undermines the right to emigrate.

Most importantly, the freedom of movement defense cannot account for the *distinctiveness* of claims to emigrate. According to this defense, the wrong of

argument on an interest in developing our Rawlsian moral powers ("Unqualified Right"). For a classic statement of the open-borders view, see Carens, *The Ethics of Immigration*.

16 For criticisms of this position, see Blake, *Justice, Migration, and Mercy*; Song, *Immigration and Democracy*; Stilz, *Territorial Sovereignty*.

17 Compare Stilz, *Territorial Sovereignty*, 205–6.

18 Ypi argues for the need to theorize emigration and immigration together ("Justice in Migration").

19 Perhaps the importance of autonomy diminishes once one has adequate options. Still, being denied further options nevertheless diminishes one's autonomy. One's autonomy interests are not exhausted by claims to an adequate range of options, even if only these are relevant to justifying human rights.

20 For a response, see Oberman, "Immigration as a Human Right."

being denied freedom of movement is a function of, and proportional to, the value of the opportunities for mobility that are foreclosed when one is denied exit. There is nothing distinctive here about one's claim to *leave* one's state. The wrong in being denied exit is just one instance of the general wrong of being cut off from various options. Yet, the wrong of being denied exit from one's state is, intuitively, not reducible to being deprived of the set of options to which one is thereby denied access.

Those barred from exiting New Zealand, e.g., are barred from many more opportunities than those barred from leaving the United States. The latter is much bigger than the former. But there is still a sense in which one core wrong in the two cases seems intuitively equivalent.²¹ Or suppose the United States were justly annexed by Canada to form Greater Canada. This would *massively* expand mobility for Canadians: they could now move freely within the current territory of the US. But suppose, simultaneously, that the Greater Canadian government prohibited emigration. Per the freedom of movement defense, Canada's new exit restriction would be less bad than a similar restriction would have been before annexation. People's *real* net opportunities for free movement would have been *expanded*. Canadians can now go anywhere in the US; before they could not. Yet, Greater Canada's emigration ban seems *in one respect* no less problematic than Canada's emigration ban would have been pre-annexation. The freedom of movement view cannot explain this. It cannot explain the *distinctiveness* of claims to emigrate.²²

One might resist this conclusion. Perhaps being denied exit usually forecloses *more* options for mobility than being denied entry. While this is *usually* true, one can easily imagine cases where it is false. Suppose the world consisted of a few tiny states, the Microlands, and one massive one, Macroland. Macroland restricts emigration and immigration. In being denied emigration, Macrolanders' options are restricted only minimally; in being denied immigration, Microlanders' options are restricted very significantly. Nevertheless, there

21 This is compatible with recognizing that the size and nature of the place one inhabits matter for, e.g., reasons of autonomy. My point is that autonomy is not the only value at stake in emigration. It is also true that people might have backward-looking autonomy interests in leaving a particular place. However, these interests are of the very same kind one has in moving elsewhere, and so would not capture something distinctive about the right to emigrate as compared to the right to immigrate.

22 A related problem holds for Blake's associative defense of exit (in Brock and Blake, *Debating Brain Drain*). On this view, the right to exit is grounded in a claim to pursue associative relationships. However, immigration restrictions impact people's associative interests just as emigration restrictions do.

seems to be something especially wrong with Macrolanders being confined to their state beyond the options to which they are thereby denied access.²³

One might reply that my argument here rests on an unfounded asymmetry.²⁴ Some deny that there is a salient moral distinction between *one's state denying one exit* and being denied exit *because no other state will let one enter*. But my argument does not depend on this asymmetry. It depends on the claim that one has a special interest in *being able to leave one's state's territory*. To accept this, one need not take a stance on whether it is worse to be denied exit by one's state's emigration restrictions than it is to be denied exit by other states' immigration restrictions. Thus, although our positive interests in freedom of movement, such as our interests in autonomy, form an important part of the justification of the right to emigrate, they are not the whole story.

While basic rights and freedom of movement interests each have a role to play in grounding the right to emigrate, neither alone constitutes the whole story. Nor do they together fully justify the right to emigrate: a hybrid account that recognized both grounds still could not capture the right to emigrate's distinctiveness where people's basic rights are not at stake. To capture the distinctiveness and generality of claims to emigrate, one needs a more general, *backward-looking* theory concerned with the distinctive relationship between the prospective emigrant and the state or society they seek to emigrate from.

2.3. Voluntarism

The most venerable backward-looking account is voluntarist. Lockeans claim that political authority arises only when people freely bind themselves to the state. As this theory is often developed, the possibility of emigration plays a crucial role. The key idea is that one's continued residence in a state's territory can constitute a form of tacit consent. But it only constitutes valid tacit consent if one has a right to emigrate, for only if one has that right is one's consent freely given. So, the right to emigrate is necessary to secure the state's authority over its citizens.²⁵

However, the voluntarist defense is, ultimately, unpromising. Even if the right to emigrate is a necessary condition for residence to constitute valid tacit consent, it is patently not a sufficient condition. For consent to be valid, it must be given knowingly and voluntarily in a clear choice situation on the basis of adequate information.²⁶ Tacit consent through residence does not meet these

23 The wrong here is a *pro tanto* one. There may also be good reasons why a single state cannot accept *all* immigrants who seek to enter (see section 6).

24 See Wellman and Cole, *Debating Immigration*, 193–205.

25 Locke, *Two Treatises of Government*; Simmons, *The Lockean Theory of Rights*; Whelan, "The Right to Leave," 647–50.

26 Simmons, *Moral Principles and Political Obligations*, 79–83.

requirements. There is not widespread awareness that residence constitutes tacit consent. Moreover, for consent to be voluntary, the consentor must have a genuine alternative to giving her consent, the costs of which are not “extremely detrimental.”²⁷ But the costs associated with emigration are often very high.²⁸ Even if states were to defray the costs of exit, there are inherent costs to emigration and people have a moral right to stay. So, even under more favorable institutional arrangements, consent via residence would likely not be valid. At least, the conditions required for it to be so are extremely demanding.²⁹ It is therefore implausible that continued residence generates political obligations through tacit consent. But if consent via residence would not suffice even under favorable conditions to generate political authority, voluntarism is not what justifies the right to emigrate.

The voluntarist might reply that a right to emigrate “seems to go *at least part way* toward making citizenship and its obligations entirely voluntaristic.”³⁰ But it is unclear what the value is of standing in a *quasi-voluntary* relation to one’s state.³¹ Insufficiently voluntary consent is still a form of invalid consent, and invalid consent cannot do the moral work that valid consent does.

More importantly, there are plausibly other grounds for political authority. But if consent is not necessary, then securing tacit consent through residence is not as important as the voluntarist alleges. The voluntarist defense depends on the claim that *only* through consent can authority be secured. This likely commits one to libertarian anarchism.³² This is too steep a price to pay to justify the right to emigrate, as this sort of libertarianism severely constrains the state’s pursuit of justice.³³ While voluntarism has its attractions, I show later that these can be captured without appeal to consent.

2.4. Freedom of Association

Christopher Heath Wellman defends an alternative backward-looking view.³⁴ He thinks individuals have strong rights of freedom of association. An essential

27 Simmons, *Moral Principles and Political Obligations*, 81.

28 Hume, “Of the Original Contract.”

29 Compare Otsuka, *Libertarianism without Inequality*, ch. 5.

30 Whelan, “The Right to Leave,” 638, emphasis added.

31 Compare Simmons, *Justification and Legitimacy*, 146.

32 This sort of view is defended in Simmons, *Moral Principles and Political Obligations*; and Kukathas, *The Liberal Archipelago*.

33 These points state *my* reservations about voluntarism. While these reservations are not decisive, I suspect many will share them.

34 Wellman, “Freedom of Movement and the Rights to Enter and Exit.” Blake defends a similar view in *Debating Brain Drain*, 198–201.

part of freedom of association is the freedom *not* to associate. But citizens' relation to their state is an associative relationship. Citizens thus have a right to stop associating with their state (and, one might add, their compatriots). Just as a husband forcing his wife to remain in a marriage would be wrong because it would violate her right to freedom of association, it would be wrong for one's state to force one to remain in a political relationship with it.

Wellman argues that this view justifies an asymmetry between exit and entry. Since states too have claims to freedom of association, freedom of association also justifies the state's right to exclude.³⁵ Wellman's critics question whether states have rights of freedom of association and whether self-determination justifies harming immigrants.³⁶ But one might endorse Wellman's defense of the right to emigrate without endorsing his defense of the right to exclude. The above criticisms do not apply to Wellman's defense of the former.

However, a different criticism of Wellman's defense of the right to exclude does. Sarah Fine argues that Wellman's defense of the right to exclude fails because Wellman lacks a theory of territorial rights.³⁷ Freedom of association is about *membership* in a political community, rather than about a general permission to *interact* with those who one chooses. The mere presence of immigrants on one's state's territory does not constitute a form of association with them.³⁸ So, Wellman's theory justifies, if anything, control over membership, not territory. A yoga group that meets in Central Park may reject new members, but it cannot bar them from making use of the park. Analogously, Fine concludes that freedom of association can only justify denying membership, not prohibiting entry, absent some further story about the connection between territory and membership.

The problem afflicts Wellman's defense of the right to emigrate. Freedom of association may explain why citizens may *renounce their membership* in their state. But it does not explain why citizens have the right to *leave the state's territory*. To tweak Fine's analogy, freedom of association may explain why one may leave the prison's book club; it does not explain why one may exit *the prison*. To do that, one must explain the link between territorial exit and membership, which Wellman does not do. So, for all Wellman says, it should be fine for the state to ban emigration as long as it ensures that everyone can, like Thoreau, isolate themselves in the wilderness and renounce their membership. This, after all, is a great way to avoid associating with anyone.

35 Wellman, "Freedom of Movement," 81–86.

36 Van der Vossen, "Immigration and Self-Determination"; Fine, "Freedom of Association Is Not the Answer."

37 Fine, "Freedom of Association Is Not the Answer," 353–55.

38 Fine, "Freedom of Association Is Not the Answer," 343.

Wellman might reply that one necessarily associates with the state as long as one remains within its territory. But this conflates two senses of association: association as membership and association as interaction.³⁹ There is no general permission to interact only with those one chooses to interact with. We have many obligations—reparative duties, duties of rescue, etc.—the discharge of which requires interacting with others in limited ways and that are not defeated by a general right not to associate.

Second, Wellman takes freedom of association to operate in a wide range of contexts.⁴⁰ On his view, the freedom not to associate is important because it is a constitutive part of one's autonomy. Yet, it is implausible that one has a completely general interest in association grounded in autonomy that is at once weighty enough to be rights grounding and general enough to apply across all associative contexts. Not all associations have an equally significant effect on one's autonomy. One's association with one's family is quite different from one's association in one's labor union. It seems implausible that a simple autonomy interest grounds a universal right to freedom of association that holds for all associations. Rather, what grounds various rights to freedom of association varies depending on the nature of the association and the way that association affects one's interests.⁴¹ That we have a right to freedom of association in some context, then, must be the *conclusion* of a more complex argument, which points to the *particular* interests at stake in that context.

Thus, Wellman's argument does not explain what about our association with the state makes it an important site of associative freedom. Indeed, at first glance, the state is a bad candidate for such associative freedom. One's claim to freedom of association in intimate relationships is usually stronger than one's freedom of association in non-intimate contexts. But one's association with the state is impersonal. Indeed, it seems like the sort of case in which one's freedom of association may be permissibly constrained, as in prohibitions against discrimination by private businesses. Without a richer account of the nature of one's associative relationship with the state and the particular interests implicated therein, the freedom of association defense of the right to emigrate fails, for it does not explain why associative principles properly apply to this relationship.

Wellman might reply that one's relationship to the state is indeed distinctive. Surely this is correct. But the real work in this argument for the right to emigrate must be done by an account of the important individual interests in

39 Brownlee, "Freedom of Association," 269.

40 Wellman acknowledges that the right to exclude has different weight in different associative contexts. But the differences, in his view, are merely in degree.

41 Brownlee, "Freedom of Association."

relating freely to the state *rather than* by a generic autonomy interest. Wellman thus treats the general right to freedom of association vis-à-vis one's state as what explains the right to emigrate, but such a right to freedom of association is *what needs to be explained*.

Finally, one's rights not to associate are plausibly circumscribed by one's (enforceable) duties toward others.⁴² One cannot generally shirk one's duties just because discharging them requires associating with others. We often have duties the discharging of which requires some kind of association. Taking care of one's child requires associating with her. But freedom of association cannot defeat one's parental obligations if no one else is willing to step in. Similarly, we have political obligations or other duties toward our compatriots and fulfillment of these obligations likely sometimes requires our continued association with the state and our compatriots.

If one's duties usually circumscribe one's associative rights, then Wellman's own theory of political obligation in conjunction with his defense of the right to exclude undermines his defense of the right to emigrate. For Wellman admits that Samaritan duties can defeat one's right to freedom of association and believes that our Samaritan duties can only be adequately discharged by our compliance with the laws of a legitimate state.⁴³ But suppose no other state is willing to accept one, as, per Wellman, is their right. Then, one could only discharge one's Samaritan duties by discharging them in one's state. In such cases, one's state would be justified in prohibiting one from leaving. This saps the force of the right to emigrate.⁴⁴

One might avoid this by embracing an extreme version of philosophical anarchism. This version denies not only that one has a general, content-independent obligation to obey the law, but also that one has any positive obligations to one's compatriots. Alternatively, one might claim that one has a right to escape one's political obligations at will. Neither position seems promising. One's obligations to one's compatriots are important duties and one cannot typically shed one's weighty obligations so easily. It also seems plausible that one sometimes has positive obligations to one's compatriots.⁴⁵ So, I do not find Wellman's view convincing.

42 Brownlee, "Freedom of Association," 275.

43 Wellman, "Samaritanism and the Duty to Obey the Law."

44 The problem of reconciling political obligation with the right to emigrate is general. Plausibly, some distributive obligations survive emigration. It is also likely that states may sometimes impose certain conditions on emigration, such as exit taxes on the super-rich (Stilz "Unqualified Right"). I discuss a similar issue briefly in Lovett and Sharp, "What Immigrants Owe."

45 Stilz, "Unqualified Right," 66–73.

I have argued that existing defenses of the right to emigrate face important limitations.⁴⁶ The basic rights theory and the freedom of movement theory both partially ground the right to emigrate. However, the former's scope is limited and the latter cannot explain what is distinctive about claims to emigrate. These accounts need to be supplemented by some additional backward-looking ground. However, prominent theories of this kind are unpersuasive. We need an alternative theory that supplements the freedom of movement view and the basic rights view and identifies some weighty interest in exiting one's state's territory.

3. A RELATIONAL EGALITARIAN DEFENSE OF THE RIGHT TO EMIGRATE

I will now offer an egalitarian defense of the right to emigrate. I first explain the conception of relational equality on which my argument relies and why it yields a complaint against political rule, and note some limitations of the standard response to this complaint (section 3.1). I then offer two arguments for the right to emigrate based on this conception (sections 3.2 and 3.3).

3.1. *The Problem of Social Hierarchy and the Transformative Solution*

My argument begins from the claim that there is something distinctively wrong with standing in hierarchical social relations with others. Consider caste societies, slave societies, and patriarchal societies. The social relationships found in such societies are deeply problematic. This is not only because of their deleterious consequences for the oppressed; nor is it only because they fail to distribute important goods fairly. Rather, such social structures are, relational egalitarians claim, also objectionable in themselves because they treat some as inferiors and others as superiors. Those confined to positions of inferiority in a social hierarchy may legitimately complain that they do not relate to their compatriots as equals; they have a complaint against "inferiority" or "subjection."

These ideas have been defended by Niko Kolodny, Elizabeth Anderson, and Daniel Viehoff among others.⁴⁷ I will not defend them further here. Instead, I

46 Rather than making a direct case for the right to emigrate, one might focus on the things that states must do to stop emigration. This requires coercion: building walls, staffing them with armed guards, threatening prospective emigrants, etc. That these actions are coercive and harmful is sufficient reason not to control emigration *in these ways*. Compare Mendoza, "Enforcement Matters." However, if this were the only problem with emigration restrictions, there would be no problem with the state building an impassable but unmanned wall to keep its citizens in. I, therefore, doubt that the wrong of restricting emigration is reducible to the harms of enforcement.

47 Kolodny, "Rule over None II" and "Political Rule and Its Discontents"; Viehoff, "Power and Equality"; Anderson, "The Problem of Equality from a Political Economy Perspective." I

will argue that *if* one accepts this anti-hierarchy conception of relational equality, one must also endorse the right to emigrate.

To make this argument, I must first say more about this strand of egalitarianism. A crucial question for these theorists is what makes a relationship objectionably hierarchical. A compelling answer has been offered by Kolodny. He suggests that social hierarchies are primarily constituted by three factors. Inequalities in power are the most obvious factor. It is the master's power over the slave that makes him the slave's superior. A second factor is inequality in *de facto* authority. The fact that the lord can issue commands and directives that must be obeyed by the peasant marks him as the peasant's social superior. A final factor is inequality of "consideration." The fact that Brahmans are held in high regard and treated with flattery and deference, while Dalits are viewed with contempt and expected to defer, marks the former as socially superior and the latter as socially inferior.⁴⁸ Disparities in power, authority, and consideration, Kolodny contends, amount to relations of objectionable social hierarchy when the disparities are particularly pronounced, are difficult to avoid, and permeate the whole of society. For our purposes, the first two factors, power and authority, are most relevant.⁴⁹

Crucially, not all disparities in power, authority, and consideration create social hierarchies of an objectionable sort. For one, such disparities are only objectionable when they occur within the context of a genuine social relationship.⁵⁰ There are no "relations of inferiority" between me and Emperor Diocletian: though he had more power than I ever will, he wields no power *over* me. Defenders of relational equality therefore pick out the social relations within particular political communities as a matter of special concern. Kolodny also claims that disparities in power and authority constitute an objectionable form of social hierarchy only where they are *unmitigated* by various *tempering factors*. Disparities in power sometimes occur in one-off encounters; they are not part of

defend such a view in Sharp, "Relational Equality and Immigration."

48 Kolodny's category of "consideration" is somewhat elusive (see "Standing and the Sources of Liberalism"). I therefore do not rely on it here.

49 I focus on Kolodny's account because it is the most developed. However, I suspect, with one exception, that nothing hinges on which account of social hierarchy one adopts. Any plausible account will likely have similar implications since such a view will frame political inequality as a threat to equal social standing in many circumstances. The exception concerns the relation between power and justification. If one thinks that appropriately justified power raises no egalitarian complaint, then this may weaken the scope of my argument. For a defense of this view, see Viehoff, "Power and Equality." I also discuss this in Sharp, "Relational Equality and Immigration," 675–78.

50 Scanlon, *Why Does Inequality Matter?* 9; Wellman, *Debating Immigration*, 62. For discussion, see Sharp, "Relational Equality and Immigration," 648.

an entrenched relationship. Other disparities, such as a teacher's power over her students, are limited in time and place or in their content. Power disparities may also be tempered by being subject to higher-order control or easily avoidable. These kinds of factors *limit* one's complaint against unequal power and authority; so, one's complaint against inequality is strongest where they are absent.

Social egalitarians contend that the complaint against social hierarchy is applicable to political relationships. Political relationships involve significant disparities in power and authority and are not usually subject to the above tempering factors. Inequalities in power and authority are constitutive of political rule. Those who hold office (or who have greater influence over who ultimately does) possess an outsized share of political power and authority. Political rule pervasively structures people's lives. There is no "higher" authority to which one may appeal; the state's power is vast and largely unbounded. Unless relations of political rule are suitably tempered or constrained, political rule thus threatens subjection.⁵¹

This provides a diagnosis of what can be problematic about one's relation to one's state and one's compatriots. State power is the power of public officials. Unless suitably constrained, it places one in a position of social inferiority vis-à-vis those who hold office. Similarly, influence over the levers of state is often unequal. When some citizens have disproportionate political influence, the state's power and authority is, to a greater degree, *their* power and authority.⁵² We, therefore, have a *pro tanto* complaint against unequal political power and authority.

The standard response to this complaint is transformative: we should *equalize* power and authority.⁵³ It is usually argued we should do so by adopting a robust form of democracy, since democracy equalizes political power. I endorse this argument for democracy. It is persuasive as far as it goes. But it suffers from some limitations.

Actual democracies are representative democracies. Yet, representative democracies inherently involve unequal power. Representatives have far more power than those they represent. They have far more influence over state policy than ordinary voters. Even the most ideal representative democracy will therefore be characterized by a great deal of political inequality.⁵⁴ What goes for

51 Kolodny, "Rule over None II" and "Political Rule and Its Discontents."

52 Kolodny, "Rule over None II."

53 Kolodny, "Rule over None II"; Viehoff, "Power and Equality."

54 Compare Lovett, "Must Egalitarians Condemn Representative Democracy?" Lovett believes that it is a mistake to think egalitarians must show representative rule is not problematic at all; the best egalitarians can show is that representative democracy minimizes the problem.

representatives goes double for the state's administrative authorities, who necessarily have considerable discretion over policy. Democracy thus does not eliminate political inequality. At best, democracy *tempers* that inequality by giving ordinary citizens a share of collective control over their representatives. It realizes equality better than other alternative political institutions. But, even in their most ideal practicable form, democracies retain significant inequalities.⁵⁵

Second, actual democracies do not live up to egalitarian ideals. Social equality requires that each person possess an equal opportunity to influence political decisions in line with their judgment. But actual democracies permit vast inequalities in informal opportunities for political influence. They distribute various goods, such as wealth, unequally and those with more of these goods typically possess much greater opportunities to influence political outcomes than others. Since inequalities in informal influence undermine social equality, no existing democracy adequately realizes equality.⁵⁶

One might reply that we should simply make existing democracies more genuinely democratic. I agree that democracy's egalitarian shortcomings must be remedied. We should distribute wealth and power more equally. Still, such transformations are not easy to achieve. Entrenched elites typically refuse to give up their power, and intractable collective action problems must be overcome to force them to do so. In most societies, there is no easy path to fully realizing political equality. Moreover, pursuing real transformative change requires much from ordinary citizens, and those who are subjected to social inequality are not always required to bear these burdens.⁵⁷ It is therefore worth thinking about *complementary* responses to the problem of social hierarchy. I will now defend the right to emigrate as one such response.

3.2. The Remedial Argument

How might those relegated to positions of social inferiority respond to the inequality of their existing political relations? One alternative to transforming relationships is to exit or break them. Call this the *avoidance solution*. Suppose one is married to a domineering partner. One response to the inequality within one's marriage is to transform it into a marriage between equals; another response is to get a divorce. In exiting one's marriage, one escapes the power

55 Perhaps full direct democracy would resolve this problem. But such a form of government is likely impracticable and not best supported by the balance of reasons.

56 Kolodny, "Rule over None II," 332–35.

57 Distributing political power fully equally may come at the costs of inefficiency or instability, or hinder the quality of decisions. The role these costs should play in debates about what form democracy should take is analogous to the role the potential negative costs of emigration should play in arguments for the right to emigrate.

and authority of one's partner: they can no longer control one's life in the same way. One can make a similar case for the right to emigrate. If one stands in unequal social relations within one's state (and suitably equalizing those relationships is infeasible), then one has a right to extricate oneself from those social relations. One can do so by leaving the state's jurisdiction. Social equality thus grounds a right to emigrate.

Let us sketch the argument more carefully. The efficacy of the "avoidance solution" stems from the underlying relational nature of the complaint. Complaints against social inequality, I argued earlier, only arise within the context of real social relationships. It is not just that someone has greater power, e.g., that forms the basis of one's complaint; it is that *they* have power over *you*. This means that one's complaint against it can be resolved *either* by equalizing the social relationships one is in *or* by dissolving those relationships.

Yet, actual states are sites of relational inequality. This inequality is entrenched enough that a transformative solution is infeasible or involves undertaking extremely high personal costs. When one's social relationships cannot be transformed into relations of equality, one's complaint against inequality does not simply disappear. Instead, one has a right to extricate oneself from unequal social relations. Yet, the state's power and authority are jurisdictionally bound. When you are in a state's jurisdiction, its laws by and large apply to you. When you are outside that state's jurisdiction, they largely do not. When you are within the state's borders, it exerts great influence over your life. When you are outside its borders, that influence is significantly attenuated.⁵⁸ Now, emigration involves exiting the state's jurisdiction. So, emigrating allows one to escape being subject to those who hold power in one's state. Social equality thus grounds a right to emigrate.

Let me emphasize two points about the remedial argument. First, the avoidance solution need not be seen as a competitor to the transformative solution. The two are not exclusive. I do not believe that the possibility of emigration renders anodyne political inequalities that would otherwise be intolerable.⁵⁹ My claim is rather that, when the transformative option is not available, one's claim against social subordination grounds a right to emigrate because doing so allows one to escape that subordination. Second, it is worth remarking on the strength of the right to emigrate that equality supports. Those who defend democracy on social egalitarian grounds typically believe that one's claim to live in a robust democracy is rather strong. If one accepts this claim, one should

58 These points are often made by subjection theorists. See, e.g., Blake, "Immigration, Jurisdiction, and Exclusion."

59 Contrast this view with Kukathas, "Exit, Freedom, and Gender."

also accept that the remedial argument grounds a claim to emigrate of *at least* equal strength, as these claims share the same ground.

3.3. *The Constitutive Argument*

The remedial argument emphasizes that *actually emigrating* can be a way to escape social inequality; therefore, one has a right to do it. This contention can be complemented by a second argument, which emphasizes the importance of *having the option* to emigrate, whether or not one exercises it. There's an important connection between exit and power. Political relationships pose a special threat to equality not only because political power is final, coercive, and pervasive, but also because political power is impossible or prohibitively difficult to avoid. This contrasts political power (ideally) with other asymmetric power relationships. Employment can be a site of domination, given the vast power bosses hold over their workers.⁶⁰ But employer power is often seen as less troubling than political power to the extent that it is more escapable.⁶¹ When labor markets offer a wide range of decent job options, workers can choose not to work for their boss. The fact that any particular boss's power is escapable makes that power more (even if not fully) compatible with social equality.

Consider again an unequal marriage. Now, compare two scenarios. In scenario one, you live in a patriarchal society that bans divorce, stigmatizes divorced persons, and provides no resources for those seeking to leave their current partner. In scenario two, you live in a society with robust divorce laws, with no stigma against divorce, and that has social policies that make leaving your partner as easy as such a consequential decision can possibly be. Being stuck in a hierarchical marriage is in case one much worse than it is in case two.⁶² But the only difference between the two cases is the way your exit options are structured. So, exit options are partially constitutive of social inequality. They are part of what *makes* unequal power and authority troubling.

Kolodny argues for this claim. He notes, for example, that "if one can exit a slave 'contract' at will, either because, as one knows, one can void it at will, or because it is already void (that is, will not be enforced by third parties), then it is not clear in what sense one really is a slave."⁶³ There is a negative and a positive lesson here. Negatively, the more inescapable a given form of unequal power is, the more troubling it is, all else being equal. This lack of exit options

60 Gourevitch, "Labor Republicanism and the Transformation of Work"; and Anderson, *Private Government*.

61 Kolodny, "Help Wanted"; Taylor, *Exit Left*.

62 Compare Brake, "Equality and Non-Hierarchy in Marriage."

63 Kolodny, "Rule over None II," 304.

is part of what makes the state's power over us so troubling since "one typically cannot escape the effects of political decisions at will, or at least not without high cost."⁶⁴ Positively, "the freer one is to exit what would otherwise be a relation of social inferiority, the less it seems a relation of social inferiority in the first place."⁶⁵ This is because "what seems to matter for relations of social inferiority and superiority is not so much equality in *actual* power, authority, and consideration, but instead equality of *opportunity* for power, authority, and consideration, where equality of opportunity is understood not as equal ex ante chances, but instead as *ongoing freedom* (both formal and informal) to exit relations of inequality."⁶⁶

Yet, Kolodny overlooks an implication of his view.⁶⁷ The ability to escape political power comes in degrees. Escaping political power can be made more feasible and less costly by altering the broader institutional arrangements in which political rule is embedded. Recognizing a robust right to emigrate makes the state's power more escapable, since the state's power is *territorially* jurisdictional, that is, particularly operative on those who reside within the state's territory. Therefore, a right to emigrate can play a *constitutive* role in reducing one's complaint against political inequality. It mitigates (the badness of) that inequality. This is especially so when the right to emigrate is accompanied by genuine exit options.

Let me clarify three things about the constitutive argument for the right to emigrate. First, again, my claim is not that having a right to exit fully alleviates the problem of hierarchy. The fact that people can emigrate does not render political inequality anodyne, for people also have a right to live as equals in their home state. Rather, my claim is that a robust right to emigrate helps temper inequality. It makes political inequality less bad than it otherwise would be. It helps make political relationships more like relations among equals. In this regard, my defense of the right to exit is on a par with egalitarian defenses of democracy. Representative democracy mitigates the badness of the inequality inherent in representative political rule by giving the ruled a share of higher-order control over their rulers; it does not eliminate the power inequality between ruler and ruled.⁶⁸

Second, I leave open the precise connection between exit and equality. There are different ways to understand this. One possibility is that when A

64 Kolodny, "Rule over None II," 305.

65 Kolodny, "Rule over None II," 304.

66 Kolodny, "Rule over None II."

67 Kolodny, "Rule over None II," 304n19. Kolodny also suggests that what matters might be the ability to exit political relations *as such*, rather than the ability to exit any particular relationship. I discuss this in section 4.1.

68 Compare Lovett, "Must Egalitarians Condemn Representative Democracy?"

has an exit option in her relationship with *B*, this mitigates the badness of *B*'s power over *A*. The thought here is that even if the extent of *B*'s power over *A* is not affected by the presence of the exit option, that the presence of a suitable exit option makes *A*'s relation to *B* *less bad* because the relation is less like an *entrenched hierarchy*. This is because social hierarchies are less bad when they are less ossified: hierarchical social relations are less bad when those subject to them are offered genuine equality of opportunity such that each person has a real option to move out of their subordinate position. On this view, the right to emigrate gives people a real option to escape relations of subordination, thereby mitigating the badness of those relations. A second possibility is that when *A* has an exit option in her relationship with *B*, *A* is less subject to *B*'s power. Suppose that *A* has power over *B* just in case *A* can get *B* to do what she wants. If *A* tries to get *B* to ϕ , *B* is likely to ϕ . This means one does not have power over people who obstinately defy one, and that one does not have power over people who one has no impact on. On this view, it is natural to think that when *B* has greater exit options this reduces *A*'s ability to get *B* to do what he wants. If *B* can easily and costlessly exit her relationship with *A*, then *B* is less likely to do what *A* wants her to do unless she desires to do so. Exit options thus reduce the power of the powerful. On either view, the right to emigrate is defensible.

Finally, note that my argument might be extended. I have focused on the inequalities in power and authority constitutive of political rule. But these are not the only relational inequalities that exist within states. Many contemporary societies contain forms of social ranking, in which some members of the state receive greater social consideration than others, due to pervasive social norms localized to that society. Examples include anti-Blackness in the United States and casteism in India. These standing-based inequalities can also be mitigated by a robust right to emigrate. This is because such inegalitarian social norms are often rather parochial. Norms about caste do not travel outside of South Asia and its diasporic communities. Anti-Blackness is particularly egregious in the United States. So, recognizing a right to emigrate can also play a limited role in ameliorating inequalities in social standing.⁶⁹ Moreover, when a state characterized by such norms hinders emigration, it is *ipso facto* enforcing those unjust social relations. This is something states have strong reason not to do.

69 My claim is *not* that a right to emigrate renders, e.g., anti-Blackness anodyne, but rather that there is value in making oppressive circumstances more escapable. Some Black political thinkers have emphasized similar ideas. Frederick Douglass, for example, although critical of the colonization movement and the Great Migration, concedes that the "right to emigrate is one of the most useful and precious of all rights" (*Negro Exodus from the Gulf States*. Baldwin emphasizes the lifesaving function of emigration in "The Art of Fiction," 78. See also Wright, "I Choose Exile."

This completes my initial defense of the claim that the value of relational equality partially grounds the right to emigrate. This view has several attractions. Like the basic rights view, it emphasizes the “protective” function of exit, though it offers a theory with wider scope. Like the freedom of movement view, the egalitarian argument entails that a merely formal right to exit is deficient, while capturing what is distinctive about exit. My view explains the appeal of voluntarism—since *unescapable* power is a special threat to social equality, having the option to exit transforms one’s relation to the state—without basing authority solely on consent. Finally, my account captures the insight that people have a claim to extricate themselves from certain relationships without treating freedom of association as an independent value.

4. OBJECTIONS

To complete the egalitarian case for the right to emigrate, however, I must address three key questions. Would such a right to emigrate really reduce inequality? Might it not instead exacerbate inequality? Does the egalitarian argument support a *right* to emigrate with sufficiently broad scope? I consider these questions in this section.

4.1. *Does the Right to Emigrate Reduce Inequality?*

One might question whether the right to emigrate reduces relational inequality. This point can be pressed in different ways in reply to the remedial and constitutive arguments. Consider first the former. Here, one might note that leaving one’s state does nothing to improve the egalitarian character of the social relations within that state. It is true that emigration does not necessarily reduce inequality for those who remain. But the thought that this undermines the remedial argument rests on two mistaken assumptions. The first is that the only way to better realize social equality is to create more equal social relationships. Rather, one can also better realize equality by *breaking* unequal ones. Early feminist advocates of the right to divorce had no illusions that it would fully transform unequal marriages into equal ones, but they nevertheless thought that the ability to escape marital domination was important.⁷⁰ Second, the objector seeks to aggregate complaints against social inequality in an inappropriate manner. Leaving one’s state may not improve political inequality for others, but it does allow one to avoid that inequality. This suffices to justify the right to emigrate, since one has a claim against social inferiority.

70 This is not the *only* case for the right to divorce; it is also justified on other grounds.

Against the constitutive argument, the objector might note that merely having a right to emigrate does little to alter one's social relationship to one's state when one does not exercise that right. After all, even if states recognize the right to emigrate, the costs of emigrating remain significant. But the right to emigrate tempers inequality only when the costs of emigration are not catastrophic.

Yet, the objection is again overstated. For one, there is a significant distinction between having no possibility of exit and there being costs to exit. Having an exit option can matter, even where that option is costly. Leaving any marriage has costs associated with it. But there is a real difference between having the option to leave and lacking it. The same holds for emigration. For another, the lesson to draw from the objection is that my account does not just entail a merely formal right to exit (i.e., a negative right against being denied exit); it requires that each person have *real* exit options (that is, a genuine option to move elsewhere). Third, the costs of emigration are variable. While they cannot be *fully* eliminated, much can be done to reduce them. What the objection shows is that states should take steps to reduce the costs of emigration. I discuss these points further in section 6.

Finally, both objections assume that having a right to emigrate does nothing to alter the social relations within one's state. Yet, this is often false. As has been remarked in a wide variety of contexts, exit options often translate into bargaining power.⁷¹ Being able to credibly threaten to exit gives one leverage. Being able to credibly threaten to leave one's state, all else being equal, heightens one's voice. If one can threaten to take one's labor elsewhere, one is more likely to be able to pressure one's employer than if one stuck in a monopsonic company town. So too do exit options give one more leverage over one's state.

Now consider a more fundamental objection to the constitutive argument. The constitutive function of the right to emigrate is, I argued, to render state power more escapable. But emigration is movement from one state to another. While a right to emigrate may allow one to escape the power of *any particular state*, it does not make *state power as such* escapable.⁷²

There are four responses to this objection that, together, explain why recognizing this important truth does not undermine the constitutive argument. The first two stem from the underlying relational structure of complaints against social inequality. Relational equality grounds both agent-neutral and agent-relative claims. It grounds claims on all to do their part to reduce social subordination: we should all work to bring about a more equal world. Yet, it also grounds

71 Taylor, *Exit Left*, 36; Hirschman, *Exit, Voice, and Loyalty*; Clark, Golder, and Golder, "An Exit, Voice, and Loyalty Model of Politics."

72 See Kolodny, "Rule over None II," 304n19.

agent-relative claims. If we stand in an unequal relationship, I have a directed claim against you that *you stop subordinating me*. This claim correlates with an agent-relative reason on the part of those who wield power to at least mitigate that power disparity—that is, a reason *for them* not to allow one to emigrate. Thus, even if one lives in a society where people of a certain group are necessarily enslaved to someone, one has a particular claim against the individual one is enslaved to as that person has elected to exercise asymmetric power over one.⁷³ So, even if one's outside options would also involve subordination, one still retains a claim to emigrate because one has a directed claim against subordination. Thus, states still have special reasons to respect the right to emigrate.

Second, the objection turns on the denial that there is a significant distinction between necessarily being subject to *someone's* power and being necessarily subject to the power of a *particular person*. It assumes that the latter lacks any independent significance. But this is not so. In the literature on employment justice, it is common to distinguish *domination by one's employer* from *structural domination in the labor market*.⁷⁴ The former occurs when one is subject to a particular boss, the latter when one is structurally forced to sell one's labor to some boss. While these problems compound one another, they are partially independent of one another. Being necessarily subject to a *particular person's* power is, in one way, worse than merely being necessarily subject to someone's power. If *A* has power over *B* to the extent that *A* can get *B* to do what she wants, then having an exit option can reduce *A's* power over *B*, just by allowing *B* to more easily escape her subjection to *A* and so avoid having to obey *A's* commands. Thus, even if the constitutive argument only resolves the former problem, this is not insignificant.

Third, the objection neglects the importance of exit options on bargaining power. The key to increased bargaining power is that one can *credibly threaten* exit. But one's threat to emigrate can be credible even where one's outside options are no better in terms of subordination than one's current options. Improving one's social relations in terms of equality is not the only reason one might want to migrate: there are a range of other reasons one might

73 It might seem that this claim stands in tension with the argument of section 5, where I contend that states' immigration policies play a constitutive role in creating unequal power relations between states and their subjects. However, *A's* claim against subordination to *B* is distinct from *A's* claim that *C* and *D* not help uphold *B's* power over *A*. An enslaved person's complaint against fugitive slave laws was that they made masters' power over enslaved persons inescapable, not that they made each white person a master. Playing an objectionable role in constituting an unequal power relationship between *A* and *B* is not equivalent to wielding that power in that relationship.

74 Compare Gourevitch, "Labor Republicanism and the Transformation of Work."

have to move and it is plausible that some other state will be better along one of these dimensions. Moreover, one's threat to emigrate can be credible *even where* the costs of exit are significant. The key thing to making an emigration threat credible is that the state *believes you might leave*. Although actual states are well aware that emigration can be costly, they are also aware that many people emigrate despite the obstacles they face. This makes most exit threats at least somewhat credible where people have a real exit option. Moreover, a just migration regime, of the sort defended in section 6, that favors ensuring admission options to democracies with low social inequality, would also further enhance the bargaining power of potential emigrants by making their exit threats more credible.

Finally, the objection assumes that all states are equally subordinating. But subordination comes in degrees. For example, one is less subject to social inequality in the average democracy than in a dictatorship. Thus, even if some amount of subordination is inescapable, emigration can be a way to escape its worst forms. Moreover, social positions are not fixed across states: there can be localized forms of status inequality that differ from state to state.⁷⁵ So, even in a world of states, the right to emigrate could mitigate relational inequality.⁷⁶

4.2. Might the Right to Emigrate Undermine Equality?

Even if the right to emigrate can mitigate inequality, one might worry that recognizing such a right might have unintended negative consequences on the distribution of power that offset its positive effects. The affluent are more mobile and can more easily absorb the costs of hopping jurisdictions. One might therefore worry that recognizing a more robust right to emigrate would give them increased bargaining power. The affluent might threaten to leave the state's jurisdiction unless political decisions are made that favor their interests.⁷⁷

75 James Baldwin explores the ambivalence of his social position in France in "Equal in Paris." He was more direct about Turkey: "Turkey saved my life" (in Zaborowska, *James Baldwin's Turkish Decade*, 8).

76 Note two further *prima facie* implications of my view. First, it gives us some reason to prefer a world in which people have the option to exit the state system altogether. Second, it gives us some reason to prefer a world of states to a world state. I doubt that considerations of equality are decisive in either case: in the former, people plausibly have a natural duty to engage in some form of social cooperation, which means opting out of the state system altogether would be unjustified; in the latter, the costs and benefits of different modes of global governance are various, and it is possible that the benefits of a world state, egalitarian or otherwise, might outweigh the egalitarian costs identified here.

77 For discussions of differential exit options, see Hirschman, *Exit, Voice and Loyalty*, 25; Taylor, *Exit Left*.

Since they command greater resources, they hold greater leverage.⁷⁸ Given this inequality, greater mobility might increase power inequalities by disproportionately increasing the wealthy's bargaining power.

It would indeed be problematic if recognizing a right to emigrate had such effects. Yet, there are reasons to doubt that it would. First, rich people do not typically flee as soon as marginal tax rates increase. There are often strong incentives for them to stay put.⁷⁹ Moreover, the right to emigrate can and should be pursued as part of a broader egalitarian agenda. The appropriate response to concerns about unequal leverage that might result from greater mobility rights is not to stop mobility. It is to address the underlying inequalities in wealth and resources that give rise to those disparities. Eliminating these disparities is independently part of what realizing relational equality requires. Egalitarians should favor redistribution *and* recognize a right to emigrate.

Bringing about greater material equality is, however, no easy task. Entrenched interests stand in the way. So, should one institutionalize such a right *absent* these broader egalitarian efforts? Note first that this worry describes the status quo. In the actual world, the affluent have greater opportunities for mobility than the poor. They can usually emigrate if they want to. The appropriate solution to this problem is not to close the border. This is in part because, on my pluralist view, there are also *nonegalitarian reasons*, in particular, reasons of autonomy, to support a right to emigrate. Since banning emigration fails to take these reasons seriously, it should not be the first-choice response to the potential differences in mobility leverage.

Instead, egalitarians should, where possible, endorse an alternative response. Bargaining power inequalities occur for two reasons: the least advantaged lack the financial and social means to move, and the affluent can use the threat of emigration as leverage. The right to emigrate can be institutionalized in ways that minimize both problems. First, the state could *resource exit* by defraying the costs of emigration for the disadvantaged.⁸⁰ It can provide resources for those who lack the means to leave. It can help them secure visas elsewhere. It can provide language courses and job training to help reduce the costs of moving. These measures make it easier for the disadvantaged to move and thus give the disadvantaged greater leverage. They also reduce the leverage the advantaged

78 Compare Lindblom, "The Market as Prison."

79 Young argues for this claim using data about the mobility behavior of millionaires both within the United States and globally (*The Myth of Millionaire Tax Flight*). Millionaires tend not to emigrate because elite social, cultural, and human capital is often place specific.

80 Similar points are developed in Taylor, *Exit Left*, 37–43 *et passim*.

have over the disadvantaged by making the wealthy's bargaining power more avoidable and making the state's power more escapable by reducing exit costs.⁸¹

Second, the state could prevent differences in mobility and wealth from being converted into differences in leverage.⁸² The simplest way to do this is to create mechanisms that stop the affluent from threatening to take their resources with them when they leave, such as exit taxes on the superrich.⁸³ Exit taxes of the right kind can block the conversion of the rich's mobility options into superior bargaining power. If one cannot take all of one's wealth when one leaves the state, then one is less able to leverage one's resources into bargaining power by threatening to leave. Since the right to emigrate can be tailored to reduce bargaining power inequalities, one need not ban emigration in order to counteract these inequalities.

Similar points apply to a second version of this worry, which concerns human capital. One might worry that recognizing a robust right to emigrate would bolster the bargaining position of those much-coveted, scarce skills, such as medical training, vis-à-vis the state, and thus generate an inequality between them and their compatriots. However, health care workers in the developing world do not have much leverage over their compatriots, even though they are much more mobile. Otherwise, one would expect them to have successfully bargained for more competitive pay. This has not, as a rule, been the case. Health expenditure remains only a fraction of GDP in most developing states. Moreover, it is not clear that leverage vis-à-vis employers, e.g., translates into leverage over other members of society. It thus seems unlikely that high-skilled workers can leverage emigration threats in a manner that will give them substantial power over their low-skilled counterparts.

This raises, however, a more general question about how my view should deal with the potential for emigration to cause brain drain and other similar phenomena. Empirically, scholars offer varying assessment of net costs and benefits of emigration of high-skilled medical workers.⁸⁴ Yet, there is little evidence that most forms of emigration exacerbate relational inequality.⁸⁵ More importantly,

81 Taylor, *Exit Left*.

82 Compare Walzer's claim that we should block the conversion of inequalities in one "sphere" into inequalities in another (*Spheres of Justice*, 17–20).

83 Stilz defends this proposal in "Unqualified Right."

84 For an overview of empirical research on the brain drain, see Brücker, Docquier, and Rapoport, *Brain Drain and Brain Gain*.

85 Consider two further reasons one might think emigration increases inequality. First, emigration may have negative distributive effects in source countries. Docquier, Ozden, and Peri argue that emigration can depress low-skilled workers' wages ("The Labour Market Effects of Immigration and Emigration in OECD Countries"). Such studies are, however,

we should not settle our general view of the right to emigrate solely by focusing on brain drain. Indeed, the debate about brain drain typically *presupposes* that there is such a right and asks when it may be permissibly restricted.⁸⁶

This is not to deny that there might be cases in which large-scale emigration would exacerbate relational (or distributive) inequality or have other negative social impacts. However, given the weighty interests at stake in emigration, ways of mitigating emigration-induced inequalities that do not involve restricting mobility, such as exit taxes or positive incentives to stay, should be favored where feasible. Cases where emigration would enhance inequality or cause serious harm and where that harm can *only* be averted by emigration restrictions at most indicate that the right to emigrate is defeasible and that there may be transition costs to moving toward a just global migration regime that can address these problems in a coordinated way. Any all-things-considered defense of emigration restrictions in a particular case requires grappling seriously with the values that justify the right to emigrate in the first place. This article thus lays the groundwork for discussions of whether and how emigration restrictions might be justified—an important issue that is beyond the scope of this paper.⁸⁷

4.3. *Does the Right to Emigrate Have Sufficiently Broad Scope?*

The right to emigrate is usually understood to apply to all persons in all states. Yet, it might seem as though my argument cannot support a *universal* right to emigrate. Consider first whether my argument applies to all states. Perhaps inequality can be reduced sufficiently without recognizing the right to emigrate.

premised on the status quo, in which migratory options are deeply unequal. Karadja and Prawitz suggest emigration can also have positive egalitarian benefits (“Exit, Voice, and Political Change”). Second, emigration might function as a safety valve for autocratic regimes. Pfaff and Kim find an inverse relation between exit and protest (“Exit-Voice Dynamics in Collective Action”). Yet, there is a reason why autocrats are so keen on controlling emigration. Alemán and Woods suggest emigration *restrictions* enhance autocratic stability (“No Way Out”). Peters paints a nuanced picture on which emigration to democracies undermines autocratic stability (“Restraining the Huddled Masses”). The autocratic safety-valve approach likely functions only when autocrats can *stop* most people from migrating—a violation of the right I defend—while *selectively* encouraging others to leave. In neither case are emigration’s overall effects clearly inegalitarian and it is anyway doubtful that these kinds of effects would offset the clear egalitarian gains of a robust right to emigrate.

86 Compare Brock and Blake, *Debating Brain Drain*; Oberman, “Can Brain Drain Justify Immigration Restrictions?”

87 Other important issues the right to emigrate raises include its relation to theories of political obligation and to the state’s power to execute just punishment to, e.g., fairly incentivize compliance with the law.

Consider an egalitarian utopia that distributes all goods, including political power, equally. In such a utopia, the right to emigrate might seem unnecessary: there is no inequality that must be rendered escapable. One might extrapolate from this. Perhaps some actual states—maybe Iceland—are already sufficiently equal. So, perhaps equality does not justify a right to emigrate for Icelanders. Thus, one might conclude, my argument cannot justify a right to emigrate *that applies to all states*.

I concede that the strength of one's claims to emigrate can vary. They can vary according to the degree of inequality to which one is subject. Yet, this is a feature of my account, not a bug. Any theory of emigration should recognize that actual claims to emigrate will vary in strength. In states less marred by inequality, one's claim to emigrate may be weaker; the worse one's social position in one's state is, the stronger one's claim. Still, the right to emigrate has an important function even in the most ideal political community. Relational equality demands that we *minimize and constrain* the power disparities in political rule to the *greatest degree possible* consistent with other values. Yet, even the most ideal political community—short of an impracticable form of full direct democracy—would likely contain *considerable* political inequality. Recognizing a right to emigrate thus has a role to play in redressing inequality in all states.

One might wonder whether this role is significant enough to justify a strong *right* to emigrate from the egalitarian utopia. I am admittedly uncertain how strong one's egalitarian claim to emigrate would be in this case. But I do not think this is damning. My argument still shows that relational equality provides a *strong reason* to protect a right to leave: it is essential to ensure that social cooperation is really free cooperation among equals. This is an important and novel conclusion about what kinds of migration policies egalitarians should favor. So, my argument is hardly trivial, even if one believes the connection between emigration and equality is purely instrumental and contingent in nature. Moreover, there are no egalitarian utopias. No actual society—not even Iceland!—comes close. So, my argument still justifies a right to emigrate from all actual states.

Still, there is a different way in which my argument might fail to be universal in scope. Perhaps it shows that all who are subjected to political inequality have a right to emigrate. But not everyone is so subjected. Those who occupy *superior* social positions are not. So, one might think, my argument cannot justify a right to emigrate for everyone.

The objection is again overstated. Even if those in positions of social superiority do not have a direct claim to emigrate, they may still have a *derivative* claim to emigrate. Standing in relations of superiority is bad: one has strong moral reason not to do it. But if one occupies such a position unavoidably due to, e.g., pervasive social norms that confer greater standing on members of one's

social group, one might have a claim to extricate oneself from such relations. Now, many such people *avoidably* occupy positions of superiority. The powerful could give up their power. But they could nevertheless acquire such a right by renouncing their power.

More importantly, my argument justifies a right to emigrate for the vast majority of people. Social and political power and authority are, in all actual societies, heavily concentrated. The fact that an egalitarian justification does not extend to everyone in all possible circumstances need not undermine its force in the cases in which it does apply. Finally, as I argued earlier, the full case for the right to emigrate is pluralistic.⁸⁸ Equality plays an important role in justifying the right to emigrate. But autonomy considerations and other positive interests in mobility have a role to play too. Equality need not do all the work.⁸⁹ So, lack of universality is not a decisive reason to reject my egalitarian argument.

5. AN ARGUMENT FOR THE RIGHT TO IMMIGRATE

I have defended the right to emigrate. I will now make a derivative argument for a right to immigrate. On the standard understanding, there is a universal but merely formal right to emigrate but no corresponding right to immigrate.⁹⁰ While it is unjust for one's state of current residence to bar one from leaving, no such injustice occurs when other states refuse to let one enter. My egalitarian argument shows that this is the wrong way to understand the relation between exit and entry.⁹¹

My argument for a limited right to immigrate begins from the observation that a merely formal right to emigrate does very little to reduce social inequality. It does nothing to make state power escapable. What is crucial is that one has *genuine exit options* that can be exercised *without prohibitive cost*. A merely formal right to emigrate does not provide such options. It does not guarantee that emigrants have anywhere to go. The right to emigrate can only serve its egalitarian function when accompanied by a corresponding right to enter at

88 The pluralist nature of my defense of the right to emigrate raises the question of how these different grounds of the right to emigrate relate to one another. A plausible view is that basic rights take lexical priority over relational equality, which in turn often, but not always, takes priority over one's positive interests in freedom of movement. This view has implications for how a just migration regime should be designed, which I discuss below.

89 Perhaps equality can do more work still: the principle of equal treatment may also require that we recognize a right to emigrate for all once we recognize it for some.

90 For a defense, see Wellman, "Freedom of Movement."

91 For other criticisms of the standard view, see Cole, *Debating Immigration*; Lenard, "Exit and the Duty to Admit."

least some other state. This provides some reason to favor a limited right to immigrate *somewhere*.

Now, one might object that this does not immediately ground a *right* to immigrate. The fact that protecting people's exit options would serve an egalitarian function does not entail that states are *under a corresponding duty* to accept immigrants. Accepting immigrants, it is often alleged, comes with costs. It must be shown that other states have a responsibility to bear those putative costs. But it is sometimes thought that the duties states have to outsiders are limited: it is not obvious that states must benefit outsiders, let alone benefit them by allowing them to immigrate. So, while states have reason to accept immigrants in order to help realize social equality by promoting effective exit, one might nevertheless deny that states have a duty to do so.

This challenge can be answered in two steps. The first step is to show why states have some special duty or reason to accept emigrants, which makes it appropriate to require them to bear the burdens of doing so. The objection alleges that when states exclude immigrants, they are merely failing to benefit them and that excluding states have no special responsibility for their predicament. But these claims are false. When states refuse to accept emigrants, they are not merely failing to benefit them; they are *actively helping to maintain* the inequality that emigrants endure. States do quite a lot to prevent people from moving. They build walls and post armed guards. They build elaborate architectures of remote control. All this means that states bear some responsibility for the plight of outsiders seeking to emigrate.⁹²

This point can be made in two different ways. The stronger version claims that states generally play an active role in constituting the very inequality at stake. Recall that the significance of the inequality between a given citizen and their state is a function of the extent of the state's power *and* of the individual's exit options. But an individual's exit options are determined by the actions of other states generally. So, by excluding immigrants, the community of states plays a constitutive role in making the power of each state more difficult to escape. But if states' exclusionary immigration policies generally play a role in constituting social inequality in other states, then states have a special responsibility to help redress that inequality. Emigrants have a complaint against *states in general* that they stop constraining their exit options.

This argument depends on the potentially controversial claim that actions of a third party, *C* (or a group of third parties, *G*) can be constitutive of a power

92 There is an even weaker way to make the case. We have weighty reason not to allow others to be subordinated when it is not costly for us to prevent it. You should jump in the pond to free a slave even if it ruins your suit! This principle applies to states. They should therefore allow some immigration, absent defeating considerations.

relation between *A* and *B*. Yet, I think this claim is defensible.⁹³ Consider the system of American chattel slavery. The relationship between enslavers and enslaved persons clearly involved a “dyadic” power relation of an asymmetric and objectionable sort. But these problematic dyadic power relations were themselves constituted by—and indeed only existed because of—a broader network of social institutions that made it possible.⁹⁴ One central institution here was the fact that slave ownership was enforced by third parties, such as slave catchers, who were backed by law and would apprehend and return enslaved persons to their enslavers. This social institution helped constitute the power relationship between the particular “master” and the “slave.” My claim is that states collectively play this kind of role in constituting power asymmetries. States typically return those they exclude to their states of origin; this helps constitute the power of their states over them.

More broadly, states are, one and all, participants in a practice of territorial sovereignty. This practice constitutively confines people to political rule in particular territories and thus subjects them to rule by those who wield power in those territories. States participating in this practice collectively *recognize* those rulers, authorize them to wield political power, and often causally support them in doing so. These facts mean that states collectively bear responsibility for the plight of those who live under the rule of the state system and, in particular, for ensuring that this system is compatible with the demands of equality.⁹⁵

However, the case for a limited right to immigrate does not hinge on these controversial claims. A weaker point suffices. Even if states’ exclusionary immigration policies do not help constitute inequality in states of emigration, such policies still help cause it. They are a key part of the reason why outsiders must endure that inequality. States thus knowingly contribute to maintaining relations of subordination. That people are generally denied the right to emigrate is a kind of structural injustice. This is a direct result of the actions of states collectively. States bring this situation about through their immigration control policies. They know that these policies reduce the exit options of those in other states, and they know (or could easily know) that this results in subordination. Greater social inequality is a foreseeable result of their actions. So, by acting to exclude immigrants, states knowingly help bring about grave structural injustice. States should not knowingly help cause structural injustice and it is only by changing their immigration policies that states could avoid doing so.

93 For a further defense, see Vrousalis, “The Capitalist Cage.”

94 Compare Gourevitch, “Labor Republicanism and the Transformation of Work,” 601–3.

95 My argument thus highlights the significance of the way states interact and are positioned within a system of states. For a critical overview of arguments of this kind, see Sharp, “Immigration and State System Legitimacy.”

Still, one might retort, this duty is too demanding. The claim that each state must accept an unlimited number of immigrants would be too high a burden to impose. This brings me to the final step of my argument. My egalitarian case for the right to emigrate straightforwardly does not entail that emigrants have a right to enter whichever country they choose. Rather it shows that states are under a *collective duty* to institute a global migration system that (a) adequately protects the right to emigrate and (b) can fairly distribute the costs of protecting this right. Under such a system, the costs of accepting emigrants would not be prohibitive; so, states are required to institute such a system. I elaborate on these claims further below as part of a broader exploration of states' duties in regulating migration.

6. IMPLICATIONS

I will conclude with a brief discussion of the implications of my theory for the duties of states. I highlight the duties that this right imposes on states of emigration (section 6.1), then consider the duties it imposes on states collectively (section 6.2). I conclude with a brief remark on what states should do under conditions of noncompliance (section 6.3).

6.1. *The Duties of States of Emigration*

The right to emigrate is often thought to impose only a negative duty on one's own state not to prevent one from emigrating. However, on my account, the right to emigrate also grounds a number of positive duties on one's state. An underlying aim of the right to emigrate is to temper the inequality inherent in political rule by making that inequality more escapable. But the extent to which political inequality is escapable depends on how costly people's exit options are. States therefore have positive duties to reduce the costs of exit for their citizens, especially the most disadvantaged ones.

There are three primary costs associated with exit—exit, mobility, and entry costs—and states can reduce all of these costs. *Exit costs* are associated with what one leaves behind: friends, family, opportunities, attachments, culture. The state can reduce these costs by recognizing a right to return, such that those who emigrate are not permanently cut off from their significant social and cultural attachments.⁹⁶ *Mobility costs* are associated with moving itself. The state can assist in securing visas and defraying the financial costs of emigrating, particularly for the disadvantaged (or states could collectively eliminate visas altogether). *Entry costs* are associated with adapting to one's new destination:

96 This right is already recognized in international law.

finding employment, learning a new language, making social contacts. Here, states can play a role in helping emigrants adapt: they can help emigrants find employment, provide job training, acquire language skills, and provide support to social organizations for emigrants.⁹⁷ States have significant latitude in how they discharge these duties, and these duties are limited by considerations of demandingness. But if there are easy ways to reduce the costs of emigration, states should do so.

6.2. *Duties on States Collectively*

I argued that relational egalitarianism requires real and robust options to emigrate and this entails a collective duty on the part of states to work to create political institutions that make this possible. What kinds of institutional arrangements would protect the right to emigrate? This section explores this issue. First, I highlight some minimal requirements that must be met to protect effective exit. I then argue that these minimal requirements do not best realize equality, and that there is strong reason to favor a regime characterized by greater openness. Finally, I argue that although a world of open borders would *best* realize relational equality, it is not clear that egalitarians must endorse open borders. The overall picture is one on which states may have some discretion in how they institutionalize a global migration regime that protects effective exit.

At a minimum, a migration system must meet two conditions to protect effective exit. First, it must ensure that each person has at least one state to which they can immigrate. To play its constitutive function in mitigating inequality, each person must have a real option to emigrate, which requires a corresponding option to immigrate somewhere. Second, it must ensure that this destination is a decent option. Otherwise, the right to emigrate would not effectively play its constitutive role. The option to immigrate to North Korea would not reduce social inequality. This means that states with comparatively democratic governments, lower political inequality, and a decent standard of living should be preferred.

These minimal conclusions entail that a strictly unilateral and discretionary system of migration control is untenable. The denial of effective exit options for some is a predictable result of states exercising unilateral control over their borders. States therefore have a duty to engage in cooperative endeavors to ensure that a system of effective exit is realized. Such a system would distribute the burdens of protecting effective exit such that no one state must bear all the costs.

97 I leave open whether the duty to defray these costs should fall on states of emigration or destination states.

Although this *minimalist system* of global migration management would make the international order more compatible with equality, it would not adequately protect the right to emigrate. This requires making inequality sufficiently escapable. To do this, effort must be made to adequately reduce the costs of emigration. But the minimalist system makes few such efforts. It provides emigrants with only one option, makes no provision for taking their destination preferences into account, and makes no provisions for group or family migration. Under such a system, the costs of emigration will remain intolerably high for many.

Therefore, a more open system is plausibly required. Such a system should provide emigrants with a range of potential destinations, protect family migration, and take their migration preferences into account. A variety of different institutional arrangements might meet these moderately more demanding requirements. A global migration regime of this kind might, for example, incorporate a preference-matching system, in which both states and emigrants “rank” destination choices. Or it might use some other mechanism for factoring in mobility preferences. Admittedly, this system would require receiving states to take on greater burdens. It is an open question how high these burdens are and whether states are required to take them on. I believe that states are generally obligated to take on these burdens. This is not only because equality is an important value and greater openness would also better realize many important freedoms. It is also because immigrants are often a net benefit to the states that accept them and the burdens of accepting them are often overstated. However, I will not argue for these claims here, since my aim is not to settle this issue.

This system of *moderate openness* is, however, nevertheless compatible with significant (though not unlimited) discretion on the part of states concerning whom to admit and thus stops short of open borders. Equality does not require that each person has a right to enter whichever state they choose. Protecting effective exit is thus compatible with a world in which states collectively manage migration, as long as states coordinate appropriately and the costs of emigration are suitably reduced.

However, there is reason to believe that a world of open borders, combined with widespread real opportunities for mobility, would best realize effective exit. This is because open borders would reduce the costs of emigration to the greatest extent. However, equality arguably does not require open borders, since states might merely be required to implement a sufficiently egalitarian migration system, rather than an optimally egalitarian one. Nevertheless, equality provides support for open borders *as an ideal*.⁹⁸ The important point for our

98 Compare Oberman, “Immigration as a Human Right,” 33n2.

purposes, however, is that all three systems of migration governance would fare much better than current arrangements at protecting the right to emigrate. How heavily one weights the values that comprise my pluralist view as well as how one regards the costs of emigration will play a crucial role in how one thinks a system protecting effective exit should be designed and thus in determining which of these systems for regulating global mobility one ultimately favors.

6.3. Duties Under Noncompliance

I have argued that states are under a collective duty to institute a fair global migration system that better protects the right to emigrate. But states have, obviously, yet to institute such a system. They have thereby failed to do their duty. What does my argument imply under such conditions? The primary thing it implies is that states are required to accept a substantially greater number of immigrants. This is because doing so would help better approximate the ideal of social equality on which the right to emigrate is partially based. This duty is not, however, unlimited. So, states may perhaps prioritize those whose egalitarian claims are most urgent. My argument thus provides yet another reason why states should loosen their immigration restrictions.⁹⁹

Ludwig-Maximilians-Universität München
daniel.sharp@lmu.de

REFERENCES

- Alemán, José, and Dwayne Woods. "No Way Out: Travel Restrictions and Authoritarian Regimes." *Migration and Development* 3, no. 2 (2014): 285–305.
- Anderson, Elizabeth. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk About It)*. Princeton, NJ: Princeton University Press, 2017.
- . "The Problem of Equality from a Political Economy Perspective." *Oxford Studies in Political Philosophy*, vol. 3, edited by David Sobel, Peter Vallentyne, and Steven Wall, 36–57. Oxford: Oxford University Press, 2017.
- Baldwin, James. "The Art of Fiction 78: James Baldwin." *Paris Review* 91 (1984):

99 For a complementary argument, see Sharp, "Relational Equality and Immigration." I thank Daniel Viehoff, Samuel Scheffler, Katharina Anna Sodoma, Matthew Lister, Matthias Hoesch, Adam Lovett, Sophie Cote, Rob Long, Clara Lingle, and Marko Malink as well as participants at Social Ontology 2021, the Ethics and Applied Philosophy Graduate Conference at University of North Carolina Charlotte, Washington Square Circle, and the NYU Equality in Immigration Workshop for helpful feedback.

- 48–82.
- . “Equal in Paris.” *Commentary* 20 (1955): 251.
- Blake, Michael. “Immigration, Jurisdiction, and Exclusion.” *Philosophy and Public Affairs* 41, no. 2 (Spring 2013): 103–30.
- . *Justice, Migration, and Mercy*. Oxford: Oxford University Press, 2019.
- Brake, Elizabeth. “Equality and Non-Hierarchy in Marriage.” In *After Marriage: Rethinking Marital Relationships*, edited by Elizabeth Brake, 100–24. Oxford: Oxford University Press, 2016.
- Brock, Gillian, and Michael Blake. *Debating Brain Drain: May Governments Restrict Emigration?* Oxford: Oxford University Press, 2015.
- Brownlee, Kimberley. “Freedom of Association.” In *A Companion to Applied Philosophy*, edited by Kasper Lippert-Rasmussen, Kimberley Brownlee, and David Coady, 356–69. London: Wiley-Blackwell, 2016.
- Buchanan, Allen. “Theories of Secession.” *Philosophy and Public Affairs* 26, no. 1 (Winter 1997): 31–61.
- Clark, William R., Matt Golder, and Sona Golder. “An Exit, Voice, and Loyalty Model of Politics.” *British Journal of Political Science* 47, no. 4 (October 2017): 719–48.
- Cole, Phillip. *Philosophies of Exclusion: Liberal Political Theory and Immigration*. Edinburgh: University of Edinburgh Press, 2000.
- Docquier, Frédéric, Çağlar Ozden, and Giovanni Peri. “The Labour Market Effects of Immigration and Emigration in OECD Countries.” *Economic Journal* 124, no. 579 (September 2014): 1106–45.
- Douglass, Frederick. *Negro Exodus from the Gulf States*. Charlottesville, VA: University of Virginia Library, 1996.
- Fine, Sarah. “Freedom of Association Is Not the Answer.” *Ethics* 120, no. 2 (January 2010): 338–56.
- Fine, Sarah, and Lea Ypi, eds. *Migration in Political Theory: The Ethics of Movement and Membership*. Oxford: Oxford University Press, 2016.
- Gourevitch, Alex. “Labor Republicanism and the Transformation of Work.” *Political Theory* 41, no. 4 (August 2013): 591–617.
- Hirschman, Albert O. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press, 1970.
- Hume, David. “Of the Original Contract.” In *Hume: Political Essays*, edited by Knud Haakonssen, 186–201. Cambridge: Cambridge University Press, 1994.
- Karadja, Mounir, and Erik Prawitz. “Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States.” *Journal of Political Economy* 127, no. 4 (August 2019): 1864–1925.
- Kolodny, Niko. “Help Wanted: Subordinates.” In Anderson, *Private Government*, 99–107.

- . “Political Rule and Its Discontents.” In *Oxford Studies in Political Philosophy*, vol. 2, edited by David Sobel, Peter Vallentyne, and Steven Wall, 35–68. Oxford: Oxford University Press, 2016.
- . “Rule over None II: Social Equality and the Justification of Democracy.” *Philosophy and Public Affairs* 42, no. 4 (Fall 2014): 287–336.
- . “Standing and the Sources of Liberalism.” *Politics, Philosophy and Economics* 17, no. 2 (May 2018): 169–91.
- Kukathas, Chandran. “Are There Any Cultural Rights?” *Political Theory* 20, no. 1 (February 1992): 105–39.
- . “Exit, Freedom, and Gender.” In *On Exit: Interdisciplinary Perspectives on the Right of Exit in Liberal Multicultural Societies*, edited by Dagmar Borchers and Annamari Vitikainen, 34–56. Berlin: De Gruyter, 2012.
- . *The Liberal Archipelago: A Theory of Diversity and Freedom*. Oxford: Oxford University Press, 2003.
- Lenard, Patti Tamara. “Exit and the Duty to Admit.” *Ethics and Global Politics* 8, no. 1 (2015): 259–75.
- Lindblom, Charles E. “The Market as Prison.” *Journal of Politics* 44, no. 2 (May 1982): 324–36.
- Locke, John. *Two Treatises of Government*. Cambridge: Cambridge University Press, 1988.
- Lovett, Adam. “Must Egalitarians Condemn Representative Democracy?” *Social Theory and Practice* 47, no. 1 (January 2021): 171–98.
- Lovett, Adam, and Daniel Sharp. “What Immigrants Owe.” *Ergo* 8, no. 36 (2021): 336–66.
- Mancilla, Alejandra. *The Right of Necessity: Moral Cosmopolitanism and Global Poverty*. London: Rowman and Littlefield, 2016.
- Mendoza, José Jorge. “Enforcement Matters: Reframing the Philosophical Debate over Immigration.” *Journal of Speculative Philosophy* 29, no. 1 (2015): 73–90.
- Miller, David. “Is There a Human Right to Migrate?” In *Migration in Political Theory: The Ethics of Movement and Membership*, edited by Sarah Fine and Lea Ypi, 11–31. Oxford: Oxford University Press, 2016.
- Oberman, Kieran. “Emigration in a Time of Cholera: Freedom, Brain Drain, and Human Rights.” *Law, Ethics, and Philosophy* 4 (2016): 87–108.
- . “Immigration as a Human Right.” In Fine and Ypi, *Migration in Political Theory*, 32–56.
- Okin, Susan Moller. “‘Mistresses of Their Own Destiny’: Group Rights, Gender, and Realistic Rights of Exit.” *Ethics* 112, no. 2 (January 2002): 205–30.
- Otsuka, Michael. *Libertarianism without Inequality*. Oxford: Clarendon Press, 2005.

- Peters, Margaret E. "Restraining the Huddled Masses: Migration Policy and Autocratic Survival." *British Journal of Political Science* 50, no. 2 (April 2020): 403–33.
- Pfaff, Steven, and Hyojoung Kim. "Exit-Voice Dynamics in Collective Action: An Analysis of Emigration and Protest in the East German Revolution." *American Journal of Sociology* 109, no. 2 (2003): 401–44.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
- Scanlon, T. M. *Why Does Inequality Matter?* Oxford: Oxford University Press, 2018.
- Sharp, Daniel. "Immigration and State System Legitimacy." *Critical Review of International Social and Political Philosophy* (forthcoming). Published online ahead of print, December 8, 2020. <https://doi.org/10.1080/13698230.2020.1860416>.
- . "Relational Equality and Immigration." *Ethics* 132, no. 3 (April 2022): 644–79.
- Simmons, A. John. *Justification and Legitimacy: Essays on Rights and Obligations*. Cambridge: Cambridge University Press, 2001.
- . *The Lockean Theory of Rights*. Princeton, NJ: Princeton University Press, 1994.
- . *Moral Principles and Political Obligations*. Princeton, NJ: Princeton University Press, 1979.
- Song, Sarah. *Immigration and Democracy*. Oxford: Oxford University Press, 2018.
- Stilz, Anna. "Is There an Unqualified Right to Leave?" In Fine and Ypi, *Migration in Political Theory*, 57–79.
- . *Territorial Sovereignty: A Philosophical Exploration*. Oxford: Oxford University Press, 2019.
- Taylor, Robert S. *Exit Left: Markets and Mobility in Republican Thought*. Oxford: Oxford University Press, 2017.
- Van der Vossen, Bas. "Immigration and Self-Determination." *Politics, Philosophy and Economics* 14, no. 3 (August 2015): 270–90.
- Viehoff, Daniel. "Power and Equality." *Oxford Studies in Political Philosophy*, vol. 5, edited by David Sobel, Peter Vallentyne, and Steven Wall, 1–38. Oxford: Oxford University Press, 2019.
- Vrousalis, Nicholas. "The Capitalist Cage: Structural Domination and Collective Agency in the Market." *Journal of Applied Philosophy* 38, no. 1 (February 2021): 40–54.
- Walzer, Michael. *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic Books, 1984.
- Wellman, Christopher Heath. "Samaritanism and the Duty to Obey the Law."

- In *Is There a Duty to Obey the Law?* edited by Christopher Heath Wellman and A. John Simmons, 3–91. Oxford: Oxford University Press, 2005.
- Wellman, Christopher Heath and Philip Cole. *Debating the Ethics of Immigration: Is There a Right to Exclude?* Oxford: Oxford University Press, 2011.
- . “Freedom of Movement and the Rights to Enter and Exit.” In Fine and Ypi, *Migration in Political Theory*, 80–101.
- Whelan, Frederick G. “Citizenship and the Right to Leave.” *American Political Science Review* 75, no. 3 (September 1981): 636–53.
- Wright, Richard. “I Choose Exile.” Department of Special Collections and Archives, Kent State University. Unpublished manuscript, 1951.
- Young, Cristobal. *The Myth of Millionaire Tax Flight: How Place Still Matters for the Rich*. Palo Alto, CA: Stanford University Press, 2017.
- Ypi, Lea. “Justice in Migration: A Closed Borders Utopia?” *Journal of Political Philosophy* 16, no. 4 (December 2008): 391–418.
- Zaborowska, Magdalena J. *James Baldwin’s Turkish Decade: Erotics of Exile*. Durham, NC: Duke University Press, 2009.

INSTITUTIONAL CONSERVATISM AND THE RIGHT TO EXCLUDE

Hallvard Sandven

IT IS A COMMON VIEW in analytical political theory that many if not most of the normative questions brought about by global migration are ultimately explained by the viability of granting states the “right to exclude.” The main thought is that if states have this right they are entitled to enforce their border regimes over outsiders, despite the fact that those outsiders are also excluded from the institutional structures that authorize that enforcement. Similarly, if states have this right, then they retain the privilege of determining how and in what sequence they reform their border regimes when they fall short of applicable standards of justice. Among those who think that migration ought to be analyzed by reference to the right to exclude, it is also common to see the grounds of that right as shaping the standards of justice that constrain its exercise. States may have the right to exclude, so the thought goes, but they cannot use that right however they want. The right to exclude is thus taken to denote both the rightful claim to regulate a given domain and the standard according to which that domain ought to be regulated: it determines both the legitimacy and justice of immigration restrictions.

There are several sophisticated theories of the right to exclude on offer, seeking to ground the right in the value of national culture, in citizens’ ownership claims, or in their claims to freedom of association.¹ These views face a common challenge, however. Even if they often provide normatively strong grounds for exclusion, they all rely on idealized conceptions of the state. This feature gives rise to a problem of applicability. To describe states in the moralized terms that satisfy the triggering conditions of the right to exclude, these theories must abstract away from central aspects of states as they currently exist. But having idealized the subject of their theories—the state—the challenge is now to explain how and why these accounts can vindicate privileges on the part of the real states that currently claim authority over actual people. In light of the

1 Miller, *Strangers in Our Midst*; Walzer, *Spheres of Justice*; Pevnick, *Immigration and the Constraints of Justice*; Simmons, *Boundaries of Authority*; Wellman, “Immigration and Freedom of Association.”

pervasiveness of this problem across different theories, it is worth taking seriously a recent methodological strategy for grounding the state's right to exclude. According to Michael Blake and Sarah Song, the right to exclude can be derived from a thin descriptive account of the state.² Thus, instead of appealing to moralized conditions that contemporary states only partly fulfill (if at all), Blake and Song appeal to minimal conception of the state as a sovereign jurisdiction.

This article offers a critical discussion of this—as Blake calls it—*institutionally conservative* approach to the right to exclude.³ It argues that, while the move to institutional conservatism convincingly avoids the problem of applicability associated with theories that moralize the state, it faces an applicability problem of its own. This problem, I will show, stems from the fact that institutional conservatism is structurally conditioned to presuppose the legitimacy of border control. Thus, institutional conservative theory cannot explain why states should retain the privilege to unilaterally enforce border control when they fall short of applicable standards of justice in migration. This means, further, that institutional conservatism is incapable of vindicating these privileges on the part of existing states—a crucial problem insofar as its proponents want their theories to shed light on the contemporary politics of migration.

Building on this analysis, the article argues that legitimacy assessments of actual instantiations of border control should be decoupled from state-based accounts of legitimacy. A general upshot of the article is that focusing on the individual state's right to exclude is an unproductive frame for assessing contemporary border regimes. Thus, the article provides further support for the thesis that the ethics of migration requires thinking about the state *system*.⁴ In other words, my critique of institutional conservatism leads to a more general critique of methodologically nationalist approaches to the ethics of migration.⁵

The article is structured as follows. Section 1 introduces the debate on the state's right to exclude, highlighting how that debate gives rise to the related, but importantly distinct, questions of justice in immigration policy and the legitimacy of border control. Section 2 briefly outlines three influential approaches to the right to exclude, shows how these idealize the state, and argues that they therefore face a problem of applicability. Section 3 outlines institutional conservatism and shows how it avoids the problem of applicability that plagues its theoretical competitors. It then argues that institutional conservatives are

2 Blake, "Immigration, Jurisdiction, and Exclusion"; *Justice, Migration and Virtue*; Song, "Why Do States Have the Right to Control Immigration?" and *Immigration and Democracy*.

3 Blake, *Justice, Migration and Virtue*, 10.

4 Bertram, *Do States Have the Right to Control Immigration?*; Brock, *Justice for People On the Move*; Owen, *What Do We Owe to Refugees?*; Sharp, "Relational Equality and Immigration."

5 Sager, "Methodological Nationalism, Migration, and Political Theory."

bound to conflate the concepts of justice and legitimacy, rendering their theories incapable of vindicating the privileges conferred by legitimacy for actually existing states. Section 4 considers an objection to my argument, which seeks to rescue institutional conservatism from the charge of simply stipulating the legitimacy of border control by way of a rational reconstruction of the state system. I argue that, while this response is available to institutional conservatives, it cannot support the kind of authority typically implied by the right to exclude. Section 5 concludes by drawing some general lessons of my discussion for the theoretical debate on the ethics and politics of migration.

1. JUSTICE IN IMMIGRATION POLICY AND THE LEGITIMACY OF BORDER CONTROL

Through their border regimes, states collectively determine who has access to which labor markets, education systems, and social security networks. They also decide who will have their basic human rights protected when other states are unwilling or unable to perform that task. By serving these functions, the institution of border control sustains and reproduces global inequality and ensures that many vulnerable individuals fail to have their human rights protected. Consequently, political philosophers and normative political theorists have scrutinized the institution of border control and asked whether, and if so how, it can be rendered compatible with the demands of political morality. In particular, they have sought to find a justification for the right each state claims to set and enforce its own immigration policy without outside interference. In the literature, this is called the state's right to exclude.⁶ The right to exclude is a Hohfeldian power in that it grants states the capacity to change outsiders' normative relationship—their claims, obligations, and liabilities—to the state itself. It is also a privilege in that states can unilaterally decide how they exercise this power.⁷ This is reflected in how states grant or deny visas, permanent residencies, or citizenship to previous nonmembers; attach distinct bundles of obligations and liabilities to each of these statuses; and unilaterally decide to whom they extend which status.⁸

The right to exclude is so central to standard conceptions of sovereignty that some philosophers have thought that it is definitionally tied to what it means for states to be legitimate. As David Copp emphasizes in his argument to this conclusion, however, having a privilege does not settle the question of

6 For an overview by a key contributor to the debate, see Fine, "The Ethics of Immigration."

7 For an exposition of the Hohfeldian incidents, see Wenar, "The Nature of Rights."

8 Fine, "The Ethics of Immigration," 255.

how that privilege should be exercised. Thus, even if legitimate states have the right to control borders, this does not settle how they ought to regulate immigration.⁹ Since a privilege entails the absence of a claim, it might be thought that justified claims for admittance on the part of some migrants—for example, refugees or members of the global poor—undercuts the existence of a right to exclude. However, even if there are individuals who have claims to migrate from their current state to wealthier and more stable parts of the world, the states in those parts of the world may retain a crucial privilege: to interpret both who has justified claims for emigration and what constitutes their fair share of the collective obligations to discharge them.¹⁰ Thus, in the absence of a claim on the part of *all* potential migrants, the thought goes, the right to exclude follows from state legitimacy. The fact that there are moral constraints on how states exercise their right to make their own policy decisions does not undercut their right to make such decisions.

Given this intuitive view, it unsurprising that the contemporary debate on the right to exclude arose in the wake of arguments seeking to establish that everyone has a claim to immigrate.¹¹ Following Copp's line of argument, the main strategy for proponents of the right to exclude has not been to deny that states may have obligations toward vulnerable migrants. Instead, it has been to show that the power to control borders is generally defensible, so long as states abide by applicable principles of justice in migration.¹² Hence, as the debate has proceeded, participants have sought to explicate the interests that a right to exclude grants a political community in order to test its weightiness vis-à-vis the interests of different categories of migrants—such as refugees and asylum seekers, candidates for family reunification, guest workers, past colonial subjects, racialized individuals, individuals displaced by climate change, members of the global poor, and those who want to move for reasons of cultural or religious affinity. The right to exclude, on this rendering of the debate, determines which migrants can permissibly be excluded. Following Lea Ypi's definition of justice in migration, on this understanding, the right to exclude "identifies permissible and impermissible restrictions on freedom of movement

9 Copp, "The Idea of a Legitimate State," 23–26.

10 Copp, "The Idea of a Legitimate State," 25.

11 The seminal piece in this debate is Joseph Carens' first article on immigration and border control, which makes a general case for open borders (Carens, "Aliens and Citizens"). See also Cole, *Philosophies of Exclusion*; and Kukathas, "The Case for Open Borders."

12 This is true even of Christopher Heath Wellman, who argues that states may permissibly exclude "even refugees desperately seeking asylum": as Wellman makes clear, that permission is conditional on states successfully "exporting justice" to refugees by assisting them in other ways ("Immigration and Freedom of Association," 109, 128–30).

and articulates how benefits and responsibilities should be distributed between the affected parties.”¹³ In other words, the right to exclude can easily be interpreted as a standard of *rightful exclusion*.

Yet, notice that principles of justice in migration are functionally distinct from the conditions of the legitimate imposition of border control.¹⁴ Whereas principles of justice evaluate the substantive content of norms, such as laws, principles of legitimacy settle the prior question of which agents have the standing to create and enforce such norms.¹⁵ Hence, even if we could agree that a particular immigration policy engendered a justified distribution of burdens and responsibilities, we would still require an independent account of why states are entitled to enforce that distribution by means of force. This matters for the debate at hand because immigration law is particular in that it directly targets individuals who, by definition, stand outside the state’s structures of authorization.¹⁶ Consequently, standard accounts of the state’s authority cannot explain why states are entitled to enforce border law, simply because these accounts generally locate that authority in relations between the state and those who reside on its territory.¹⁷ This is not simply a theoretical point. Legitimacy matters in practical terms because it confers crucial privileges onto its holder. The first is the privilege to interpret *which* requirements of justice apply in a given domain. This privilege matters under circumstances where there are several competing conceptions of justice in migration on offer and at least some of them qualify as reasonable. Second, legitimacy entails a presumption of reform, so that, even in the case where its rules fall short of applicable standards of justice, the addressees of the institution’s rules have weighty reason to comply with it and others

13 Ypi, “Justice in Migration,” 391.

14 Miller, “Justice in Immigration,” 392; Yong, “Immigration Rights and the Justification of Immigration Restrictions,” 463–64; Ypi, “Justice in Migration,” 408, n38.

15 Pettit, *On the People’s Terms*, 130–31.

16 This is a core insight in Arash Abizadeh’s seminal argument that border control is illegitimate unless and until potential immigrants receive *democratic* justification for exclusion. As Abizadeh writes, “justice in the liberal sense is not a sufficient condition for democratic legitimacy: a set of laws may pass the test of hypothetical justification but still lack democratic legitimacy if the laws were simply the edicts of an enlightened autocrat” (“Democratic Theory and Border Corecion,” 42). Abizadeh’s democratic illegitimacy argument turns on his claim that, because border law subjects all non-citizens and is enforced by means of coercion, it thereby coerces all non-citizens (“Democratic Theory and Border Coercion,” 57–60, and “The Scope of the All-Subjected Principle”). For the purposes of this article, I suspend judgment on this claim: my argument should be read as a further and independent legitimacy-based critique of contemporary border regimes.

17 Hidalgo, “Resistance to Unjust Immigration Law,” 460–66.

from interfering with it.¹⁸ The upshot is that if a state wields legitimate power in the enforcement of its border regime but falls short of applicable requirements of justice, then it retains the privilege to interpret the relevant requirements and decide how to sequence the relevant reforms. To put the same point in a different way, legitimacy generates *content-independent* moral reasons for subjects to comply, and for outsiders reasons not to interfere, with the institution's rules.

The question of legitimacy—what grants states the standing to enforce their interpretation of justice in migration—is thus conceptually and practically distinct from the question of what constraints justice puts on a justifiable immigration policy. A successful account of the right to exclude should, therefore, be able to explain why states retain the privilege to interpret the requirements of justice in migration and to make decisions about the sequencing of reform when their border regimes fall short of these requirements. This theoretical desideratum is especially important in light of one core characteristic of much of the philosophical literature on migration: its participants want their accounts to shed light on the contemporary politics of migration. Proponents of the right to exclude thus want their theories to vindicate privileges on the part of actual states, in the real world.¹⁹ However, even by the standards of the most restrictive accounts of the right to exclude, the abysmal track records of contemporary nation-states in the Global North at protecting the most basic human rights of refugees and asylum seekers means that their border regimes fail to qualify as just. Thus, if the legitimacy of border control is tied to a theory of justice in migration, that would undercut the moral claims of Australia, the EU Member States, and the United States to control their borders without outside interference.²⁰

2. MORALIZING THE STATE: THE PROBLEM OF APPLICABILITY

The most important defenses of the state's right to exclude appeal to moral relations that give rise to claims for self-determination. Since having the privilege to decide how to act on one's obligations of justice is a plausible interpretation of what it means to be self-determining, these accounts—if successful—would provide the needed explanation for why contemporary nation-states retain the privilege to enforce border law, even when their regimes fall short of standards

18 Buchanan, "Institutional Legitimacy," 57.

19 See, for example, Blake, *Justice, Migration and Virtue*, 1–6; Miller, *Strangers in Our Midst*, 166–74; Pevnick, *Immigration and the Constraints of Justice*, 1–18; Song, *Immigration and Democracy*, introduction.

20 David Miller has recently argued that the authority of immigration law can be derived from a Rawlsian natural duty to respect just institutions, and explicitly notes that this means that many states will simply not have such authority ("Authority and Immigration," 9, 12).

of justice in migration. However, these accounts depend on idealized conceptions of the state, which make them vulnerable to a problem of applicability: it is unclear how they can vindicate privileges for the states that inhabit the actual world.

One well-known version of this self-determination argument invokes the value of freedom of association. According to Christopher Heath Wellman, citizens associate with one another and, given the central liberal commitment to let them do so freely, they must have the right to make rules about membership.²¹ Like the members of a sports club or an orchestra have central interests in seeing their priorities and values reflected in their organizations, so do citizens have an interest in seeing their priorities and values reflected in their states. Moreover, this interest cannot be protected without granting groups the privilege to control membership rules because this is what secures control over the constitution of the relevant group. Thus, Wellman argues, freedom of association is tied to self-determination and is practically meaningless if it does not include a right to disassociate.

Another version of the self-determination argument appeals to the necessary conditions of upholding valuable cultural ties between citizens. The most sophisticated version of this argument is offered by David Miller, who argues that the function of national culture is to allow citizens to conceive of their political community as a shared project.²² This national *identity* is instrumentally valuable because it fosters the kind of social cohesion and trust that enables the achievement of progressive social justice, and intrinsically valuable because it becomes the vehicle through which collective self-determination is expressed. Since immigration necessarily changes the composition of the citizenry by introducing individuals who may not identify with the state's national culture, the state needs the privilege to control borders in order to uphold the social cohesion that national culture supplies.

A final version of the self-determination argument appeals to ownership over the state's institutions. Ryan Pevnick thus argues that, by creating and maintaining common political institutions, citizens earn ownership claims in those institutions.²³ These ownership claims grant special entitlements to have a say about what the state's institutions should look like in the future, which includes decisions about the composition of membership. Arguments that ignore this

21 Wellman, "Immigration and Freedom of Association."

22 Miller, *Strangers in Our Midst*, ch. 4. For another nationalist argument, see Walzer, *Spheres of Justice*, ch. 2.

23 Pevnick, *Immigration and the Constraints of Justice*. For another argument that appeals to ownership but with an emphasis on individual rather than collective ownership, see Simmons, *Boundaries of Authority*.

historical process, Pevnick argues, “deny the importance of self-determination” because they “unhelpfully abstract from the relationship between state institutions and the concerted collective effort that brought them into existence.”²⁴

These arguments have all faced extensive critical discussion seeking to dislodge the substantive moral grounds—freedom of association, liberal nationalism, and collective ownership—each account supplies as justification for exclusion. However, much of this substantive criticism can be redescribed as turning on an empirical or descriptive objection to how these accounts conceive of the state. Take Wellman’s freedom of association argument. Critics have pointed out that citizens in modern-day states do not associate with one another in the way that is protected by the freedom of association law to which Wellman appeals.²⁵ That law protects intimate and expressive associations, which are the kind of associations whose very purpose would be undermined if they could not control membership. Yet, the same cannot be said for citizens, who neither associate freely nor are likely to ever interact with the vast majority of their compatriots. Now consider the liberal nationalist argument. Critics of liberal nationalism have targeted the argument’s empirical component by questioning the relationship between national culture and social cohesion.²⁶ More generally, critics have pointed out that citizens of modern-day states are not members of single, unitary national cultures. Instead, they often belong to several overlapping cultures—sometimes as a result of the history of colonialism.²⁷ Lastly, critics have pointed out that the histories of colonialism, territorial annexation, and transnational exploitation entail that it is far from obvious that Pevnick’s account can support clean—and equal—ownership claims on the part of citizens in actual states.²⁸ The contributions of noncitizens to the institutions of many old states will mean that there are ownership claims outside of the state’s membership. Conversely, for young states, it will often be unclear whether there is a shared history between its citizens and their current institutions at all, since these institutions were often created by others.

24 Pevnick, *Immigration and the Constraints of Justice*, 39.

25 Blake, *Justice, Migration and Virtue*, 66; Fine, “Freedom of Association Is Not the Answer,” 349–51; Miller, *National Responsibility and Global Justice*, 211; Song, *Immigration and Democracy*, 44–45.

26 Indeed, the liberal nationalist argument has inspired interesting empirical studies, one of which has found that nationalist values *detract from* social cohesion (Breidahl, Holtug, and Kongshøj “Do Shared Values Promote Social Cohesion?”).

27 Amighetti and Nuti, “A Nation’s Right to Exclude and the Colonies”; Erez, “Liberal Nationalism, Immigration, and the Problem of Multiple National Identities”; Fine, “Migration, Political Philosophy, and the Real World”; Song, *Immigration and Democracy*, 34.

28 Blake, *Justice, Migration and Virtue*, 63–64; Kukathas, “Why Open Borders?”

These various lines of critique can be condensed into a general objection, which goes as follows. One could accept the substantive normative premises of these accounts but deny that they extend to the states that inhabit the actual world. Although they are all internally consistent, and may even supply normatively appealing grounds for exclusion, each account idealizes the subject of their theory: the state. Thus, the objection is not that they present morally implausible cases for exclusion, but rather that these cases do not apply to contemporary states. Hence, they cannot vindicate the privileges conferred by the right to exclude *for* states because their accounts turn on descriptive premises that states only fulfill partly (if at all). Thus, they face a problem of applicability: they assume predicates as a condition for the applicability of their theories that cannot be reintroduced at the level of application without thereby changing the substantive normative content of their theories.²⁹ Consequently, these theories face a steep *empirical* challenge if they are to provide a successful explanation for why the legitimacy of border control follows from state legitimacy more generally.

3. INSTITUTIONAL CONSERVATISM AND THE STATE

The common problem faced by the accounts above is that they invoke moralized and idealized conceptions of the state. That this problem is common across a range of influential theories makes a recent theoretical development especially noteworthy. According to two recent arguments by Michael Blake and Sarah Song, the right to exclude can be derived from a minimal descriptive account of the state.³⁰ More specifically, both accounts hold that the privilege to control borders follows from facts that are entailed by sharing a jurisdiction. The appeal of this methodological strategy should be obvious: if the right to exclude can be supported by an appeal to relations that uncontroversially obtain between citizens of actually existing states, then we have an explanation for why these states retain the privilege to control borders even when they fall short of applicable standards of justice in migration. In line with the arguments rehearsed above, Blake and Song both find that the existing general theories of the right to exclude are insufficiently attuned to the political nature of states.³¹ As Blake puts it in a critical appraisal of Song, “Most defenses of the right spend too little time on what a state is, *qua* state—a political community,

29 O’Neill, *Towards Justice and Virtue*, 40–41; Valentini, “On the Apparent Paradox of Ideal Theory,” 351–55.

30 Blake, “Immigration, Jurisdiction, and Exclusion”; *Justice, Migration and Virtue*; Song, “Why Do States Have the Right to Control Immigration?” and *Immigration and Democracy*.

31 Blake, *Justice, Migration and Virtue*, 56, 59, 66; Song, *Immigration and Democracy*, 31–46.

with a defined jurisdictional reach.”³² To rectify this lack, they both point to minimal conditions that unite citizens living under the same state.

For Blake, “states exist whenever there is an effective government able to exert political and legal control over a particular jurisdiction.”³³ At a minimum, therefore, the state is “a jurisdictional project, in that it is defined with reference to a particular sort of power held over a particular sort of place.”³⁴ If there is a right to exclude, then it must be grounded in what is distinctive about sharing a jurisdiction. According to Blake, this distinctiveness lies in the fact that everyone who shares a jurisdiction becomes collectively responsible for each other’s rights (71). When a new person enters that jurisdiction, they gain a claim against those who were there from before. The right to exclude is grounded, Blake argues, in the claim to be free from being so obliged. According to this (as we might call it) “unwanted obligations principle,” “we have a presumptive right to be free from others imposing obligations onto us without our consent” (74).

For Song, the right to exclude is grounded in the value of self-determination. But, unlike the theorists reviewed above, Song’s account of the “self” that holds this right to self-determination is made by reference to a “people,” which in turn is defined along the thin descriptive lines advocated by Blake. More precisely, Song’s peoples are defined by three conditions. Peoples (i) are “engaged in a common project that aims at collective self-rule,” (ii) share “a history of political participation and contestation,” and (iii) have “the capacity to establish and maintain political institutions.”³⁵ It is important to emphasize that Song explicitly distinguishes her account from what she calls the “strong statist” view, which she associates with Blake’s early work, and that holds that “the state is prior to and necessary and sufficient for a people.”³⁶ This is because she wants to open space for the possibility that peoples can exist below the state level and thus might incur rights of jurisdiction by themselves. However, while living under the same state is not a necessary condition for qualifying as a people on Song’s view, the triggering conditions of her three participation requirements are so low that it is certainly sufficient: individuals participate in the practice of peoplehood when they support a common political project, which for Song

32 Blake, “Jurisdiction and Exclusion,” 70.

33 Blake, “Immigration, Jurisdiction, and Exclusion,” 110.

34 Blake, *Justice, Migration and Virtue*, 68. From here, I will refer to Blake’s book by page numbers in the main body of text.

35 Song, *Immigration and Democracy*, 58–59.

36 Song, *Immigration and Democracy*, 57. For Blake’s “strong statism,” see his “Distributive Justice, State Coercion, and Autonomy.”

involves “observing traffic laws, responding to jury summons, paying taxes, and respecting the legal rights of others.”³⁷

For Blake and Song alike, the right to exclude is, as Song puts it, “a jurisdictional right.”³⁸ Their accounts are functionally equivalent in that they grant privileges to states, as they currently exist, by way of the state’s structuring of the relationship between citizens and residents.³⁹ There is much that can be said about the substantive normative basis of both accounts.⁴⁰ However, what is interesting for the present discussion is the structure their shared methodological strategy gives their theories. Since Blake has defended this—as he calls it—“institutionally conservative” strategy most elaborately, I will mostly focus on his arguments below. However, insofar as she relies on the same reasoning, my arguments will be equally applicable to Song. As I will try to show, institutionally conservative theories are bound to presuppose the legitimacy of border control.

As Blake describes it, the core of institutional conservatism is that the theorist takes an existing social institution as their starting point and asks what it would take for this institution to be justified.⁴¹ The reason why this institutionally conservative method is called for, he argues, is both pragmatic and normative: “if we can adjust what we have and meet the tests of justice, then we should do so—where that *should* refers both to the conceptual difficulties in building new institutional forms and to the practical difficulties engendered by revolutionary changes in institutional framework” (10). Only if our current institutions are structurally incompatible with the demands of justice do we have reason to think of new institutional forms. This seems to make institutional conservatism vulnerable to the problem of second best, according to which theories err when they assume that, when one or more conditions of optimal arrangements are missing, one should still try to approximate those arrangements instead of looking to alternatives.⁴² Against such appearance,

37 Song, *Immigration and Democracy*, 60.

38 Song, “Why Do States Have the Right to Control Immigration?” 42.

39 For a comparison between their views, see Blake, “Jurisdiction and Exclusion.”

40 In Blake’s case, the unwanted obligations principle is controversial because, by implying that it is generally wrong to bring about the triggering conditions of moral norms, the principle can seem to cut against the very purpose of moral norms (see Hidalgo, “Immigration Restrictions and the Right to Avoid Unwanted Obligations”; and Kates and Pevnick, “Immigration, Jurisdiction, and History”). For Song, the main challenge will be to explain why the *interest* of peoples to be self-determining provides content-independent moral reasons for outsiders to respect their border laws (see Blake *Justice, Migration and Virtue*, 56–60).

41 In his earlier work, Blake referred to this method simply as “institutional theory” (“Distributive Justice, State Coercion, and Autonomy,” 261–65).

42 In David Estlund’s definition, theories commit a “fallacy of approximation” when they “infer (*à la* Superset) from the value-contributing conditions of any model scenario, that

however, Blake argues that questions of migration and borders are not only especially well-suited for an institutionally conservative analysis, but that they positively demand it:

Indeed, it seems to me that the very question of immigration itself makes sense only under where this assumption holds true; if the world contained only one government, ruling over all habitable land, the concept of immigration would seem to be inapplicable. (70)

Thus, on the institutional conservative approach, political theorists should presuppose the existence of states and ask what it would require for a state to wield its current powers defensibly.

Notice that institutional conservative theory on border control resembles John Rawls's approach to questions of justice. As Aaron James has convincingly argued, the seeming tension between Rawls's egalitarian conception of social justice and his sufficientarian conception of global justice can be resolved by assigning to Rawls a particular practice-based methodology. James demonstrates that Rawls identified principles of justice with rules applying to specific practices and that, although he held that individuals held moral priority, Rawls was clear that participants to such practices could also be nations, churches, corporations, and so on.⁴³ Thus, instead of misapplying his own normative logic, Rawls refrained from extending the individualist interpretation of his Original Position to the global realm because, in that realm, the relevant subject to which principles of justice apply is relations between peoples, not individuals.⁴⁴ This practice interpretation of Rawls, James argues further, helps explain another novel feature of the Rawlsian approach to justice: it contains no optimality condition according to which some institution can always be rendered more just, for example, by becoming more equal.⁴⁵ This suggests that, on the Rawlsian view, justice functions as a moral constraint on actions taking place within predetermined practices.

Blake, in fact, identifies Rawls as an institutional conservative but without further elaboration (228n10). The practice interpretation of Rawls can explain why. On both approaches, we identify some existing practice and ask what it would require for that practice to become substantially justified. Further, as both Blake and James emphasize, invoking practices in this sense cuts across Rawls's well-known distinction between ideal and nonideal theory because the

among alternatives that lack at least some of those conditions, supersets of those subsets are better" (Estlund, *Utophobia*, 275).

43 James, "Constructing Justice for Existing Practice," 283.

44 James, "Constructing Justice for Existing Practice," 299. Compare Valentini, "Global Justice and Practice-Dependence," 405–6.

45 James, "Constructing Justice for Existing Practice," 294.

theoretical subject can be described in more or less idealized terms.⁴⁶ The point is to draw on moral values to give a rational reconstruction of the subject of our theory and—if that subject is not fundamentally incompatible with those values—thereby provide a normative benchmark against which we can identify permissible and impermissible actions within our currently existing practice. Hence, the commitment to reform of existing institutions over institutional innovation. Institutional conservatism is a form of the “practice-dependent” methodology James assigns to Rawls.⁴⁷

The constraint interpretation of Rawlsian principles of justice, which follows from his practice dependence, coheres well with many of Blake’s substantial normative claims. Take, for example, Blake’s argument that states cannot permissibly exclude refugees: on his view, the right to exclude places constraints on the kind of policies states may permissibly adopt when they set their immigration policies.⁴⁸ This explains the following passage:

Even if legitimate states have a right to exclude unwanted would-be immigrants, much work needs to be done to figure out . . . the contours of that right. It is possible to have the right to exclude, after all, and still question whether or not that right is able to ground a particular exclusionary policy. (79)

Here, the term “right” is clearly invoked in the sense of “rightfulness” that is associated with justice. However, the cited passage also raises a question: Can institutionally conservative theories like Blake’s establish the sense of “right” that is associated with legitimacy—namely, of the privilege to retain control over borders when obligations of justice in migration are not met? Or will the legitimacy of border control simply have to be assumed to follow from state legitimacy more generally? Despite suggestions to the contrary, Blake ultimately ends up having to stipulate that border control follows from legitimacy more generally. This becomes explicit in his discussion of irregular migration:

46 Blake, “Distributive Justice, State Coercion, and Autonomy,” 263n7; James, “Constructing Justice for Existing Practice,” 291–93.

47 On practice dependence as a meta-normative position about the grounding of principles of political morality, see Sangiovanni, “Justice and the Priority of Politics to Morality”; and James, *Fairness in Practice*, ch. 4.

48 The constraint interpretation also coheres well with the other part of Blake’s overall project, which is to make conceptual space for evaluating immigration policy in terms of whether it is *merciful* (chapter 10). The central thought is that, even if an immigration policy is just, we might still find it objectionable because it is *unmerciful* (213). Thus, on Blake’s final view, justice places baseline constraints on morally *acceptable* policy, whereas mercy places further constraints on morally *virtuous* policy. The scope of the latter is narrower than the scope of the former.

I will assume that the state trying to exclude is the sort of state to issue authoritative commands; however we identify the line of decency below which states lose the right to expect obligation, our state is above that line. . . . I will not assume . . . that the law mandating exclusion is itself just; I will merely assume that the state is the sort of entity that might create an obligation to obey the law, regardless of whether or not that law is just. (167)

Although Blake's defense of the right to exclude does not assume that immigration restrictions are just, it does assume that states wield legitimate power in the enforcement of their border regimes. It thus fails to provide an *explanation* for why states should be entitled to regulate migration even when they fail to discharge the applicable principles of justice in migration.

This problem is not, I argue, local to Blake's particular theory. It is a structural problem that derives from his reliance on institutional conservatism.⁴⁹ Employing the language of practice dependence, we can see why. For any political theory, we can ask: Does the theory aim to identify principles for regulating a given practice; to justify that practice; or, more fundamentally, to identify an existing practice as subject for justification and/or regulation? This systematization is useful because it allows us to see how theories that, in a Rawlsian mode, seek to answer the first of these questions do not contain the resources for justifying the practices that they take as their starting point. The principles they generate are importantly limited because, as James points out, if "a principle applies to the world only insofar as an appropriate kind of social practice exists, that principle cannot itself be used to criticize either the existence or the non-existence of the kind of practice that conditions its application."⁵⁰ The problem with institutionally conservative approaches to the right to exclude—as with all approaches that analyze that right only in terms of justice in migration—is that they treat their theories as *also* responding to the second question when they only contain the resources for responding to the first.

Therefore, institutionally conservative approaches to the right to exclude do not explain why the states that inhabit the actual world wield legitimate power in the enforcement of their border control. Arguments for legitimacy must explain why subjects of rules have content-independent moral reasons

49 A parallel argument can be made against Song on the grounds of her stipulation of territorial rights (*Immigration and Democracy*, 61–65). As Blake puts it in his critical discussion, Song's argument can explain why it is *good* for individuals who share a territorial jurisdiction to hold unilateral control over their borders, but it cannot explain why they have a *right* ("Jurisdiction and Exclusion," 74–75; *Justice, Migration and Virtue*, 57–60).

50 James, "Constructing Justice for Existing Practice," 313.

for compliance, and outsiders have reasons for noninterference, with the actual institutions that seek to regulate their actions.⁵¹ Constructing an argument to the effect that these institutions *could* be rendered compatible with the demands of justice if they were organized differently provides an unconvincing answer to that question. Thus, institutional conservatism turns out to lack a key function that its proponents need their theories to perform: to explain why states are entitled to enforce border control when their border regimes fall short of the requirements of justice in migration.

Two important objections to my argument can be made on the part of institutional conservatives. The first is to argue that one can draw on the resources of institutional conservatism to evaluate the practice of border control itself, as distinct from the rules that ought to regulate that practice. The second is to argue that considerations of feasibility justify the institutionally conservative stance. In the next section, I tackle both objections by arguing that, although an evaluation of border control is perfectly coherent on a practice-based methodology, it cannot plausibly deliver the strict unilateral kind of control over borders supported by Blake and Song.

4. PRACTICE DEPENDENCE AND STATE-SYSTEM LEGITIMACY

To summarize the argument thus far, I have shown that institutional conservatism is an appealing approach to the ethics of migration because it straightforwardly avoids a core problem facing approaches that moralize the state. However, I have argued that institutional conservatism cannot yield a plausible account of the legitimacy of border control, as distinct from a theory of justice in migration, because it cannot explain why states retain the privilege to control borders when their regimes fall short of applicable requirements of justice. To make this argument, I argued that institutional conservatism is a form of the practice-dependent method developed by James and others. This section draws on the resources of practice dependence with the aim of demonstrating that a reconstructed version of institutional conservatism that seeks to evaluate the practice of border control itself, while perfectly coherent, cannot yield strict unilateralism over borders.

Having read my argument above, a reader might wonder why those who are sympathetic to institutional conservatism could not draw on the resources of practice dependence to offer a defense of border control. On this view, the only mistake institutional conservative writing on migration has made is to conduct their discussions at the wrong level of analysis. Instead of evaluating the practices

51 Buchanan, "Institutional Legitimacy."

of each individual state, so this argument goes, one could vindicate the privileges held by states by engaging in an analysis of the “practice of statehood.” Hence, if we gave a rational reconstruction of the *state system* and convincingly showed that unilateral border control would be part of such a system, then we would seemingly have an account that could confer legitimacy onto border control. Thus, is this a viable strategy for vindicating the legitimacy of border control?

Although it is perfectly possible to offer a rational reconstruction of the state system along the lines envisaged by the practice-dependent methodology, it is unlikely to deliver the substantial conclusion that institutional conservatives are after. Indeed, it has become increasingly popular to draw on precisely this kind of argument to show that strict territorial sovereignty as it currently exists is indefensible.⁵² The reasoning behind these accounts is straightforward and based on a widely shared normative premise: that the function of the state system is to determine who is responsible for protecting whose human rights. This premise is institutionally enshrined in international law and forms the basis of international refugee law, where it is recognized that states are obligated to assist individuals whose states are no longer willing or able to protect their rights from persecution.⁵³ To recall, the premise also forms the basis of Blake’s substantive normative account, since the fact that the state system allocates responsibility for human rights protection is precisely the reason why a person who enters a new jurisdiction imposes obligations on those already present in that jurisdiction (36). It is also accepted by Song, who argues that the core legitimacy criterion for states is the maintenance of order and protection of human rights of inhabitants.⁵⁴

However, whereas Blake goes on to argue that the fact that the state system assigns human rights protection through territorial presence can ground the right to exclude (by way of the right to *avoid* being responsible for human rights), a group of authors has recently drawn on the logic of this way of assigning responsibility for human rights to argue for a different conclusion. Gillian Brock and David Owen thus argue that, given the state system’s normative core function of allocating responsibility for human rights, the justification of the system itself turns on successful human rights protection.⁵⁵ In other words, when individuals fail to have their human rights protected, the state system falls short of its purpose. These “state system legitimacy” theories, as Daniel

52 See, for example, Pavel, *Divided Sovereignty*.

53 Criddle and Fox-Decent, “The Authority of International Refugee Law.”

54 Song, *Immigration and Democracy*, 55, 62. Song also invokes the human rights-based logic of the state system when she points to international human rights law to motivate her case for collective self-determination (*Immigration and Democracy*, 53). See also Song, “Political Theories of Migration,” 395.

55 Brock, *Justice for People on the Move*; Owen, *What Do We Owe to Refugees?*

Sharp has usefully labeled them, then point out that this insight has important implications when it comes to migration.⁵⁶ If the purpose of the state system is to ensure that each individual has some political authority that is in charge for the protection of her basic rights, then emigration becomes an important remedy for individuals when the political authority that is currently responsible for those rights fails to discharge its responsibility.

Based on this link between the function of the state system and migration, state system legitimacy theories make the following argument.⁵⁷ States, as a matter of moral principle, depend for their privileges on a state system that is premised on protecting human rights. When states fail their task of protecting the rights of those for whom they have been allocated responsibility, this creates a problem for the system as a whole and, by extension, for the individual states that depend for their privileges on the justifiability of the state system. This legitimacy problem can be solved by erecting effective institutions that ensure that those who require protection through migration receive that protection—hence, Owen’s description of the refugee regime as a “legitimacy repair mechanism.”⁵⁸ This means, Brock and Owen argue, that the various privileges associated with sovereignty—including and especially control over borders—is conditional on the creation and maintenance of supranational institutions for securing migrants’ rights.⁵⁹

Sharp has recently provided a different argument in support of Brock’s and Owen’s conclusion.⁶⁰ On his view, the erection of supranational institutions is not required on human rights grounds, but to rectify the problematic discrepancy in power between citizens of wealthy states and individuals who could significantly better their lot by immigrating to them. This power could be rectified, Sharp argues, if these individuals were granted access to institutions that would give them a say in the border policies of wealthy states in the Global North. Such procedural inclusion would be desirable because it would help ameliorate the

56 Sharp, “Immigration and State System Legitimacy.” Sharp also attaches the “state system legitimacy theory” label to Christopher Bertram. However, as Sharp emphasizes, although Bertram also invokes the concept of legitimacy to discuss border control and ends up endorsing conclusions similar to Brock and Owen, his account is distinct from theirs in that he does not rely on the reconstruction of the state system (Sharp, “Immigration and State System Legitimacy,” 6). Instead, Bertram identifies the legitimacy problem in contemporary border controls with the power each state claims over each potential immigrant (*Do States Have the Right to Control Immigration?* 52–53).

57 Brock, *Justice for People on the Move*, 38–40; Owen, *What Do We Owe to Refugees?* 45–47; Sharp, “Immigration and State System Legitimacy,” 3–6.

58 Owen, “In Loco Civitatis,” 275, and *What Do We Owe to Refugees?* 47.

59 Brock, *Justice for People on the Move*, 226–27; Owen, *What Do We Owe to Refugees?* 107.

60 Sharp, “Relational Equality and Immigration,” 673–75.

problematic relationship of inequality that exists between the rich and poor and because it would, by ensuring ongoing exit options, ameliorate the problematic relationship the latter group stands in with their current states of residence.⁶¹

These arguments in favor of the necessity of supranational institutions for regulating migration point in the opposite direction than the one that institutional conservatives need them to. Far from providing a straightforward justification for the sovereign privileges states currently enjoy, these arguments entail that a rational reconstruction of the state system would include institutional mechanisms that would guarantee the possibility of migration in the face of a state that fails to discharge its responsibilities toward those who reside within its jurisdiction. Yet, this guarantee would necessarily revoke the privilege to unilaterally decide on all immigration policy. For at least some vulnerable migrants, political communities might be *made* to take responsibility for their human rights by the institutions regulating global migration. Thus, while the practice-based methodology can certainly be used to evaluate border control, it is not clear that it can support the legitimacy of full unilateral control over borders. Quite to the contrary, these sophisticated arguments point in a different direction: for at least some immigration decisions, the interpretation of applicable principles of justice, and the sequence of the reform to which they give rise, would be determined by a supranational institution.

Institutional conservatives can object to this line of argument by pointing to the infeasibility of moving away from the state system. According to this objection, I have misunderstood the main point of institutional conservatism—namely, that it takes seriously that there are “conceptual and practical” difficulties associated with creating new institutional forms (9–10).⁶² Thus, given these difficulties, it will always be better to prioritize the reform of states than to advocate the delegation of parts of their competences to new institutions above (or below) the state level. This appeal to feasibility is unsatisfactory as it stands, however. This is because it is vulnerable to the following countercharge. Since institutional conservatives accept that human rights are a baseline constraint on state actions, and that this gives rise to obligations toward

61 Sharp, “Relational Equality and Immigration,” 669. This procedural proposal is distinct from Abizadeh’s because it does not imply a need for a global democratic constituency: only members of the global poor would be entitled to a say and only in the border policies of wealthy states (Sharp, “Relational Equality and Immigration,” 674). Compare Abizadeh, “Democratic Legitimacy and Border Coercion.”

62 It is worth noting that, when making this argument, Blake only considers the creation of a world state as the alternative to the current state system (228n30). As the argument above demonstrates, however, a world state is not the only relevant alternative to our current arrangements.

refugees and asylum seekers, the appeal to feasibility cannot simply be that it is easier to keep things as they are: that position would amount to a simple endorsement of the status quo. Instead, it must be that it is *more* feasible to secure human rights protection under the current state system than under alternative arrangements.

However, the claim that it would be more feasible to secure human rights protection under our current system of unilaterally enforced borders than under some alternative system involving supranational institutions is a controversial claim that requires justification. Its controversy stems from the incentive structure inherent to our current system. In particular, the fact that governments have far more to gain from pandering to anti-immigration sentiments in their own populations than they have from taking steps to protect refugees and asylum seekers, entails that a system of unilateral control over borders licenses short-sighted and poorly informed decisions in the regulation of global migration. For example, Brock reviews empirical literature that suggests that citizens in Western democracies hold beliefs that immigration is bad for citizens' economic prospects; that immigrants have negative effects on public finances; and that immigrants increase crime.⁶³ These beliefs are, at best, contingent on the policy choices that receiving states make and, in many cases, they are straightforwardly false.⁶⁴ In conjunction with the structural bias that governments have toward their own citizens, these contingent or false beliefs create strict immigration regimes that predictably exclude migrants with justified claims for protection under international law.⁶⁵

One institutional expression of the perverse incentives engendered by our current system is the externalization of border control. Over the past decades, states in the Global North have made a decisive shift in how they conceive of the practice of border control. Instead of a mere concern with controlling their territorial borders, states are increasingly preoccupied with controlling "migration flows" heading toward their territories.⁶⁶ Thus, in

63 Brock, *Justice for People on the Move*, 204.

64 Brock, *Justice for People on the Move*, 205.

65 Brock and Hidalgo use these premises in an *epistemic* argument for the supranational regulation of migration (Brock, *Justice for People on the Move*, 202–9; Hidalgo, "The Case for the International Governance of Immigration," 144–52). As Hidalgo puts it, "if an agent has biases that impair this agent's capacity to make morally risky decisions in reliable ways, this agent has moral reason to transfer decision-making authority to a more reliable party" ("The Case for the International Governance of Immigration," 152). What matters for my purposes, however, is the more minimal claim that these structural biases make reliable human rights protection less feasible than under a different system.

66 FitzGerald, *Refuge beyond Reach*; Gammeltoft-Hansen, *Access to Asylum*; Longo, *The Politics of Borders*; Shachar, *The Shifting Border*.

addition to exercising power *at* their territorial borders, states are employing a host of techniques to exercise power over potential immigrants far beyond their physical territory. Important measures include the imposition of carrier sanctions and of uniform visas, the administration of which is often externalized to private contractors.⁶⁷ In addition, states create so-called migration compacts, where poorer third countries agree to host asylum seekers and to police known routes of emigration in exchange for political and financial benefits. Lastly, states create and uphold deep information-sharing networks that let them track migrants even as they move far beyond their own borders. The EU's "Integrated Border Management" strategy, for example, explicitly aims to track "the movement of third-country nationals from the point of departure in countries of origin, all throughout transit, and up to their arrival in the EU"—and to deploy the measures above to deter the movement of those migrants at each possible juncture.⁶⁸

The main reason that states choose to externalize their border regimes in these ways is grounded in the norms of territorial sovereignty. Since responsibility for human rights protection is triggered by territorial presence, the externalization of border-controlling measures allows wealthy states to wield power over vulnerable migrants without incurring the responsibilities they otherwise would incur if they had exercised that power within their own jurisdictions. It has been robustly demonstrated in legal and empirical literatures that these practices create forms of overlapping power and functional jurisdiction that lead to responsibility gaps.⁶⁹ Practically, it is often unclear who is responsible for whose rights in this new bordering landscape—is it the state on whose orders migrants are constrained or detained? Or is it the state on whose territory the constraining and detaining occurs? The overall effect has been to weaken protection for vulnerable migrants and to create an effect where these migrants are pushed to take increasingly dangerous routes to reach a state that can reliably protect their rights so that they can launch an asylum claim once territorially present. The sociologist David Scott FitzGerald describes this as the "catch-22" of modern migration control: rich states respect the principle of *non-refoulement*, which means that those who reach their physical territories

67 To his credit, Blake offers an interesting discussion of carrier sanctions and makes a convincing argument that the practice often involves violation of a positive duty to assist needy migrants (102). However, he also argues that the practice is generally permissible because it fails to be coercive and does not consider its broader implications for reliable human rights protection.

68 Moreno-Lax, *Accessing Asylum in Europe*, 3.

69 Gammeltoft-Hansen, *Access to Asylum*; Moreno-Lax, *Accessing Asylum in Europe*; Shachar, *The Shifting Border*.

will be able to depend on their protection, but at the same time create conditions where it is exceedingly dangerous to reach their territories.⁷⁰

These practices, and the incentives underlying them, are fully explained by our current norms of strict territorial sovereignty. Moreover, their effect on human rights protection is clear.⁷¹ Therefore, I argue, the feasibility calculation does not pull in the favor of the preferred conclusions of institutional conservatives. Rather, a rational reconstruction of the state system would plausibly entail that states should be stripped of their privilege to *full* unilateral discretion over immigration policies: for at least some groups of migrants, the privilege to interpret the requirements of justice in migration and to sequence reforms should be wielded by supranational institutions. In conclusion, then, it is perfectly possible to construct a practice-dependent evaluation of the state system, but it does not provide a plausible defense of the privileges associated with legitimacy.

5. CONCLUDING REMARKS

Where does this leave us? The first conclusion that can be drawn from the above is that, for all its strengths as a methodology for constructing theories of justice in migration, institutional conservatives seem forced to scale back their ambitions to vindicate the privileges claimed by the states in the actual world. One possibility for institutional conservatives would, of course, be to tie the legitimacy of border control to their preferred standards of justice in migration and argue that all border regimes that fall short of that standard thereby fail to generate reasons for compliance and noninterference. However, this is not a palatable option for institutional conservatives, since it would undercut their ambition of vindicating privileges on the part of the states that inhabit the actual world since most of these regimes fall short of even very permissive standards of justice in migration. This is especially true for the states that theorists in the analytical tradition, either implicitly or explicitly, assume as their subjects of concern: wealthy democracies in the Global North.

The more general lesson I think can be drawn from my discussion above is methodological. The ethics of migration has long been characterized by a *methodologically nationalist* approach, which centers individual states in their normative analyses and asks what makes the claims of individuals who live within them weighty enough to ground rights to exclude outsiders.⁷² But since this

70 FitzGerald, *Refuge beyond Reach*, 10.

71 For a comprehensive normative critique of these developments in contemporary border control, see my “The Practice and Legitimacy of Border Control.”

72 Sager, “Methodological Nationalism, Migration, and Political Theory.”

approach has led theorists to either moralize the state or stipulate the legitimacy of its claim to regulate immigration, a more fruitful analysis of the ethics and politics of migration should take systemic effects into account. In particular, I have sought to offer further support for the view that normative accounts would do well to think about the state system and the incentives (not) to protect vulnerable migrants that it engenders. Engaging in such an analysis does not run counter to Copp's claim that the concept of legitimate statehood invoked by political theorists and philosophers—and scholars in other disciplines—has included control over borders. However, it does entail that the sovereign privileges entailed by such a concept of legitimate statehood cannot be evaluated as one bundle. Instead, we should disaggregate the analysis of the distinct claims to authority that states make over their citizens, and over potential immigrants.⁷³

University of Oslo
h.h.sandven@stv.uio.no

REFERENCES

- Abizadeh, Arash. "Democratic Theory and Border Coercion: No Right to Unilaterally Control Your Own Borders." *Political Theory* 36, no. 1 (February 2008): 37–65.
- . "The Scope of the All-Subjected Principle: On the Logical Structure of Coercive Laws." *Analysis* 81, no. 4 (October 2021): 603–10.
- Amighetti, Sara, and Alasia Nuti. "A Nation's Right to Exclude and the Colonies." *Political Theory* 44, no. 4 (August 2016): 541–66.
- Bertram, Christopher. *Do States Have the Right to Exclude Immigrants?* Cambridge: Polity, 2018.
- Blake, Michael. "Distributive Justice, State Coercion, and Autonomy." *Philosophy and Public Affairs* 30, no. 3 (Summer 2001): 257–96.
- . "Immigration, Jurisdiction, and Exclusion." *Philosophy and Public Affairs* 41, no. 2 (Spring 2013): 103–30.

73 For comments and discussion that much improved the article, I am grateful to Shai Agmon, Rufaida Al Hashmi, Rebecca Buxton, Jamie Draper, Henrik Kugelberg, Cécile Laborde, Alejandra Mancilla, David Miller, Laura Valentini, and one anonymous reviewer for this *Journal*. I owe special thanks to Michael Blake, who revealed his identity as the second reviewer, and who showed great intellectual generosity toward an article with which he probably still disagrees. Finally, I am grateful to Lior Erez, who in his editorial capacity for a different journal encouraged me to expand on what was originally conceived as a shorter piece.

- . “Jurisdiction and Exclusion: A Response to Sarah Song.” In *NOMOS LVII: Immigration, Emigration, and Migration*, edited by Jack Knight, 69–76. New York: NYU Press, 2017.
- . *Justice, Migration, and Mercy*. Oxford: Oxford University Press, 2020.
- Breidahl, Karen N., Nils Holtug, and Kristian Kongshøj. “Do Shared Values Promote Social Cohesion? If So, Which? Evidence From Denmark.” *European Political Science Review* 10, no. 1 (February 2018): 97–118.
- Brock, Gillian. *Justice for People on the Move: Migration in Challenging Times*. Cambridge: Cambridge University Press, 2020.
- Buchanan, Allen. “Institutional Legitimacy.” In *Oxford Studies in Political Philosophy*, vol. 4, edited by David Sobel, Peter Vallentyne, and Steven Wall, 53–78. Oxford: Oxford University Press, 2018.
- Cole, Philip. *Philosophies of Exclusion: Liberal Political Theory and Immigration*. Edinburgh: Edinburgh University Press, 2000.
- Copp, David. “The Idea of a Legitimate State.” *Philosophy and Public Affairs* 28, no. 1 (Winter 1999): 3–45.
- Criddle, Evan J., and Evan Fox-Decent. “The Authority of International Refugee Law.” *William and Mary Law Review* 62, no. 4 (March 2021): 1067–1136.
- Erez, Lior. “Liberal Nationalism, Immigration, and the Problem of Multiple National Identities.” *Critical Review of International Social and Political Philosophy* 23, no. 4 (2020): 495–517.
- Estlund, David. *Utopophobia: On the Limits (if Any) of Political Philosophy*. Princeton: Princeton University Press, 2020.
- Fine, Sarah. “The Ethics of Immigration: Self-Determination and the Right to Exclude.” *Philosophy Compass* 8, no. 3 (March 2013): 254–68.
- . “Freedom of Association Is Not the Answer.” *Ethics* 120, no. 2 (January 2010): 338–56.
- . “Migration, Political Philosophy, and the Real World.” *Critical Review of International Social and Political Philosophy* 20, no. 6 (2017): 719–25.
- FitzGerald, David Scott. *Refuge beyond Reach: How Rich Democracies Repel Asylum Seekers*. Oxford: Oxford University Press, 2019.
- Gammeltoft-Hansen, Thomas. *Access to Asylum: International Refugee Law and the Globalisation of Migration Control*. Cambridge: Cambridge University Press, 2011.
- Hidalgo, Javier. “The Case for the International Governance of Immigration.” *International Theory* 8, no. 1 (March 2016): 140–70.
- . “Immigration Restrictions and the Right to Avoid Unwanted Obligations.” *Journal of Ethics and Social Philosophy* 8, no. 2 (2014): 1–9.
- . “Resistance to Unjust Immigration Law.” *Journal of Political Philosophy* 23, no. 4 (December 2015): 450–70.

- Kates, Michael, and Ryan Pevnick. "Immigration, Jurisdiction, and History." *Philosophy and Public Affairs* 42, no. 2 (Spring 2014): 179–94.
- Kukathas, Chandran. "The Case for Open Borders." In *Contemporary Debates in Applied Ethics*, edited by Andrew I. Cohen and Christopher Heath Wellman, 207–20. Oxford: Blackwell, 2005.
- . "Why Open Borders?" *Ethical Perspectives* 19, no. 4 (December 2012): 650–75.
- Longo, Matthew. *The Politics of Borders: Sovereignty, Security, and the Citizen after 9/11*. Cambridge: Cambridge University Press, 2018.
- James, Aaron. "Constructing Justice for Existing Practice: Rawls and the Status Quo." *Philosophy and Public Affairs* 33, no. 3 (July 2005): 281–316.
- . *Fairness in Practice: A Social Contract for a Global Economy*. New York: Oxford University Press, 2012.
- Miller, David. "Authority and Immigration." *Political Studies* (forthcoming). Published online ahead of print, September 20, 2021. <https://doi.org/10.1177/003232172111046423>.
- . "Justice in Immigration." *European Journal of Political Theory* 14, no. 4 (October 2015): 391–408.
- . *National Responsibility and Global Justice*. Oxford: Oxford University Press, 2007.
- . *Strangers in Our Midst: The Political Philosophy of Immigration*. Cambridge, MA: Harvard University Press, 2016.
- Moreno-Lax, Violeta. *Accessing Asylum in Europe: Extraterritorial Border Control and Refugee Rights under EU Law*. Oxford: Oxford University Press, 2017.
- O'Neill, Onora. *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. Cambridge: Cambridge University Press, 1996.
- Owen, David. "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 Lea Ypi, 269–89. Oxford: Oxford University Press, 2016.
- . *What Do We Owe to Refugees?* Cambridge: Polity, 2020.
- Pavel, Carmen. *Divided Sovereignty: International Institutions and the Limits of State Authority*. Oxford: Oxford University Press, 2014.
- Pettit, Philip. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press, 2012.
- Pevnick, Ryan. *Immigration and the Constraints of Justice: Between Open Borders and Absolute Sovereignty*. Cambridge: Cambridge University Press, 2011.
- Sager, Alex. "Methodological Nationalism, Migration, and Political Theory." *Political Studies* 64, no. 1 (March 2016): 42–59.
- Sandven, Hallvard. "The Practice and Legitimacy of Border Control." *American*

- Journal of Political Science* (forthcoming). Published online ahead of print, August 28, 2022. <https://doi.org/10.1111/ajps.12736>.
- Sangiovanni, Andrea. "Justice and the Priority of Politics to Morality." *Journal of Political Philosophy* 16, no. 2 (June 2008): 137–64.
- Shachar, Ayelet. *The Shifting Border: Ayelet Shachar in Dialogue*, edited by Peter Nielsen. Manchester: Manchester University Press, 2020.
- Sharp, Daniel. "Immigration and State System Legitimacy." *Critical Review of International Social and Political Philosophy* (forthcoming). Published online ahead of print, December 8, 2020. <https://doi.org/10.1080/13698230.2020.1860416>.
- . "Relational Equality and Immigration." *Ethics* 132, no. 3 (April 2022): 644–79.
- Simmons, A. John. *Boundaries of Authority*. Oxford: Oxford University Press, 2016.
- Song, Sarah. *Immigration and Democracy*. Oxford: Oxford University Press, 2019.
- . "Political Theories of Migration." *Annual Review of Political Science* 21, no. 1 (May 2018): 385–402.
- . "Why Do States Have the Right to Control Immigration?" In *NOMOS LVII: Immigration, Emigration, and Migration*, edited by Jack Knight, 3–50. New York: NYU Press, 2017.
- Valentini, Laura. "Global Justice and Practice-Dependence: Conventionalism, Institutionalism, Functionalism." *Journal of Political Philosophy* 19, no. 4 (December 2011): 399–418.
- . "On the Apparent Paradox of Ideal Theory." *Journal of Political Philosophy* 17, no. 3 (September 2009): 332–55.
- Walzer, Michael. *Spheres of Justice: A Defence of Pluralism and Equality*. New York: Basic Books, 1983.
- Wellman, Christopher Heath. "Immigration and Freedom of Association." *Ethics* 119, no. 1 (October 2008): 109–41.
- Wenar, Leif. "The Nature of Rights." *Philosophy and Public Affairs* 33, no. 3 (Summer 2005): 223–52.
- Yong, Caleb. "Immigration Rights and the Justification of Immigration Restrictions." *Journal of Social Philosophy* 48, no. 4 (Winter 2017): 461–80.
- Ypi, Lea. "Justice in Migration: A Closed Borders Utopia?" *Journal of Political Philosophy* 16, no. 4 (December 2008): 391–417.

FREEDOM, DESIRE, AND NECESSITY

AUTONOMOUS ACTIVITY AS ACTIVITY FOR ITS OWN SAKE

Pascal Brixel

IN THE *Nicomachean Ethics*, Aristotle takes seriously only three candidates for the good human life: the life of pleasure, the life of politics, and the life of contemplation. These are the serious contenders because each of them has some claim to be not merely instrumentally good but desirable for its own sake:

If ... there is some end of the things we do, which we desire for its own sake (everything else being desired for the sake of this), and if we do not choose everything for the sake of something else (for at that rate the process would go on to infinity, so that our desire would be empty and vain), clearly this must be the good and the chief good.¹

Hannah Arendt argues that Aristotle is here guided not just by a view of the good but by a view of freedom.² For he rejects other ways of life, such as “the life of money-making,” on the basis that they are “undertaken under compulsion,” wealth in particular being “merely useful and for the sake of something else.”³ Elsewhere, he says that wealth acquisition “may be studied by a free man, but will only be practiced from necessity,” and that the occupations it encompasses are “the most servile” arts.⁴ And he is similarly dismissive of the life of labor aiming at the immediate sustenance and reproduction of life itself—cooking, eating, procreating, cleaning, etc. In the *Ethics* he simply ignores it, and in the *Politics* it appears as the life proper to slaves and animals: “both with their bodies administer to the needs of life.”⁵

The compulsion that Aristotle takes such activities to involve is not involuntariness in some inner sense, such as unawareness of what one is doing, as in

1 Aristotle, *Nicomachean Ethics*, 1094a.

2 Arendt, *The Human Condition*.

3 Aristotle, *Nicomachean Ethics*, 1096a.

4 Aristotle, *Politics*, 1258b.

5 Aristotle, *Politics*, 1254b.

sleepwalking, or capitulation to overpowering psychological forces, as perhaps in some addictions or phobias. Nor, notwithstanding the reference to slavery, is it fundamentally a matter of subjection to interpersonal coercion. It is rather the compulsion that consists in having to do something merely as a means—out of mere necessity for an end external to the activity. A truly free activity, on this view, is one chosen for its own sake and not merely for the sake of further ends.

A version of Aristotle's idea appears again in the work of Marx. In volume 3 of *Capital*, Marx famously identifies "the true realm of human freedom" with "the development of human powers as an end in itself," which he opposes to "the realm of necessity," or of "labor determined by necessity and external expediency."⁶ Marx takes a more expansive view of the human good than Aristotle: the highest good consists no longer in pure contemplation but in the development and exercise of the totality of a human being's powers. But Marx shares Aristotle's sense that a truly free activity cannot proceed from mere instrumental necessity. Free activity must be, at a minimum, activity for its own sake.

Despite its pedigree, this Aristotelian-Marxian idea about freedom has not found its way into mainstream discussions of freedom and autonomy in contemporary analytic philosophy, even as an object of criticism. Perhaps this is because neither Aristotle nor Marx ever developed or defended the idea in a systematic, comprehensive, and fully explicit way. Be that as it may, I will argue in this paper that it constitutes a serious contribution to thought about autonomy and promises to solve important and abiding problems in the contemporary literature on that topic.

Autonomy—freedom, in one traditional sense of the word—is self-determination.⁷ To act autonomously is to be determined by one's own will rather than by alien compulsions. Since such compulsions can take multiple forms, autonomy has multiple dimensions. In its internal dimension, autonomy is opposed to the psychological unfreedom that stems from phobias and addictions, maladaptive preferences, brainwashing, and the like. In its external

6 Marx, *Capital*, vol. 3, 958–59.

7 Throughout the paper, when I speak without further specification of "autonomy," I mean "local" rather than "global" autonomy. That is, I am interested in what it is for a particular activity to be autonomous, rather than what it is for a person to be autonomous over the course of their whole life. Put another way, I am interested in what it is for someone to *do something autonomously*, rather than what it is for someone to be autonomous *simpliciter*. Note also that I will have nothing to say about various important adjacent issues, such as the role of autonomy in a specifically political ideal, or whether there is a right to autonomy, or under what conditions we can hold people morally or legally responsible for their doings. Approaches to autonomy that are motivated primarily by concerns about rights and responsibility are thus outside the scope of this paper. Perhaps it will turn out that little progress can be made if we put these concerns aside. But let us see.

dimension, it is opposed to the unfreedom associated paradigmatically with coercion by other people and duress by unfavorable circumstances.

My concern here is with autonomy in its external dimension.⁸ Our ordinary thought about this dimension is structured by two conceptual oppositions. First, as mentioned, autonomy is opposed to external unfreedom; that is, both coercion and duress compromise autonomy. Second, whereas to act autonomously is in some sense to do what one wants or desires, to do something under coercion or duress is in some sense to be forced or necessitated to do it. An adequate theory of external autonomy, I argue, ought to make sense of both of these platitudes. It ought to tell us why coercion and duress compromise autonomy, and it ought to identify the relevant senses of desire and necessity that structure our ordinary thought about these matters.

In the first half of the paper, I defend these theoretical desiderata and show that standard contemporary accounts of autonomy struggle to deliver the requisite explanations. I then defend an alternative necessary condition of autonomous activity along Aristotelian-Marxian lines. One does something autonomously, I argue, only if one does it for its own sake and not *merely* for the sake of a further end. This idea turns out to steer an attractive middle path between two dominant families of contemporary theories of autonomy. In this way, I argue, it puts us in a position to explain how coercion and duress compromise autonomy, and to articulate the sense in which to act autonomously is to act under the guise of desire, while to act unfreely is to act under the guise of necessity.

1. AUTONOMY AND EXTERNAL UNFREEDOM

I begin in this section by identifying two pieces of common sense that a complete theory of autonomy should try to respect and indeed explain. First, coercion and duress can compromise autonomy. Second, while autonomous activity is, at a minimum, a matter of doing what one really or truly *wants* to do, victims of coercion or duress are *forced* to act as they do—where the relevant kinds of desire and necessitation are conceptually opposed to each other.

1.1. Coercion and Duress

If autonomy is self-determination, it is hard to deny that coercion by another person compromises autonomy. When I speak of coercion, I have in mind not the direct use of force against someone's body but rather what is sometimes

8 I will thus put aside, for instance, the question of how to distinguish those kinds of internal compulsion that make us unfree from those that do not. An example of the latter might be the sort of inner compulsion from which Martin Luther is said to have acted in declaring, "Here I stand; I can do no other."

called “volitional” coercion, which operates by means of threats. Think of the slave who does her master’s bidding to avoid a beating, or the mugging victim who hands over her money to avoid being shot. Action of this kind is a paradigm case of heteronomy, determination not by one’s own will but by an alien compulsion, in the form of another’s will.

Though coercion is certainly the most obvious and egregious type of external unfreedom, it is very natural to think that autonomy can also be compromised by duress. When I speak of duress, I have in mind compulsion by unfavorable circumstances that force an agent to choose the lesser of two evils, where the relevant circumstances do not result from another’s deliberate machinations. Think of a sailor who throws goods overboard in a storm to save herself and her ship, or someone who abandons her house to escape a fire. Like (volitionally) coerced actions, these actions are intentional. But like coerced actions, they fall short of unqualified self-determination, though the alien compulsion in this case originates not in another will but in the agent’s exigent circumstances.

No doubt there are deep and important differences between interpersonal coercion and impersonal duress, with respect to both their nature and their moral implications. No doubt our interest in (being free of) each is motivated to some extent by concerns distinctive to each. But equally, there certainly appears to be a generic unity underlying these specific differences. As I have observed, it is perfectly natural to speak of both coercion and duress as sources of unfreedom, inasmuch as both seem to compromise self-determination. If possible, a complete account of autonomy ought to vindicate and explain this appearance.

1.2. Desire and Necessity

In pursuing this explanatory task, we should attend to another piece of common sense. Our ordinary thought about coercion and duress alike is structured by a conceptual opposition between a sort of desire and a sort of necessity. Whereas autonomous activity is an expression of the agent’s desire, its opposite—heteronomy, unfreedom—is a reflection of necessitation alien to such desire.

Thus, it is natural to think that victims of coercion and duress act non-autonomously precisely because, in some salient sense, they do not act as they really want to, but are instead forced, compelled, or necessitated to act as they do. Indeed, the fact that we readily affirm this sort of thought about both the interpersonal and the impersonal cases lends further, more concrete support to the suggestion that coercion and duress have something in common, and that both are opposed to autonomy.

The problem, though, is to explain the relevant kind of desire, as well as the distinctive kind of necessity that is opposed to it. The solution is not obvious. After all, there is a straightforward sense in which people acting under coercion

or duress do just what they want to do. Given their nonideal circumstances, the mugging victim handing over her money, the slave doing her master's bidding, the sailor throwing goods overboard, and the person fleeing her house all act intentionally, and presumably they all prefer this action to any of the available alternatives. Now, one wants to say: these agents do not *really* or *truly* want to do these things. But the point is to explain what this means. What we can say at this stage is that the kind of desire that makes an activity autonomous must amount to something more robust than mere intention.

On the flip side, there is an equally straightforward sense in which people acting under coercion or duress are not literally forced to act as they do. The slave could opt to be beaten, and the mugging victim, the sailor, and the person fleeing her house could all choose to take their respective goods to the grave rather than part with them. So the kind of necessitation that is the mark of external unfreedom must amount to something weaker than the absolute impossibility of doing otherwise.

Coercion and duress compromise autonomy; to do something autonomously is in some sense to do what one wants to do, but victims of coercion and duress do what they are forced to do, where the relevant kind of necessitation is opposed to the relevant kind of desire. These platitudes function as theoretical constraints. We have reason to be dissatisfied with a theory that does not vindicate and explain them. But as we will see, they also function as useful footholds. We can try to make progress by starting with these fairly abstract ideas and giving them greater determinacy.

2. SUBJECTIVISM AND OBJECTIVISM

In this section, I argue that standard theories of autonomy struggle to explain the platitudes I just described. Since my subject matter is autonomous activity, I focus in the first instance on theories of "local" rather than "global" autonomy: that is, theories of what it is to do a particular thing autonomously, rather than what it is to be an autonomous person over the course of one's life as a whole. These theories tend to be deeply subjectivist, making the problem with coercion and duress fundamentally a function of the agent's reflective attitudes. At the end of the section, I also consider an alternative, more objectivist approach that is more usually associated with global autonomy. The two types of approach, I argue, struggle for opposite reasons. The subjectivist approach is so fundamentally dependent on the agent's subjective attitudes that it leaves us unable to do justice to the role that the concept of autonomy can and should play in the justification and criticism of those attitudes. The objectivist approach, on the other hand, is so divorced from the agent's subjective attitudes that it leaves us

unable to capture the sense in which the unfree agent does not what she wants but what she is forced to do.

2.1. Subjectivism

What does the autonomy of an activity require beyond mere intention? If autonomy is self-determination, it seems that the desire from which the agent acts must be internal rather than alien to her: it must be *her own*, in some especially robust sense. To put the point another way, the desire must express her practical identity, or who she is as an agent. Only then does the realization of this desire count as *self-determination*. But what is it for a desire to be robustly “one’s own,” or to express one’s “practical identity”?

The most influential approach to this question—pioneered by John Plamenatz and later developed by Gerald Dworkin, Harry Frankfurt, and many others—centers on our capacity for self-reflection: our capacity to turn our attention toward our own motives and to form higher-order attitudes about them.⁹ We are not always moved by our desires automatically, as it were, but have the power to step back from at least some of them and appraise them critically. On the basis of such appraisal, whether in the light of reason or in the light of other elements of our psychology, we can endorse or disavow our first-order desires, approve of being moved by them, or resent being moved by them. But to endorse or approve of one’s first-order desire is in a sense to identify with it, to recognize it as truly one’s own. To disavow or resent it, on the other hand, is to be alienated from it, to look upon it as a force external to one’s true will.

On this account, then, it is the agent’s higher-order attitudes that determine whether a given desire expresses her practical identity or constitutes something more like an alien force acting upon her. It is her higher-order attitudes that make the difference between autonomy and heteronomy. To quote Plamenatz, whose version of the view is especially clear and simple: “freedom must be defined, in its primary . . . meaning, as action from a motive from which a man desires to act, or, at least, does not desire not to act, and the lack of freedom as action from a motive from which he does not desire to act.”¹⁰ The suggestion is thus that an agent who gives in to an unwanted addiction, for example, acts non-autonomously because, although she may do what she desires, the relevant desire is not her own in the fullest sense, because she does not identify with it; on the

9 Plamenatz, *Consent, Freedom and Political Obligation*, ch. 5; Dworkin, “Acting Freely”; and Frankfurt, “Freedom of the Will and the Concept of a Person,” “Identification and Wholeheartedness,” and “The Faintest Passion.” Some in this literature speak of “freedom,” some of “autonomy,” and some simply of “identification.” But beneath the superficial diversity of expression lies an obvious unity of subject matter and content.

10 Plamenatz, *Consent, Freedom and Political Obligation*, 122.

contrary, she wishes that she did not have it, or at least that it did not move her to action on this occasion. She resents being moved to action in this particular way.

There are many sophisticated and idiosyncratic variations on this approach to which I cannot do justice here.¹¹ Broadly speaking, the variations tend to modify the basic view along three dimensions. First, theories in this tradition vary in terms of what specific types of reflective attitude play the decisive role in identifying an agent with, or alienating her from, her own motives. Some speak simply of higher-order desires; others employ richer notions such as “resentment”; others still use more technical terms such as “commitment” and “policy” (for example, about what to treat as a reason for action). Second, theories vary, too, in terms of the objects of these reflective attitudes. In some versions, the relevant reflective attitudes are attitudes toward one’s own desires. In others, the reflective attitudes instead (or additionally) concern the historical process by which one’s desire came to be. In others still, the (or another) relevant object of reflective attitudes is the choice situation that gave rise to the desire. Finally, many theories stipulate some kind of cognitive condition, to the effect that what really matters is the reflective attitude that the agent would have toward the relevant aspect of her motivational structure *if*, for instance, she were minimally rational and fully informed. Typically, however, these conditions of rationality and the like are quite minimal; to the extent that they are not, the view begins to depart in a more substantial way from the basic Plamenatzian model.

What unites all of these versions of the approach is a deeply subjectivist orientation, in the sense that they all conceive of autonomy as largely a function of what the agent thinks and feels, or would think and feel, about elements of her motivational structure or the situation that produces this structure. The relevant reflective attitudes ultimately divide the agent’s first-order desires into “authentic” desires that are truly her own, and “inauthentic” desires that are hers only in a qualified way, whether because they are immediately objects of a relevant negative reflective attitude or because they arise from a process or from a situation that is the object of such an attitude (or would be if the agent were aware of it). Fundamentally, what makes an activity fully autonomous is that it is an expression of authentic rather than inauthentic desires in this sense.

How can the subjectivist picture help us to understand what is wrong with coercion and duress? For the sake of illustration, let us work in the first instance

11 See, e.g., Neely, “Freedom and Desire”; Thalberg, “Hierarchical Analyses of Unfree Action”; Christman, “Autonomy and Personal History”; Ekstrom, “A Coherence Theory of Autonomy”; Bratman, “Identification, Decision, and Treating as a Reason”; Velleman, *The Possibility of Practical Reason*, ch. 1; J. S. Taylor, “Autonomy, Duress, and Coercion,” 138; Christman, “Procedural Autonomy and Liberal Legitimacy” and “Introduction”; and Killmister, *Taking the Measure of Autonomy*.

with Plamenatz's straightforward version of the view, according to which autonomous activity requires that one does not act from an "undesired" motive, that is, a motive from which one actively desires not to act. This condition seems to deal in an obvious way with some cases of coercion—namely, those involving what the law calls an "overborne will." An agent may divulge a secret under torture to avoid further suffering, while wishing that she had the inner strength to take the secret to her grave. This agent suffers what Frankfurt calls an "inner defeat."¹² She does not do what she *really* or *truly* wants—she acts "against her own will"—inasmuch as she acts from a motive from which she desires not to act. She therefore acts non-autonomously.

Not all volitional coercion involves an overborne will. The victim of a mugging may not suffer an "inner defeat" like the victim of overbearing torture. She may not wish that she had the strength of will to take her money to the grave. Given the circumstances, she presumably not only wants to hand over the money, but very much wants to have this desire and to act from it. Her first-order desire is not straightforwardly an object of her aversion in the sense that she would prefer not to be moved by it, *given* the circumstances.¹³

Even so, however, we can say that she acts from an "undesired" motive in the sense that there is some description of her motivational structure under which she has a negative attitude toward this structure. For instance, it is often suggested that people resent acting not in order to improve their condition but merely to keep it from becoming worse, or that they resent being motivated by "duress," or by "menacing" or "coercive" situations.¹⁴ In any case, what makes action under coercion or duress unfree is that the agent has some negative attitude toward some distinctive aspect of the way in which she is being moved to action. It is because of this negative attitude that an activity that flows from the relevant kind of motivational structure is not a true expression of its agent's will but alienated, importantly foreign to her will.

2.2. *The Wrong Order of Explanation*

The notion that victims of coercion and duress resent or are otherwise psychologically alienated from the way they are motivated is supposed to capture the sense in which the relevant desires are not really or truly their own—which in turn explains why these agents act unfreely. But reflective attitudes, such as the

12 Frankfurt, "Coercion and Moral Responsibility," 81.

13 This is the point of Irving Thalberg's objection in "Hierarchical Analyses of Unfree Action."

14 Dworkin, "Acting Freely," 377; Frankfurt, "Coercion and Moral Responsibility," 82–83; J. S. Taylor, "Autonomy, Duress, and Coercion," 138; Thalberg, "Hierarchical Analyses of Unfree Action," 217–18; Christman, "Introduction," 8–9. Compare Killmister, *Taking the Measure of Autonomy*, 109–10.

desire not to act from certain types of motive, cannot play the explanatory role demanded of them in this account, for, properly understood, such attitudes are *responses* to unfreedom rather than *grounds* of it.

Consider, for a start, what we should say if a victim of coercion or duress lacks the relevant kind of reflective aversion to her own motive. Imagine a Stoic slave as described in the discourses of Epictetus. This slave feels no aversion at being threatened and ordered around. When ordered to hold her master's chamber pot, she obeys with equanimity, reasoning simply that "it is a more valuable thing to get a dinner than not; and a greater disgrace to be given a thrashing than not to be."¹⁵ Her lack of reflective aversion is not the product of an impaired capacity for reflection. It is not even the product of a socially inculcated sense of her own inferiority. Rather, it springs from her reflective and authentic moral conviction that she should not rail against that which she cannot control. Does this person act autonomously?

As several philosophers have argued, subjectivist theorists of autonomy must say yes, since from the point of view of the slave's reflective attitudes, everything is in order.¹⁶ But this conclusion is difficult to take seriously. The Stoic slave may indeed feel or believe that she is free. Feeling free, however, or believing oneself to be free, even when this feeling or belief is formed in a minimally rational and fully informed way, is not the same as being free. When the slave holds the chamber pot for her master—something she has no independent desire to do, and which she does only to avoid being beaten—she acts not autonomously but under the compulsion of an alien force. This is a paradigm case of heteronomy.¹⁷

Superficially, this looks like a problem concerning the extensional adequacy of the theory, and that is how it has tended to feature in the literature.¹⁸ The real problem, however, is more fundamental; it concerns the *structure* of the subjectivist explanation. What the case of the contented slave highlights, I think, is that when we feel aversion at being coerced, for instance, this is not a brute psychological fact about us. For the contented slave does not differ from a slave who resents her condition simply in being constituted psychologically *differently*. Rather, there is something that the latter gets right and the former

15 Epictetus, *The Discourses of Epictetus*, bk. I, ch. 2.

16 Zimmerman, "Making Do"; Kristinsson, "The Limits of Neutrality"; J. S. Taylor, "Autonomy, Duress, and Coercion," sec. 4. As Kristinsson points out, some subjectivists seem to embrace this conclusion quite explicitly; see, e.g., Dworkin, "The Concept of Autonomy," 23; and Christman, "Liberalism and Individual Positive Freedom," 359.

17 For a more thorough discussion of counterexamples in the same vein, see Oshana, "Personal Autonomy and Society" and *Personal Autonomy in Society*, ch. 3.

18 For some responses to the charge of extensional inadequacy, see Killmister, *Taking the Measure of Autonomy*, esp. ch. 6.

gets wrong. That is, the former fails to feel a kind of aversion that her situation warrants, indeed demands. Aversion to being coerced—aversion at being motivated in this kind of way—is thus not a brute psychological response to a stimulus but a *warranted response* to this kind of treatment.

But what is it that warrants this negative reflective attitude? In other words, what is objectively wrong with being subjected to coercion or duress, such that this condition normally and properly gives rise to reflective aversion? Most obviously, what is wrong with it is precisely that to act in this way is to act unfreely. Such action is not self-determined, and this is why we can be expected to do it only with aversion. What the Stoic slave fails to recognize, then, is that her condition warrants or demands aversion because it is a condition of unfreedom. Her equanimity is a mark of insufficient concern for her own autonomy.

From a subjectivist point of view, however, we cannot make sense of these natural and attractive thoughts. For from that point of view, as Dworkin says quite explicitly, “we do not find it painful to act because we are compelled; we consider ourselves compelled because we find it painful to act for these reasons.”¹⁹ We can now see that this gets things the wrong way around. The various modifications of the basic subjectivist approach do not help with this problem. For instance, it makes no difference here whether what a victim of coercion or duress reflectively resents is the motive from which she acts or the process or situation that gives rise to this motive. In any of these cases, we are owed an explanation of why she happens to resent just *these* kinds of motive, process, or situation and not others, and this explanation should make the resentment intelligible as a *warranted response* to the relevant kind of motive, process, or situation. But the prospects for a successful explanation from a subjectivist point of view are dim, given the unavailability of the most obvious such explanation—namely that the relevant kind of motive, process, or situation makes us unfree.

The Plamenatz-Dworkin-Frankfurt approach to understanding the unfreedom of action under coercion and duress, then, is not promising on reflection. Because of the depth of its subjectivist dependence on the agent’s psychology, this approach seems to leave us unable to account adequately for the crucial role that the concept of autonomy itself can and should play in the justification and criticism of our reflective attitudes. The problem is not simply that we sometimes fail to feel resentment when we act under coercion or duress. It is that when we do feel such resentment, it is normally, properly, and fundamentally because coercion and duress are making us act unfreely—because they involve forms of motivation that compromise our autonomy. But then it cannot

19 Dworkin, “Acting Freely,” 378–79. Compare Epictetus: “Nobody, then, who lives in fear, pain, or distress is free; but whoever is delivered from pains, fears, and distresses, by the same means is delivered likewise from slavery” (*The Discourses of Epictetus*, bk. II, ch. 1).

be the case, conversely, that they compromise our autonomy fundamentally because we resent them. Resentment is downstream of unfreedom.²⁰

2.3. *The Wrong Kind of Necessity*

In our ordinary thought, the sense in which the unfree agent does not do what she really or truly wants is opposed to a correlative sense in which she is instead forced to do it. A secondary shortcoming of the subjectivist approach is that it suggests no plausible account of the relevant kind of necessity.

To act from a motive that is an object of one's aversion is not in itself to act under the guise of compulsion or necessity. The mugging victim may choose to hand over her money merely in order to avoid the greater evil of being shot but, strictly speaking, she could have chosen instead to take the money to her grave. Whether or not she happens to feel aversion at the motive from which she in fact acts simply seems to be neither here nor there, as far as the necessity or non-necessity of her action is concerned. Of course, since people cannot be expected voluntarily to act from motives to which they are averse, it stands to reason that she must indeed be forced to act as she does. But we already knew this. The problem was, and still is, to explain what it means and why it is true.

Perhaps a further proposal from Frankfurt could help. In "Coercion and Moral Responsibility," he suggests that the desires to which coercion and duress give rise in their victims are "irresistible." The idea is that the victim of a mugging, for instance, will typically not only acquire a desire to acquiesce to the mugger's demands so as to save her life, but will find it psychologically impossible not to act on this desire, even if she wants to resist it.²¹ "The victim's desire or motive to avoid the penalty with which [she] is threatened," Frankfurt says, "is—or is taken by [her] to be—so powerful that [she] cannot prevent it from leading [her] to submit to the threat."²² The kind of necessity that characterizes external unfreedom, on this view, is the inexorability with which the desire leads to its own expression in action.

Though Frankfurt at least takes the problem of necessity seriously, his proposal looks like a step in the wrong direction. The suggestion that victims of

20 The same problem arises for idealized versions of the subjectivist picture. Thus, Robert Nozick explains the unfreedom of coerced action by appeal to the aversion that the "Rational Man" feels at being threatened ("Coercion," 460). But of the Rational Man, too, we should be able to say that he feels aversion at being threatened *because* this makes his action unfree.

21 Frankfurt, "Coercion and Moral Responsibility," 80–81.

22 Frankfurt, "Coercion and Moral Responsibility," 77. For the sake of simplicity, I am largely ignoring Frankfurt's caveat that the victim might merely *think* that she is unable to resist the desire. I do not think this makes a difference to my arguments.

coercion generally act from desires that are literally irresistible wrongly assimilates all cases of coercion to those special cases that involve an overborne will.²³ A mugging victim might happen to be so frightened by the prospect of imminent death that she is unable to resist the desire to acquiesce. But another victim, perhaps endowed with an unusually proud and rebellious temperament, might find herself pulled more or less equally toward acquiescing on the one hand and resisting her coercer on the other. She might be perfectly *capable* of choosing the option of dying rather than acquiescing. If this victim ends up acquiescing, however, her action, too, is forced and less than fully autonomous.

Relatedly, the appeal to the irresistibility of desire seems to misidentify the source of necessitation in typical cases of external unfreedom. On Frankfurt's proposal, a coerced action is necessitated by the irresistible desire that motivates it. When we speak of coerced actions as forced, however, what we have in mind is that they are necessitated not by the agent's own overpowering urges but by something altogether external to the agent, such as her coercer or her environment. We have yet to make sense of this thought. These first two mistakes—that coercion in general involves an overborne will, and that the source of necessitation is internal to the agent herself—appear to be symptoms of a single error: the attempt to assimilate the external dimension of autonomy too directly to its internal dimension.²⁴

Finally, the structure of Frankfurt's account should give us pause. On his view, coercion and duress compromise autonomy because of the confluence of two conditions: the agent's aversion to the desire from which she acts, and the psychological irresistibility of this desire. The former is supposed to explain the sense in which she does not do what she really wants, and the latter the sense in which she is forced to act as she does. But these two conditions are independently intelligible, and their confluence is fundamentally accidental.

By contrast, in our ordinary thought about these matters, the relevant kind of desire and the relevant kind of necessity are internally related, as the poles of a conceptual opposition. When we say that the victim of coercion or duress acts unfreely because she does not do what she really wants *and* because she is forced to act as she does, we are offering not two separate explanations, or even two separate parts of an explanation, but one and the same explanation in two different ways. To make sense of this thought, we will need a different sort of account, one on which there is an internal relation between doing what

23 For a version of this objection, see, e.g., Zimmerman, "Making Do," 38.

24 This is related to James Stacey Taylor's charge that Frankfurt's theory of autonomy is "metaphysical" rather than "political." (J. S. Taylor, *Practical Autonomy and Bioethics*, ch. 3.) See also Oshana, *Personal Autonomy in Society*, *passim*.

one does not really want to do on the one hand and being forced to do it on the other.

2.4. *Objectivism*

Given that subjectivist theories struggle to make sense of external autonomy and unfreedom, perhaps we should instead try formulating some objective conditions of autonomy that do not depend on the agent's desires or other attitudes. Such conditions have typically been defended in the context of theories of global autonomy—but conceivably they could also help us with our local-autonomy problem.

Joseph Raz argues that a person is autonomous, over the course of her life as a whole, only if she is independent—that is, not subjected to the will of another person—and she has an adequate range of options to choose from.²⁵ As marks of global autonomy, these conditions seem plausible enough. A person who lives her life under the arbitrary power of another cannot truly be said to be self-governing; she lacks the final authority to choose how to live her life, making such choices only at the whim of the other person. Likewise, we might think that if someone lacks the opportunity to choose how to live her life from among a range of diverse and valuable options, she thereby lacks the opportunity for one important kind of exercise of her agency. In the absence of valuable alternatives to the life she in fact leads, she can say yes to this life but there is a sense in which she cannot say no to it.²⁶

I do not doubt that some version of the independence condition is also true as a condition of the autonomy of a particular activity. Someone who acts under coercion is obviously subjected to the will of another person, and this violation of her independence compromises the autonomy of her action. Nevertheless, I

25 Raz, *The Morality of Freedom*, ch. 14, sec. 1. For a similar account, see Oshana, *Personal Autonomy in Society*, ch. 4. For other defenses of something like the independence condition, see, e.g., Pettit, *Republicanism*, ch. 1; Lovett, "Domination"; and List and Valentini, "Freedom as Independence." In a similar vein, Al Mele suggests a non-coercion condition for local autonomy, comparable to the "procedural independence" condition that Dworkin defends in his later work, though Dworkin does not apply it to coercion specifically. (See Mele, *Autonomous Agents*, ch. 10, sec. 4; and Dworkin, "Autonomy and Behavior Control," 25, and *The Theory and Practice of Autonomy*, 18–19. Compare Christman, "Procedural Autonomy and Liberal Legitimacy," sec. 1.) For other defenses of something like Raz's options condition, see, e.g., Cohen, "The Structure of Proletarian Unfreedom"; van Parijs, *Real Freedom for All*, ch. 1; and Goodin et al., *Discretionary Time*, ch. 2. I should reiterate at this point that I am largely putting aside those approaches whose principal aim is to explain why coercion violates rights or undermines responsibility.

26 Hurka, "Why Value Autonomy?"

do not find this condition helpful, at least in the form in which Raz presents it and with respect to the problem I am trying to solve in this paper.

It is worth emphasizing, to begin with, that in the absence of further analysis the condition amounts to little more than a restatement of a theoretical desideratum, or perhaps a placeholder for an explanation. Since, as we saw, the coerced agent does just what she intends to do, pursuing her own ends as best she can in admittedly nonideal circumstances, the difficulty is to understand in what sense she is indeed subjected to the will of another person rather than being directed by her own will. Significant theoretical work, then, remains to be done here. More to the point, though, even a helpfully fleshed-out independence condition will only tell us something about interpersonal unfreedom in particular. That is, perhaps it could ultimately explain what makes interpersonal unfreedom special vis-à-vis other forms of unfreedom—an important problem but one outside the scope of this paper.²⁷ It will not solve the problem I am trying to solve, which is to explain what interpersonal coercion and impersonal duress have in common such that they are both forms of unfreedom.²⁸

On its face, the options condition seems more promising in this respect.²⁹ Impersonal duress does not violate independence, but is characterized by a lack of valuable alternatives to the chosen course of action. The sailor who must choose between throwing his goods overboard and allowing his ship to sink lacks an adequate range of options to choose from, and this makes his choice less than fully autonomous. And victims of coercion, of course, also frequently lack an adequate range of options. On reflection, however, things are not so clear.

First, I doubt that an adequate range of options, or the availability of valuable alternatives, is necessary for a particular activity to be done autonomously. We generally lack a valuable or even minimally acceptable alternative to eating. In reasonably favorable circumstances, it is true that we can normally choose what and when to eat, but we cannot reasonably choose never to eat. Nevertheless, we do not normally eat unfreely. The lack of a valuable alternative to eating does not in itself compromise the autonomy of this activity. The problem seems to be that the options condition is too disengaged from the agent's desires. For, arguably, the reason we can eat autonomously even in the absence of valuable alternatives is that it is possible for us to really or truly *want* to eat, and that if this desire takes the relevant form (still to be clarified), this may make the activity autonomous.³⁰

27 Thanks to an anonymous reviewer for this suggestion.

28 Thanks to Brookes Brown for this suggestion.

29 Thanks to an anonymous reviewer for urging me to consider Raz's options condition.

30 Perhaps we could try a more modest version of the options condition, according to which a range of options counts as adequate so long as one has at least one valuable option. This certainly goes against the spirit of Raz's idea, but if we understand "valuable" as *intrinsically*

Second, the options condition also does little to explain why victims of duress are forced to act as they do. Raz points out that “a choice between survival and death,” for instance, “is no choice from our perspective.”³¹ Thus, one who does something merely as a condition of survival has no real choice but to do it, and is effectively forced to do it. Of course we want to say this.³² But what entitles us to say it? Why is the language of necessitation appropriate? After all, a choice between survival and death actually *is* a choice, and the agent is not literally, absolutely forced to choose survival. Even putting aside the worry about its extensional inadequacy, then, the options condition merely kicks the can down the road.

To be clear, I do not think that the problem I am trying to solve is the problem that Raz is trying to solve—so none of this is intended as a direct criticism of his theory. Nevertheless, it is instructive to consider why his resources prove so unhelpful for my purposes. The root of the problem, it seems to me, is that objectivist conditions of autonomy are not promising when they become, as it were, too objective. The options condition, in particular, does not make reference to the agent’s desires in a way that would help us understand the sense in which a victim of duress does not act as she really or truly wants to. As a result, it becomes very hard to articulate the sense in which her action is genuinely forced and non-autonomous. The fact that such an agent does make a genuine choice—that she does what she intends to do in the pursuit of her own ends—keeps annoyingly rearing its head.

In order to make progress, then, I propose that we return to the dialectic of desire and necessity that structures our ordinary thought about autonomy and external unfreedom. In particular, I propose to start again with the question what kind of desire is necessary for autonomous activity.

valuable, and we suppose further that the agent chooses the valuable option *on account* of its perceived value, then we arrive at roughly the view I end up defending in section 3.

31 Raz, *The Morality of Freedom*, 376.

32 At least this is common ground between me and Raz. Following James Stacey Taylor, one might argue instead that a victim of impersonal duress, of the sort to which Raz’s options condition might be intended to apply, suffers not a genuine impairment of her autonomy but merely a diminution of the value of her autonomy to her (J. S. Taylor, *Practical Autonomy and Bioethics*, 108). Alternatively, one might argue, following another proposal of Taylor’s, that there is a tripartite distinction at work, so that one can lack autonomy without yet being unfree: perhaps, then, a victim of the relevant kind of duress does lack autonomy but does not act unfreely. As I argued in section 1, however, a theory of autonomy should aim to vindicate both the appearance that impersonal duress genuinely compromises autonomy and the appearance of a bipolar conceptual opposition between the kind of desire associated with autonomy and the kind of alien necessitation associated with unfreedom. While we may be forced to revise these desiderata on reflection, I think we ought to try to satisfy them if possible. Thanks to an anonymous reviewer for pressing me on these points.

3. AUTONOMOUS ACTIVITY AS ACTIVITY FOR ITS OWN SAKE

In this section, I introduce and defend a different approach to autonomy and external unfreedom, following the Aristotelian-Marxian line of thought I discussed in the introduction. One does something freely, I argue, only if one does it for its own sake—that is, on account of its perceived intrinsic value—and not merely instrumentally. I motivate this idea on the basis of the same guiding intuitions as the standard subjectivist approach, but I show that it ultimately allows us to steer a promising middle path between subjectivism and objectivism, putting us in a position to explain how action under coercion or duress is non-autonomous.

3.1. *Intrinsic and Extrinsic Desires*

Let me begin with a familiar distinction between different modes of activity. On the one hand, we do some things partly or wholly for their own sake, as final ends. These can include activities, such as some forms of play and contemplation, that do not have any further end at all. But they also typically include some instrumental activities, such as meaningful work. We do not want to live lives of pure leisure; we value intrinsically the pursuit of productive activities that allow us to exercise and develop our skills while making valuable social contributions. On the other hand, there are those things, such as cleaning the toilet, that we do *merely* instrumentally, purely for the sake of further ends distinct from the activities themselves.

These modes of activity correspond to importantly different kinds of desire. A final end is something we do on account of its perceived intrinsic value: value that it has in its own right and that does not reduce to the value of other things, such as its effects.³³ Insofar as we do something for its own sake, then, we have what can be called an “intrinsic” desire to do that thing, for the perceived good that we are after lies in the activity itself. By contrast, instrumental activity is done on account of the perceived value of something else: the further end to which it is a means. Insofar as we do something instrumentally, then, we can be said to have an “extrinsic” desire to do that thing, and insofar as we do something *merely* instrumentally, we have *only* an extrinsic desire to do it.

An extrinsic desire is really a desire only in a qualified way. It is always parasitic on an intrinsic desire for something else, and in a sense it is always this other,

33 Christine Korsgaard argues that to value something as a final end is not necessarily to value it intrinsically (“Two Distinctions in Goodness”). But her argument relies on a gratuitously narrow identification of intrinsic value with unconditional value, value that an object would have in any context whatsoever; see Tannenbaum, “Categorizing Goods.” Thanks to Dan Moerner for drawing my attention to Korsgaard’s and Tannenbaum’s papers.

intrinsic desire that is the agent's "true" desire, the desire whose object is what she is really after. When someone intentionally cleans a toilet, for example, there is an obvious sense in which she "wants" to clean it. But the reason it sounds odd to say this is that her desire to clean the toilet is likely to be purely extrinsic. The real object of value for her—the object of her true and unqualified desire—is likely not the activity of cleaning the toilet but the further ends to which this activity is a mere means, such as living in a hygienic and attractive environment.

With this distinction in hand, we can begin to see more clearly what is at stake in our pursuit of autonomy, in view of familiar considerations of identification and alienation. An autonomous activity, it seems, ought to be a suitably unqualified expression of its agent's will. We said that this means that the agent ought to be able to recognize in the action, and in the desire that motivates it, her own practical identity, or "who she is" as an agent. On the subjectivist picture, this idea was spelled out, unsuccessfully, in terms of a distinction between authentic and inauthentic desires. But we can appeal more fruitfully to our new distinction.

The commitments that are definitive of an agent's practical identity without qualification—the commitments with which she identifies in the first instance—are surely her intrinsic desires, the ends that she pursues for their own sake and not merely for the sake of other things. We care about autonomy because we wish to be the authors of our own lives—but it is our final ends that really determine what we are after in life, what we want our lives to look like. It is when our activity is an immediate realization of our final ends, then, that it expresses our practical identity in the world without qualification. Of such activity we can truly and simply say that it *is* the presence of the agent's will in the world.

We cannot *simply* say this of merely instrumental activities, valued merely extrinsically. To be sure, such activities can be said to reflect the agent's agency and practical identity in a thin sense.³⁴ After all, they are still intentional activities that proceed from the agent's instrumental reasoning in the pursuit of her own ends. However, this thin expression of agency is evidently insufficient for autonomy. It is present, for instance, in actions taken under coercion or duress—paradigm cases of unfreedom. The problem is that, insofar as it is *merely* instrumental, an activity expresses its agent's practical identity and will only in a qualified, privative way, for the full realization of her will in the world lies not within this action but outside it.

To illustrate the point, consider someone who performs routine labor for a living and is utterly indifferent to their job, doing it only to pay the bills. It would be a cruel joke to say of someone in such a position that their work expresses who they are as an agent and a valuer—that they choose their work

34 Thanks to Brookes Brown for pressing me on this point.

autonomously because what they want is precisely to perform such labor. Here is a striking passage from an interview with T, a call-center worker:

Would T be happy to think that his identity came from what he does all day? “I really hope not. I could not say enough how I hope not. I used to like who I was, and if this place is now my identity, then I don’t like myself. Literally, apart from the few people that I can sit and have a chat with and a gas with, the money is only just passable as the reason I come here. So, if the money changed, or certain people didn’t work here anymore, I can safely say I would probably be at the Job Centre looking.”³⁵

T draws an immediate and natural connection between the fact that his work in some way fails to express his identity and the fact that it is a mere means to other ends. To be sure, he hedges a little, speaking only of what he “hopes” his identity is. But this ambivalence simply reflects the fact that there is *some* sense in which “who he is” is a matter of his actual doings—what he “does all day.” Evidently, the “self” constituted by these doings, though, is not one with which he identifies. And he explains why. The fundamental problem is not that he disavows any of his own first-order desires, or that these desires are in some other way “inauthentic.” It is rather that the relevant desires are merely extrinsic and so fail properly to express his practical identity. This is the sense in which he is alienated from them, and from the corresponding activities. We can thus do much of the philosophical work that we had hoped to do using higher-order desires, but while staying within the realm of first-order desire.³⁶

We have arrived, then, at the Aristotelian-Marxian necessary condition of autonomy: one does something autonomously only if one does it for its own sake and not merely for the sake of further ends. This is not to say that all instrumental action is non-autonomous. There is nothing wrong as such with having to do one thing in order to do another. Indeed, many instrumental activities are loci of great and irreplaceable intrinsic value, and motivated in part by this value. The point is that an activity fails to be autonomous insofar as it is done *merely* instrumentally, for the sake of a further end alone.

I do not claim that doing something for its own sake is also a sufficient condition of autonomy. Our final ends are themselves vulnerable to distortion by the sorts of heteronomous psychological processes already familiar from the literature. The purpose of the new condition is to shed light on the comparatively

35 Biggs, *All Day Long*, 87.

36 In according a central role to the agent’s final ends, my account resonates with Gary Watson’s theory of autonomy, centered on a robust conception of the agent’s “values” (see Watson, “Free Agency”). Since Watson does not try to explain *external* unfreedom, however, the exact relation between my account and his is not straightforward.

poorly understood external dimension of autonomy. In this dimension, the question is not whether an agent has the right desires or whether they were formed in the right way, but whether her activity is an unqualified expression of her intrinsic desires.

3.2. *Making Sense of External Unfreedom*

The idea of autonomous activity as activity for its own sake promises an explanation of how coercion and duress normally compromise autonomy—an explanation that is structured, like our ordinary thought about these matters, by an opposition between a kind of desire and a kind of necessity.

Insofar as one is coerced to do something by means of a threat, one does it in order to avoid the threatened penalty. Volitionally coerced action is thus a species of instrumental action. Clearly, however, it must normally be a case of *merely* instrumental action. For, in general, if the coerced agent were already motivated to perform the action for its own sake, then it would be unnecessary to motivate her further by means of a threat. It follows that coerced action is normally non-autonomous. Its non-autonomy is a species of the non-autonomy of merely instrumental action in general, and can be explicated by saying that the coerced agent does not do what she really wants, in the sense that she does not do what she intrinsically wants, which is the robust kind of wanting that matters from the point of view of autonomy.

Much the same can be said of classical cases of duress, such as that of the sailor who throws goods overboard in a storm, the patient who chooses lifesaving surgery, or the person who abandons her house to flee a fire. Each of these seems to be a paradigm case of merely instrumental action: the agent does it intentionally but not for its own sake. In that sense, it is not something she really or truly wants to do.³⁷

This explanation puts higher-order attitudes in their place. The agent acting merely instrumentally may well resent the way she is being motivated, whether by another person or by her circumstances. But we can now say, plausibly, that her resentment is a response to her unfreedom rather than the explanation of it. Thus, the Stoic slave does not become autonomous just by virtue of her equanimity. And even when an agent does resent being coerced, her non-autonomy is not a function of her resentment but its cause.

So much for the “desire” pole of the conceptual opposition. We also commonly say that agents acting under coercion or duress are forced, compelled,

37 Compare Aristotle’s discussion of the sailor’s action. Such actions, he says, seem voluntary inasmuch as they “are worthy of choice at the time when they are done,” but they cannot be called voluntary without qualification because “no one would choose any such act in itself” (Aristotle, *Nicomachean Ethics*, 1110a).

or necessitated to act as they do. Now we are in a position to explain what this means. Initially, I characterized my account in terms of a distinction between different sorts of desire, intrinsic and extrinsic. But to do something merely instrumentally, which is to act merely from an extrinsic desire, is to do it only because it is instrumentally necessary (other things equal) for the realization of one's further ends. It is to do what one *has* to do in order to do what one really wants to do. The distinctive kind of necessity that is the mark of external unfreedom, in other words, is *mere instrumental necessity*.

Unlike Frankfurt's attempt to account for the same thing, this explanation does not rely on the fiction of a psychologically irresistible desire, which would assimilate volitional coercion in general to special cases involving an overborne will. Moreover, the explanation locates the source of necessitation in the right place: not in the agent's own desire but in circumstances of her environment—those that account for the necessity of the means for the end. Finally, the relevant kinds of desire and necessity are not two separate, independently intelligible conditions but two sides of the same conceptual coin. To do something merely instrumentally is on the one hand not to do what one really wants to do (in the sense of what one intrinsically wants to do), and on the other hand—by the same token—to do merely what one is forced to do (in the sense of what one has to do for the sake of one's further ends).

The Aristotelian-Marxian account steers a path between subjectivism and objectivism. The subjectivist picture, guided by a preoccupation with psychological threats to autonomy, revolves around a contrast between authentic and inauthentic desires. Since this is essentially a division within the agent's mind, it is not surprising that the subjectivist picture struggles to make proper sense of external sources of unfreedom such as coercion and duress. The objectivist picture errs in the opposite direction. It takes more seriously the distinctiveness of external threats to autonomy, but it untethers this dimension of autonomy excessively from the agent's subjectivity. If duress is not understood fundamentally in conceptual opposition to some kind of desire, it proves difficult to understand how it can compromise autonomy.

The Aristotelian-Marxian picture does revolve around a conceptual opposition between (intrinsic) desire and (mere instrumental) necessity. This contrast remains tethered to the mind of the individual agent, in the form of her intrinsic desires or final ends. But at the same time, it situates the agent essentially within an external world, which is not just a place where we realize our desires but also a source of compulsions altogether alien to our desires. It is therefore no surprise that this alternative picture should be better placed than either purely subjectivist or purely objectivist approaches to make sense of external threats to our autonomy.

4. DOUBTS ABOUT NECESSITY

In the rest of the paper, I qualify and refine the basic Aristotelian-Marxian idea in important ways. I approach this task indirectly, by discussing some possible challenges to the view. In this section, I consider whether merely instrumental activity can really be said to be “necessitated” and hence unfree when there are other possible means to the same end or when that end is unimportant. I argue that mere instrumental necessity can be understood as a genuine form of necessity and a genuine source of unfreedom in both cases. I suggest, however, that the degree of this unfreedom is higher when the end in question is relatively important.

4.1. *Alternative Means*

Mere instrumental necessity is a species of instrumental necessity, or necessity for an end. It is rare, however, that a particular instrumental activity is the only possible means to its end; at least at a sufficiently specific level of description, there are usually bound to be alternatives. But if there are alternative means to a given end, it might seem that the means cannot be considered necessary for that end. Consider T, who works at the call center to pay his bills. If he could not pay his bills without working in some job or other, then we can say that he is forced *to work*: at this level of description, his activity is necessary for the further end (and done purely on account of this necessity). But if he could pay his bills by working at a fast-food restaurant instead, then it seems we cannot say that he is forced *to work at the call center*. He could have done something different and still achieved his end of paying the bills.

If this were right, then the core of my view could remain unchanged, but the relation between autonomy and necessity would be more complicated than I suggested in section 2.2. For I would have to say something like the following. An activity is autonomous only if it is done for its own sake, and not merely for the sake of further ends. But not all activities that fail to be autonomous are thereby *forced* or *necessitated*; or at any rate, they are not forced under all levels of description. A non-autonomous activity under a given description is forced only if there are no alternative means to the same end.

Perhaps we could put this in terms of a distinction between the absence of autonomy and the presence of unfreedom. Since T works purely for the sake of paying the bills, he does not work autonomously; and since he furthermore has no alternative means to this end, he is forced *to work* and does so unfreely. Similarly, T works at the call center purely for the sake of paying the bills, so he does not work at the call center autonomously; but since he has alternative means to the same end, at this level of description, he is not forced *to work at the*

call center and does not do so unfreely. His activity, at this level of description, falls short of autonomy but is not quite unfree.

I do not think any of this is obviously implausible. Yet there is something to be said for a more straightforward line of thought. First of all, it is important to see that there is a meaningful sense in which T's activity is genuinely necessary for its end even under the more specific description. While not necessary for the end without qualification, the activity is necessary for the end, *other things equal*, that is, assuming T does not take some other means to his end instead. This is a qualified but, I think, genuine sense of necessity, for it amounts to a genuine kind of counterfactual dependence.

By analogy, suppose a forest fire started because someone dropped a lit match on some dry leaves. The lit match may not have been necessary for the fire to start in the strictest possible sense. Perhaps, for example, the fire could equally have been started by the application of a flamethrower. Yet we are happy to say, I think, that the lit match being dropped was in some genuine sense a necessary condition of the fire. The reason we are happy to say this is that the latter event is counterfactually dependent on the former. That is, if the match had not been dropped, then the fire would not have started. This counterfactual clearly presupposes an other-things-equal condition, but it seems unproblematic.

The end of an instrumental action displays a similar counterfactual dependence on the action. That is, when someone does X for the sake of a further end Y, it is generally the case that she would not attain Y if she did not do X, other things equal. For instance, if T did not work at the call center, he would not get paid, other things equal. If the counterfactual dependence of the forest fire on the dropping of the match corresponds to a sense in which the latter was necessary for the former, the same seems true of the relation between an instrumental action and its further end. The action is necessary for its end, other things equal.³⁸

Arguably, this qualified kind of necessity is the kind of necessity relevant for external unfreedom. A slave who is told to cook dinner might be left with a choice between cooking various specific meals. If she decides to cook stew, this action—cooking stew—is necessary for avoiding punishment only in the qualified, *ceteris paribus* sense: it is necessary for her end *if* she does not take some other means to her end—in particular, if she does not cook some other meal—instead. But to me, it seems absurd to deny that the slave is forced to cook the stew and cooks the stew unfreely. Accordingly, it seems to me on balance that the necessity in play when we say that someone is forced to do

38 Thanks to an anonymous reviewer for pressing me on this point.

something is not incompatible with the existence of alternative means to the same end. It is necessity, other things equal.³⁹

4.2. "Optional" Ends

Be that as it may, the end-relative nature of (mere) instrumental necessity raises a second worry. A low-wage call-center worker might work merely instrumentally in order to provide himself with food, shelter, and access to health care. A wealthy investment banker, with enough savings to retire comfortably, might also work merely instrumentally, perhaps in order to fund her dream of going on a commercial space flight or indulge some other "expensive taste." But there is surely a crucial difference here. Unlike the call-center worker, the banker seems to work for the sake of an end that is itself not essential but optional. Whereas food, shelter, and access to health care are necessities, the opportunity to go on a commercial space flight is a comparative luxury—an object, it is tempting to say, of *mere desire* as opposed to *need*. Is it not implausible, then, to say that the banker is forced to do her work? Surely a necessary condition of an optional end is itself optional, not necessary.

The idea that one of these ends but not the other is necessary without qualification, however, is a fiction that does not stand up to scrutiny. The call-center worker, for his part, could in principle choose to give up the end of obtaining food, shelter, and health care; it is not literally impossible for him to do this. To be sure, in giving up these ends, he would also be giving up the perceived goods he is after: survival itself, and everything to which survival is a means. But for the investment banker, too, to give up on the project of satisfying her expensive taste would be to give up the perceived good she is after. If there is a difference between the two cases, it does not consist in the fact that the call-center worker cannot do otherwise than pursue his end whereas the banker can do otherwise than pursue hers.

It does seem reasonable to say that the worker's end is likely to be more necessary for his good than the banker's end is for hers. Here, we are conceiving of the ends in question—survival and luxury space flight, respectively—also as means to some highly general end such as living a good life. This general good comes in degrees. And the point is that dying would likely result in a greater loss of it for the call-center worker than remaining permanently on Earth would for the banker. To the extent that this is true, it seems reasonable to say that the call-center worker's work is more necessary for his good, and that his work is in this sense necessitated more stringently. But given that he works *merely*

39 Note that this is consistent with allowing that the availability of (meaningfully different) alternative means is an important constituent of *global* autonomy.

instrumentally, this necessitation is a compulsion alien to his will. It therefore seems reasonable, too, to say that he works more unfreely than the banker does.

This picture has a subjectivist component, for I am suggesting that the degree to which a person's instrumental activity is necessitated is a matter of the degree to which her *perceived* good depends on it.⁴⁰ Conceivably if improbably, our banker might in fact place such a high value on luxury space travel that she would consider her life barely worth living without it. In that case, the work she did merely as a means to this further end would be necessitated and unfree to a very high degree. This strikes me as a plausible implication.

Notwithstanding this subjectivist component of my proposal, it is no accident that instrumental activities done for the sake of the things we usually call "basic necessities" are often necessitated to a distinctively high degree. This is not just because it is highly unusual for people to value luxuries such as space travel so obsessively that they would not consider their lives worth living without them. More importantly, it is because basic necessities tend to be highly *load bearing*: they tend to be necessary not just for this or that particular perceived good but for a very large range of perceived goods.⁴¹

For instance, food and shelter are necessary for survival, which is obviously necessary for very many of the things people might want to do with their lives—including, indeed, the enjoyment of luxuries such as flying to space for fun. Thus, even for the banker who greatly values this particular luxury, any work she does merely in order to secure her survival is likely to be still more necessary, and still more unfree, than the work she does merely in order to pay for her space flight. By the same token, we can say more generally that merely instrumental activity for the sake of procuring basic necessities tends to be highly unfree.

This claim rests on the following general principle: a merely instrumental activity is more stringently necessitated and hence more unfree, other things equal, when more is at stake for the agent. The principle also applies in an independently plausible way to classical cases of coercion and duress. Coercing someone by threatening her life makes her action more unfree than coercing her by threatening to destroy her stamp collection. There is greater unfreedom in leaving one's house to escape a fire than in fleeing to escape a rat infestation. I take it to be a strength of my account that it is able to vindicate and explain these comparative judgments of unfreedom.

Importantly, however, the difference in each case is altogether a matter of degree. For all that has been shown, there is no *categorical* difference between the call-center worker and the banker as far as the autonomy of their activity is

40 Thanks to Brookes Brown and to an anonymous reviewer for urging me to clarify this point.

41 Compare Rawls's notion of "primary goods" (*A Theory of Justice*, sec. 15).

concerned. In particular, it is not the case that the banker works autonomously. To be sure, for pragmatic reasons, we may shy away from rhetorically powerful words such as “forced” to describe cases like hers, where what is at stake for the agent is likely fairly trivial. Strictly speaking, though, both workers work unfreely, albeit to different degrees. Both work not under the guise of (intrinsic) desire but under the guise of (mere instrumental) necessity.

5. TOO MUCH UNFREEDOM?

I have defended the Aristotelian-Marxian approach to freedom on grounds of general principle, as a straightforward interpretation of the idea that free activity is done from desire as opposed to necessity. I have also defended it on the basis of its explanatory power with respect to paradigm cases of unfreedom, for it makes sense of cases of duress and coercion that standard theories of autonomy struggle to illuminate. Yet the view considerably shrinks the domain of autonomous activity compared with competing theories, for if it is true, then merely instrumental activity is as such unfree. And this implication will strike many people as unacceptable. First, it might be argued, the implication is counterintuitive in itself, for many instrumental activities do not ordinarily strike us as unfree. Second, the implication may seem to devalue the concept of freedom, for many merely instrumental activities seem intuitively unobjectionable, so if they really are unfree, then their unfreedom cannot be of a kind worth caring about—or so one might argue.

Now, in general, that a view has counterintuitive implications is not necessarily a decisive objection against the view. Beyond paradigm cases, our intuitions are often unstable, subject to disagreement, and undoubtedly fallible. Sometimes, systematic theory should lead us to revise our intuitive judgments rather than vice versa. That said, I do not think that the Aristotelian-Marxian view requires as drastic a revision of our intuitions as one might think. In this section, I try to take some of the edge off, as it were, by arguing that the implications that appear most counterintuitive are largely confined to trivial cases where the relevant intuitions can be either accommodated or explained away.

5.1. *Trivial Cases*

The intuitive case against the Aristotelian-Marxian picture is strongest, I think, with respect to very trivial sorts of action. At a sufficiently specific level of description, life seems full of merely instrumental activities.⁴² In order to complete any number of tasks throughout my day, I have to push a variety of buttons—on

42 See, e.g., Jaeggi, *Alienation*, 207–8.

phones, elevators, washing machines, etc. Presumably, when I push such a button I normally do this not for its own sake but *merely* for the sake of its useful effect. On the Aristotelian-Marxian understanding, it follows that such activities are done unfreely—yet this implication seems strange in itself. Moreover, even if these activities are unfree, they seem obviously unobjectionable, which calls into question the normative significance of this category of supposed unfreedom.

I acknowledge that the unfreedom of such trivial activities seems untroubling in these sorts of cases. But this appearance has a straightforward explanation that is compatible with the Aristotelian-Marxian view. I have already argued that merely instrumental activities are less unfree if they serve unimportant *ends*; and presumably a lesser degree of unfreedom matters less than a greater degree, other things equal. I would now add that the unfreedom of merely instrumental activities likewise matters less, other things equal, when the *actions* themselves are insignificant, in the sense that they occupy a trivial place in the agent's life—as pushing the occasional button in the course of daily life surely does.

There are bound to be comparatively trivial instances of unfreedom on any plausible view of freedom and unfreedom. The distinction between those activities that really matter—those that play a significant role in our lives—and those that are comparatively unimportant is central to ethical and political thought in general. We presuppose it whenever we identify certain domains of human activity as salient subjects of freedom or autonomy in political philosophy: labor, sexuality, religion, political participation, and so on. As Charles Taylor puts it, “we make discriminations between obstacles as representing more or less serious infringements of freedom. And we do this, because we deploy the concept against a background understanding that certain goals and activities are more significant than others.”⁴³ The fact that a certain kind of unfreedom—such as the unfreedom of merely instrumental activity—is unimportant when the activities in question are trivial, then, hardly implies that it is not very important in other cases. The point is that good moral and political philosophy is partly a matter of focusing on the right things and asking the right questions.

The same considerations, I submit, can explain away our intuitive sense that these are not cases of unfreedom at all. The concept of unfreedom has its primary domain of application in the sphere of significant rather than trivial human activities, and in this primary domain it carries strong normative implications. When we apply the concept instead to very trivial activities where these implications are for all intents and purposes absent, we naturally feel reluctant to speak of genuine unfreedom, since some of the typical—normative—marks

43 Taylor contrasts the restrictions imposed by traffic lights with those imposed by a law limiting freedom of religion (“What’s Wrong with Negative Liberty,” 217–18).

of unfreedom are more or less missing. None of this shows, however, that the actions in question are without qualification *free*.

Certainly there are *specific* ways in which the activities in question can be said to be “free”—specific dimensions on which they may not be unfree. The ordinary, everyday sort of button pushing is free, for instance, in the specific sense that it is not coerced. This is the most important specific way in which an activity can be free or unfree, and in some contexts it will be all that matters. But uncoerced activities are not necessarily free *simpliciter*, or free in all ways. The way in which I have argued merely instrumental activities are unfree is the very generic sense that they are done under the guise not of desire properly speaking but of mere necessity, and therefore fail to be full and unqualified expressions of their agent’s will. Viewed in this light, the claim that even trivial sorts of merely instrumental activity are not free without qualification—that human freedom does not flourish in such activity—strikes me as an admittedly unfamiliar but ultimately principled and plausible upshot of our ordinary thought about freedom and compulsion.

5.2. *Nontrivial Cases*

What about less trivial cases of merely instrumental activity? It does not seem odd to me to speak of objectionable unfreedom in such cases. Consider what I take to be an importantly representative example, a different and rather more troubling case of button pushing. Here is Phil Stallings, describing his job as a spot welder in a car factory:

The welding gun’s got a square handle, with a button on the top for high voltage and a button on the bottom for low. . . . I stand in one spot, two- or three-foot area, all night. The only time a person stops is when the line stops. We do about thirty-two jobs per car, per unit. Forty-eight units an hour, eight hours a day. Thirty-two times forty-eight times eight. Figure it out. That’s how many times I push that button.⁴⁴

The number is 12,288. Stallings’s button pushing would seem to be a nontrivial instance of merely instrumental activity if anything is. But the claim that it is unfree seems reasonable. Examples of this sort suggest, to my mind, that when a merely instrumental activity like button pushing comes to play a significant role in the agent’s life, its unfreedom is not unintuitive.⁴⁵

44 Terkel, *Working*, 159.

45 For other arguments linking autonomy with meaningful work, see Schwartz, “Meaningful Work”; and Roessler, “Meaningful Work.”

Of course, there will be any number of cases between the extremes of my everyday button pushing and Stallings's all-day button pushing. Many ordinary household chores, for instance, probably fall somewhere in the middle. Here, too, I do not find the suggestion of unfreedom unintuitive; and in proportion as the activities are nontrivial, their unfreedom seems to me a genuine and obvious cause for concern. According to one study,

the proportion of working-class women spending more than 9 hours a week in washing and ironing dropped from 61 per cent before the introduction of the washing machine to 24 per cent afterwards.⁴⁶

The work of washing and ironing, like factory labor, is arguably a form of toil: intrinsically unchoice-worthy labor. But if that is true, it does not seem a stretch to think of its reduction—other things equal—as a significant and highly desirable expansion of freedom. Of course, such work, like factory labor, is unequally distributed. This is an aggravating factor, and it partly explains why so many working-class women were spending absolutely so much time in this activity in the first place. But it does not explain the apparent unfreedom of the activity. On the contrary, the unequal distribution of toil is especially troubling precisely because it is a distribution of unfreedom.

Technology and the social reorganization of work can reduce the amount of extreme toil in human life. But perhaps merely instrumental activity, even of a nontrivial sort, cannot be altogether eliminated. Perhaps human life will always contain some measure of nontrivial labor done out of mere necessity. Both the Aristotle of book x of the *Ethics* and the Marx of volume 3 of *Capital* seem to have thought so. Aristotle admits that we cannot spend literally our whole lives in contemplation for its own sake, while Marx for his part suggests that the presence of a “realm of necessity” is an unavoidable fact “in all forms of society and under all possible modes of production.”⁴⁷ If this moderately pessimistic assessment of the human condition is right, then human life must always fall somewhat short of the full and unqualified realization of the ideal of freedom. All we can realistically aspire to is to approximate the ideal as closely as possible. But is there anything wrong with that aspiration? Perhaps Aristotle is on to something when he declares that

we must not follow those who advise us, being men, to think of human things, and, being mortal, of mortal things, but must, so far as we can, make ourselves immortal, and strain every nerve to live in accordance

46 Lynd and Lynd, *Middletown in Transition*, quoted in Gershuny, *Changing Times*, 67, quoted in Goodin et al., *Discretionary Time*, 75n29.

47 Marx, *Capital*, vol. 3, 959. Compare Veltman, *Meaningful Work*, ch. 5.

with the best thing in us; for even if it be small in bulk, much more does it in power and worth surpass everything.⁴⁸

6. CONCLUSION

My aim in this paper has been to put on the table a neglected but attractive approach to understanding external freedom inspired by Aristotle and Marx. To do something autonomously, I have argued, one must do it for its own sake and not merely instrumentally. This necessary condition explains the sense in which doing something autonomously is a matter of doing what one wants. It also explains the distinctive and correlative sense in which externally unfree action is marked by alien necessitation.

As finite creatures, we will always live our lives partly in a realm of instrumental necessity, having to do one thing because we want to do another. There is nothing wrong with this condition *per se*: nothing necessarily wrong, for example, with having to cook in order to eat, or having to eat in order to live. Indeed, as I have mentioned, many instrumental activities are loci of great and irreplaceable intrinsic value. But the point is that our subjection to such necessity is compatible with autonomy only insofar as it amounts to more than *mere* necessity. To be autonomous, instrumental activity must be at the same time an object of intrinsic desire—an unqualified expression of our will in the world.

The Aristotelian-Marxian view is not without its difficulties. It has seemingly counterintuitive implications, since there are many merely instrumental activities that we may not usually think of as unfree. However, I have argued that the view is not as counterintuitive as it seems. In trivial cases, I suggested, the relevant intuitions can be either accommodated or explained away. In non-trivial cases, on the other hand, the view's radical implications seem plausible on reflection. Indeed, I take these implications to be a strength of the view. In addition to shedding light on traditional paradigm cases of coercion and duress, it fruitfully expands our philosophical understanding of autonomy by bringing into view distinctive species of external unfreedom.

Chief among these is the unfreedom of toil. Toil as such is unfree, not because it is necessarily coerced or done from especially dire economic need but because it is an impoverished, intrinsically worthless form of activity done only as a means to further ends. While the political implications of the Aristotelian-Marxian idea are beyond the scope of this paper, this thought might certainly give us pause. Perhaps we will want to reevaluate how well capitalist societies—which have brought us the steam engine and the computer but also

48 Aristotle, *Nicomachean Ethics*, 1177b.

factory labor and scientific management—live up to their promise of protecting and promoting the autonomy of the individual.⁴⁹

Clemson University
pbrixel@clemson.edu

REFERENCES

- Arendt, Hannah. *The Human Condition*, 2nd ed. Chicago: University of Chicago Press, 1998.
- Aristotle. *Nicomachean Ethics*. In *The Complete Works of Aristotle*, edited by Jonathan Barnes, 1729–1867. Princeton, NJ: Princeton University Press, 1984.
- . *Politics*. In *The Complete Works of Aristotle*, edited by Jonathan Barnes, 1986–2129. Princeton, NJ: Princeton University Press, 1984.
- Biggs, Joanna. *All Day Long: A Portrait of Britain at Work*. London: Profile Books, 2015.
- Bratman, Michael E. “Identification, Decision, and Treating as a Reason.” *Philosophical Topics* 24, no. 2 (Fall 1996): 1–18.
- Christman, John. “Autonomy and Personal History.” *Canadian Journal of Philosophy* 21, no. 1 (March 1991): 1–24.
- . “Introduction.” In *The Inner Citadel: Essays on Individual Autonomy*, 3–23. Brattleboro, VT: Echo Point Books and Media, 2014.
- . “Liberalism and Individual Positive Freedom.” *Ethics* 101, no. 2 (January 1991): 343–59.
- . “Procedural Autonomy and Liberal Legitimacy.” In *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, edited by James Stacey Taylor, 277–98. Cambridge: Cambridge University Press, 2005.
- Cohen, G. A. “The Structure of Proletarian Unfreedom.” *Philosophy and Public Affairs* 12, no. 1 (Winter 1983): 3–33.
- Dworkin, Gerald. “Acting Freely.” *Noûs* 4, no. 4 (November 1970): 367–83.
- . “Autonomy and Behavior Control.” *The Hastings Center Report* 6, no. 1 (February 1976): 23–28.
- . “The Concept of Autonomy.” In *The Inner Citadel: Essays on Individual*

49 For extended and valuable feedback that improved this paper, I want to thank Amichai Amit, Anastasia Berg, Brookes Brown, Dan Brudney, Jacob Butcher, A. J. Julius, Ben Lawrence, Gabriel Lear, audiences at the Practical Philosophy Workshop and the Graduate Research Symposium at the University of Chicago, several anonymous reviewers, and especially Anton Ford and Claire Kirwin.

- Autonomy*, edited by John Christman, 54–62. Brattleboro, VT: Echo Point Books and Media, 2014.
- . *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press, 1988.
- Ekstrom, Laura Waddell. “A Coherence Theory of Autonomy.” *Philosophy and Phenomenological Research* 53, no. 3 (September 1993): 599–616.
- Epictetus. *The Discourses of Epictetus*. Translated by Elizabeth Carter and Robin Hard. London: Everyman, 1995.
- Frankfurt, Harry. “Coercion and Moral Responsibility.” In *Essays on Freedom of Action*, edited by Ted Honderich, 63–86. London: Routledge and Kegan Paul, 1973.
- . “The Faintest Passion.” *Proceedings and Addresses of the American Philosophical Association* 66, no. 3 (November 1992): 5–16.
- . “Freedom of the Will and the Concept of a Person.” *Journal of Philosophy* 68, no. 1 (January 1971): 5–20.
- . “Identification and Wholeheartedness.” In *The Importance of What We Care About*, 159–76. Cambridge: Cambridge University Press, 1988.
- Gershuny, Jonathan. *Changing Times: Work and Leisure in Post-Industrial Society*. Oxford: Oxford University Press, 2000.
- Goodin, Robert E., James Mahmud Rice, Antti Parpo, and Lina Eriksson. *Discretionary Time: A New Measure of Freedom*. Cambridge: Cambridge University Press, 1998.
- Hurka, Thomas. “Why Value Autonomy?” *Social Theory and Practice* 13, no. 3 (Fall 1987): 361–82.
- Jaeggi, Rahel. *Alienation*. Translated by Frederick Neuhouser and Alan E. Smith. Edited by Frederick Neuhouser. New York: Columbia University Press, 2014.
- Killmister, Suzy. *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. Milton Park, UK: Routledge, 2018.
- Korsgaard, Christine M. “Two Distinctions in Goodness.” *Philosophical Review* 92, no. 2 (April 1983): 169–95.
- Kristinsson, Sigurdur. “The Limits of Neutrality: Toward a Weakly Substantive Account of Autonomy.” *Canadian Journal of Philosophy* 30, no. 2 (June 2000): 257–86.
- List, Christian, and Laura Valentini. “Freedom as Independence.” *Ethics* 126, no. 4 (July 2016): 1043–74.
- Lovett, Francis N. “Domination: A Preliminary Analysis.” *Monist* 84, no. 1 (January 2001): 98–112.
- Lynd, Robert S., and Helen Merrell Lynd. *Middletown in Transition: A Study in Cultural Conflict*. New York: Harcourt Brace, 1937.
- Marx, Karl. *Capital: A Critique of Political Economy*. Vol. 3. Translated by David

- Fernbach. London: Penguin, 1981.
- Mele, Alfred R. *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press, 1995.
- Neely, Wright. "Freedom and Desire." *Philosophical Review* 83, no. 1 (January 1974): 32–54.
- Nozick, Robert. "Coercion." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, edited by Sidney Morgenbesser, Patrick Suppes, and Mary Terrell White. New York: St. Martin's Press, 1969.
- Oshana, Marina. "Personal Autonomy and Society." *Journal of Social Philosophy* 29, no. 1 (Spring 1998): 81–102.
- . *Personal Autonomy in Society*. Aldershot, UK: Ashgate Publishing, 2006.
- Pettit, Philip. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press, 1997.
- Plamenatz, J. P. *Consent, Freedom and Political Obligation*. London: Oxford University Press, 1938.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.
- Roessler, Beate. "Meaningful Work: Arguments from Autonomy." *Journal of Political Philosophy* 20, no. 1 (March 2012): 71–93.
- Schwartz, Adina. "Meaningful Work." *Ethics* 92, no. 4 (July 1982): 634–46.
- Tannenbaum, Julie. "Categorizing Goods." In *Oxford Studies in Metaethics*, vol. 5, edited by Russ Shafer-Landau, 257–94. Oxford: Oxford University Press, 2010.
- Taylor, Charles. "What's Wrong with Negative Liberty." In *Philosophical Papers*, 211–29. Cambridge: Cambridge University Press, 1985.
- Taylor, James Stacey. "Autonomy, Duress, and Coercion." *Social Philosophy and Policy* 20, no. 2 (July 2003): 127–55.
- . *Practical Autonomy and Bioethics*. New York: Routledge, 2009.
- Terkel, Studs. *Working: People Talk about What They Do All Day and How They Feel about What They Do*. New York: Random House, 1972.
- Thalberg, Irving. "Hierarchical Analyses of Unfree Action." *Canadian Journal of Philosophy* 8, no. 2 (June 1978): 211–26.
- Van Parijs, Philippe. *Real Freedom for All: What (If Anything) Can Justify Capitalism?* Oxford: Oxford University Press, 1995.
- Velleman, J. David. *The Possibility of Practical Reason*. Oxford: Oxford University Press, 2000.
- Veltman, Andrea. *Meaningful Work*. Oxford: Oxford University Press, 2016.
- Watson, Gary. "Free Agency." *Journal of Philosophy* 72, no. 8 (April 1975): 205–20.
- Zimmerman, David. "Making Do: Troubling Stoic Tendencies in an Otherwise Compelling Theory of Autonomy." *Canadian Journal of Philosophy* 30, no. 1 (March 2000): 25–54.

TOWARD A PERCEPTUAL SOLUTION TO EPISTEMOLOGICAL OBJECTIONS TO NONNATURALISM

Preston Werner

STANCE-INDEPENDENT nonnaturalist moral realism is subject to two related epistemological objections.¹ First, there is the metaethical descendant of the Benacerraf problem.² Second, there are evolutionary debunking arguments.³ Standard attempts to solve these epistemological problems have not appealed to any particular moral epistemology. This makes sense: a response to these particular epistemic concerns that is otherwise epistemologically neutral is preferable to one only available to those willing to take on other epistemological commitments. On the other hand, the focus on these epistemologically neutral responses leaves many interesting theoretical stones unturned. Exploring the ability of particular theories in moral epistemology to handle these difficult epistemological objections can help illuminate strengths or weaknesses within these theories themselves, as well as opening up potentially unexplored avenues for responding to deeply entrenched concerns about our epistemic access to the moral properties.

This paper is a case study in the latter kind of project. I assess the prospects of a perceptualist model of moral knowledge for responding to epistemological arguments against non-skeptical moral realism. I argue that *Moral Perceptualism* (MP), as I will call it, has powerful responses to these objections that are not available to other moral epistemologists. Furthermore, the

- 1 Shafer-Landau defines *stance independence* as the claim that “the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective” (*Moral Realism*, 15). “Nonnaturalism” is also subject to different understandings. I will not define “nonnaturalism” precisely for the purposes of this paper, except to say that on a nonnaturalist view, the moral facts are not identical or reducible to natural facts. (This conflicts with the epistemological characterization of “nonnaturalism” that Shafer-Landau favors.)
- 2 Benacerraf, “Mathematical Truth.”
- 3 The two most famous examples of evolutionary debunking arguments are those of Joyce, *The Myth of Morality*, ch. 4, and *The Evolution of Morality*; and Street, “A Darwinian Dilemma for Realist Theories of Value.”

uniquely perceptualist responses are arguably *more* compelling than other approaches to the epistemic objections that have cropped up in the literature. The upshot is that *if* some version of MP is correct, then the realist has less to fear from Benacerraf and evolutionary debunking–style epistemological objections. Insofar as one is already a committed realist, then, this provides some indirect support for MP.

The structure of the paper is as follows. In section 1, I discuss two important—and what I take to be the two most powerful—ways of understanding the epistemic constraint that nonnaturalists’ moral beliefs cannot meet. After a brief overview of MP in section 2, in section 3, I clarify and consider the claim that nonnaturalists cannot explain our *epistemic access* to nonnatural facts. I argue that MP can meet the epistemic access constraint in a way that appears unavailable to traditional *a priori* nonnaturalist moral epistemologies. This requires a slight digression to discuss the causal nature of perceptual experience. In section 4, I consider a second way of understanding the epistemological objection to nonnaturalism—the idea that nonnaturalists cannot illustrate an *explanatory connection* between our moral beliefs and nonnatural facts. Here again I claim that the proponent of MP is better placed to meet the challenge than its *a priori* counterparts. Finally, in section 5, I sum up what I take myself to have shown.

1. EPISTEMIC PRINCIPLES BEHIND SKEPTICISM ABOUT NONNATURAL NORMATIVE FACTS

Most philosophers agree that there is something epistemically questionable about nonnatural moral knowledge, given the genealogy of our moral beliefs and the metaphysical status of those facts. These facts are alleged to undercut some necessary condition on the possibility of knowledge about some domain. Just what is this necessary condition? Different authors have proposed different ideas. Here are what I take to be the two most powerful:

Epistemic Access: In order for our beliefs about some domain *D* to constitute knowledge, we must have epistemic access to the *D*-facts.⁴

Explanatory Connection: In order for a belief *B* to constitute knowledge that *P*, *P* must play an ineliminable role in an explanation about why *B* exists.⁵

4 Benacerraf, “Mathematical Truth” (on one reading); and Timmons, “On the Epistemic Status of Considered Moral Judgments.”

5 See Jenkins, “The Analysis of Knowledge”; Woods, “Mathematics, Morality, and Self-Effacement”; and Lutz, “The Reliability Challenge in Moral Epistemology.”

These two principles are a long way from exhausting the possibilities.⁶ My claim, which I will not defend here, is that these principles represent two of the most powerful yet non-question-begging grounds for raising skeptical worries about nonnaturalist normative realism. Both of them can be met in the case of epistemically uncontroversial domains such as knowledge of ordinary objects and scientific knowledge, but not in the case of moral facts if those facts are construed nonnaturalistically. The key here is to find a principle that genuinely puts nonnaturalist moral knowledge in doubt without overgeneralizing to something more closely resembling a global skepticism. And of course, the principle should itself be a plausible, independently motivated constraint on knowledge of some domain.

Before turning to a discussion of why and how a perceptualist moral epistemology can meet these two principles, a brief explication of MP (as I will understand it) is necessary.

2. MORAL PERCEPTUALISM: AN OUTLINE

MP, as I will use the notion here, consists of two substantive claims. First, MP is a version of *Ethical Foundationalism* (EF).

EF: Most ethical agents have at least some non-inferentially justified first-order ethical beliefs.

As stated, EF is just the claim that foundationalism—understood in the epistemologist’s sense—is true of the structure of at least some ethical beliefs, and that some ethical beliefs are members of the set of foundational beliefs. However, EF does not entail that the non-inferentially justified ethical beliefs are grounded in *intuitions*, whatever those turn out to be.

The second claim that constitutes MP is *Ethical Empiricism* (EE):

EE: The non-inferential justification of first-order ethical beliefs is grounded in perceptual experiences that represent the instantiation of evaluative properties.

6 Two other principles often raised in this context have to do with whether the nonnaturalist can explain our reliability with respect to the moral facts. On one reading, the claim is that nonnaturalists cannot explain our actual reliability in a non-question-begging way (see, e.g., Vavova, “Debunking Evolutionary Debunking”). On another reading, the claim is that nonnaturalists cannot explain how we could possibly be reliable with respect to the non-natural facts (see Enoch, *Taking Morality Seriously*). I think that others have convincingly argued that neither of these principles will make for a powerful but non-question-begging challenge to nonnaturalism, so I will not discuss them in detail here. For discussion, see Vavova, “Debunking Evolutionary Debunking”; Jonas, “Access Problems and Explanatory Overkill”; and Baras, “Our Reliability Is in Principle Explainable.”

EE says that non-inferentially justified ethical beliefs are justified in the same way as other perceptual justification. The basic picture is as follows. Under certain circumstances, evaluative properties figure in the contents of perceptual experience. Furthermore, at least sometimes, the evaluative properties that figure in the contents of perceptual experience can provide non-inferential justification for beliefs about the instantiation of evaluative properties. This is compatible with the claim that sometimes evaluative perceptual experiences fail to non-inferentially justify. First, there may be defeaters for the justification that an evaluative perceptual experience would otherwise provide. Second, some evaluative perceptual experiences may be epistemically dependent in the sense that they cannot provide justificatory force independently of some prior justified evaluative belief. MP only claims that, in at least some circumstances, neither of these things holds. When they do not, an evaluative perceptual experience can ground a non-inferentially justified moral belief.⁷

3. EPISTEMIC ACCESS AND MP

3.1. What Is Epistemic Access?

Let us turn now to epistemic access. One worry is that epistemic access is itself a technical notion that is often not given further characterization. A complete analysis of the notion of “epistemic access” cannot be given here. But let me say a little bit about the general idea. Epistemic access, as I understand it, involves establishing that some (metaphysical) relation holds between the *D*-beliefs and the *D*-facts that can ground positive epistemic status. Epistemic access is both weaker and stronger than a notion such as reliability. It is weaker because it does not require accuracy—an epistemic access relation can hold without a majority of beliefs being true. But it is stronger because it requires some such relation to hold; even beliefs that are reliably true (because for example their contents are necessary) may not meet an epistemic access condition. Finally, note that the sort of relation that underwrites epistemic access need not be causal. Consider a few noncausal examples.⁸

Introspective Access: Though the reliability of introspection has been questioned, it is plausible that we have some special access to our own mental states, however fallible it may be.⁹ I take it that even though

7 It is perhaps worth noting that MP is compatible with a number of views on the metaphysics of moral properties, as well as with a number of views in normative ethics.

8 None of these are going to be completely uncontroversial; they are only meant to be illustrative.

9 See, e.g., Schwitzgebel, “Introspection,” sec. 4.

introspection involves underlying causal brain processes, it is from an epistemic standpoint a different kind of access than causal access.¹⁰ On a traditional sort of model of introspection, our introspective beliefs are responsive to the facts they are about via a relationship of direct acquaintance.

Conceptual Access: A certain philosophical school of thought claims that we can learn a lot about the concepts we possess competently by conceptual analysis, which may involve reflecting on how we would apply them in various scenarios.¹¹ This is one potential explanation and defense of analytic knowledge (assuming that there are analytic truths).¹² Again, though this kind of *a priori* reflection would be underwritten by causal (and possibly also introspective) processes, the *access* in question is not causal or introspective, because of the nature of the truths in question. On this view, analytic truths are not causally related to us, nor are they merely facts about our own mental states.¹³

Constitutive Access: We have constitutive access to a truth *t* when something about our coming to believe *t* is partially constitutive or provides evidence for what is partially constitutive of its being the case that *t*. Arguably, many beliefs about response-dependent properties involve constitutive access.¹⁴ Suppose that something is beautiful iff it is believed to be beautiful by all/many/some normal adult human beings.¹⁵ A normal adult human being comes across a Chuck Close painting and comes to believe that it is beautiful. She has constitutive

- 10 I do not want to take a stand on whether, or to what extent, introspection should be subsumed under the category of causal access. I include it in the list because it seems to have been thought to be epistemically distinct in some special way by many philosophers, and my intention here is only to give a list of possibly different forms of epistemic access. See Schwitzgebel, "Introspection."
- 11 Jackson, *From Metaphysics to Ethics* and "Locke-ing onto Content"; Audi, "Skepticism about A Priori Justification"; Russell, *The Problems of Philosophy*.
- 12 Robert Audi, for example, believes that at least some substantive moral knowledge is conceptual, in the sense that the wrongness of certain actions is "contained in" the moral concepts alone. See *The Good in the Right* and "Intuition, Inference, and Rational Disagreement in Ethics."
- 13 For a book-length defense of analyticity, see Russell, *Truth in Virtue of Meaning*.
- 14 As with introspection, I do not intend to take a stand on the epistemology of response-dependent properties. My intention here is only to give a list of possibly different forms of epistemic access.
- 15 There are many complications I am ignoring here, not the least of which is how to define "normal" in a noncircular way.

access to the fact that the painting is beautiful insofar as her belief is partially constitutive of that fact.

This list is not meant to be exhaustive. What is important here is that bearing the kind of substantive relation that can meet *epistemic access* involves a plurality of options, but that this requires more than coincidental accuracy.

3.2 Epistemic Access and MP

I now turn to considering whether MP is better equipped to meet the epistemic access challenge. I argue that it is, although there remain some wrinkles to be ironed out. It may initially seem like the proponent of MP has a simple but complete answer: our access is perceptual. We perceive the nonnatural moral facts and, after all, there is nothing implausible about claiming that perception can provide access to mind-independent properties. So, according to MP, the nonnatural facts are epistemically accessible.

This response may sound too good to be true. While it is true that, according to MP, our access to the nonnatural facts is perceptual, just how this is possible is much more unclear than it is in the case of tables, cats, or shapes.¹⁶ This is because, unlike tables, cats, and shapes, the nonnatural properties are widely thought to be noncausal. It seems as though the quick and dirty response given above just pushes the problem of epistemic access back a step. The skeptic can now ask: How could we have perceptual access to a causally inefficacious property, when perception is essentially a causal relationship?¹⁷

The proponent of MP could deny that perception is essentially causal, but without further motivation, this would appear *ad hoc*. She could also deny that nonnatural properties are causally inefficacious, but that would raise its own problems.¹⁸ It might seem that these two options, both unpalatable, are the only routes available for the proponent of MP. And so it may look as if, initial appearances aside, MP is not well placed to provide an adequate account of our epistemic access to the nonnatural properties.

What we have, then, is a seemingly inconsistent triad:

Causally Inefficacious (CI): Nonnatural moral properties are causally inefficacious.

Perceptual Access (PA): We have epistemic access to nonnatural moral properties through perception.

16 Schroeder raises a similar worry for the view that desires are appearances of the good (“How Does the Good Appear to Us?” sec. 4).

17 See McGrath, “Causation By Omission.”

18 For this strategy, see Oddie, *Value, Reality, and Desire*.

Causal Condition on Perception (CCP): Perception is an essentially causal relation.

As we have just seen, we should be reluctant to give up either CI or CCP. Rejecting PA appears to be the only option left. But this is not right. Contrary to initial appearances, the triad above is not inconsistent. We can simultaneously accept CI, CCP, and PA—or so I presently argue.

The appearance of inconsistency arises because perception is essentially causal, while nonnatural properties are noncausal. However, once we focus on what precisely CCP says (and does not say), it becomes clear that CCP is actually compatible with moral perception, and thus compatible with PA, even if the moral properties are noncausal. To see this, notice that the proponent of MP need not—and in fact *should not*—deny that moral perception is causal. If Norma perceives that Tibbles's being lit on fire is bad, and this perception is not hallucinatory, she surely must stand in some causal relation to Tibbles. Thus, Norma's perception is essentially causal; CCP is met. And yet her perceptual experience represents *badness*, a causally inefficacious property; so we have not given up CI either.

Sarah McGrath, a moral perceptualist of a sort, has bolstered this claim by a kind of partners in innocence argument.¹⁹ Imagine a non-skeptical Humean about perception, who argues that we cannot visually perceive anything other than two-dimensional color splotches. What should such a theorist say about our knowledge of trees, tables, and chairs? If she does not want to fall into skepticism, she has to say one of two things. Either our knowledge of these objects is somehow *a priori*, or we can gain perceptual knowledge of things even if we do not perceptually experience them. Since the former idea is absurd (“there is a tree in this room” is surely not *a priori*), we should think we can gain perceptual knowledge without perceptual experience. But then the moral epistemologist can say the same thing about moral knowledge, and such a move is not at all *ad hoc*.

I think McGrath is onto something here, but there are a couple of things that necessitate further discussion of this point. First, McGrath's view is that we can gain non-inferential moral knowledge on the basis of perception *despite* the fact that we cannot have perceptual experiences with moral content. And this feature of her view is a requirement for her response to this objection to work. This conflicts with my reading of MP discussed in section 2. Second, there may be a reasonable fear here that there is some faulty philosophy of perception going on in the background—how could we have non-inferential moral knowledge

19 McGrath is a moral perceptualist in an important sense—she thinks that moral beliefs can be justified on the basis of perceptual experience alone. However, unlike the “moral perceptualism” defended in this paper, McGrath rejects the idea that moral properties are part of the content of perceptual experience. See McGrath, “Moral Perception and its Rivals.”

from wholly nonmoral content without some bridge principle? I have some sympathy for this cautiousness. So it is worth a small digression to say a bit about the underlying philosophy of perception issues going on beneath the surface. Once they are brought out, it becomes clear that even the proponent of MP in my preferred sense can allow for moral experiences compatible with CCP and CI.

3.3. *Interlude: Just What Is Essentially Causal about Perception?*

There is a vast literature in the philosophy of perception concerning what properties figure in the contents of perceptual experience. Call *Conservatism* the view that only low-level properties—such as shapes, colors, and tones—are represented. Call *Liberalism* the view that some high-level properties—such as natural kinds, artifacts, and relations—can also be represented. Conservatives and liberals disagree about what properties feature in perceptual experience, but they widely agree that perception is an essentially causal relationship.²⁰

I cannot adjudicate the conservative/liberal dispute here. But it seems safe to assume that MP is only going to be even initially plausible to liberals. Assuming that moral properties are high-level properties, conservatives are going to reject MP from the get-go. In what follows, I assess how best to understand the essentially causal nature of perception from within a liberal framework. In the bigger picture, this is a contentious assumption. But since proponents of MP are already committed to liberalism, it is a safe assumption to make in this context. The idea, then, is to home in on the essentially causal nature of perception by considering some causally unique cases of properties thought to be perceivable by liberals about perceptual experience.

Consider one natural way to understand the causal constraint on perception:

Strict CC: Necessarily, if a property F is part of the contents of S 's perceptual experience e , then F (or the fact that F is instantiated) is at least partially causally responsible for e .

Strict CC is a relatively robust causal constraint on perceptual representation. But it is also an initially intuitive way of characterizing the causal nature of perception in a precise way. Nevertheless, I now argue that Strict CC should be rejected by liberals about perceptual experience. I will argue this by considering three sorts of properties that liberals have defended as perceivable that could not be, if Strict CC were true: absences, Gibsonian affordances, and the mental states of others. I consider each in turn.

Many liberals have recently argued that perceptual experience extends beyond the representation of positive properties to the representation of what we

20 Though not universally—see Snowdon, “Perception, Vision, and Causation.”

can call *absence properties*. For example, you may perceive a gap in an otherwise predictable pattern of coins arranged on a table, the holes in a slice of Swiss cheese, darkness inside a cave, or the sound of silence.²¹ Suppose that these liberals are right—that we do perceive at least some absence properties. It is unclear whether this is compatible with Strict CC, since it is unclear that the *lack* of something can figure in a genuine causal relationship. It is plausible that silence, for example, does not involve the existence of some causal property, but rather the lack of any causally efficacious property of a certain sort. So, while it is not uncontentious, the perception of absence properties does provide some *prima facie* reason to favor a less robust causal constraint on perception than Strict CC.

A second set of properties that appears to conflict with Strict CC includes what I will call *affordance properties*. The idea of affordances in perceptual experience goes back to the psychological research of James J. Gibson, but it has also been the subject of quite a bit of recent work in the philosophy of perception.²² In Gibson's words, affordances are properties that tell an animal what an environment "offers the animal, what it *provides* or *furnishes*, either for good or ill."²³ Others—both philosophers and psychologists—following in Gibson's footsteps have attempted to refine the idea of affordance properties in various ways.²⁴ But paradigmatic instances of affordance properties should illustrate the idea clearly enough for present purposes. For example, an animal's prey may be seen as *to-be-killed*, a cup as *able-to-be-picked-up*, and the liquid in the cup as *drinkable*.²⁵ In brief, affordance properties relate agents and their abilities to the environment. They represent something like potential actions.

Though affordance properties are surely grounded in causal properties (for example, the structure of the cup grounds or constitutes its ability to be picked up), they are arguably not themselves causal. However, according to at least many

21 See, respectively, Farennikova, "Perception of Absence and Penetration from Expectation," 2; Casati and Varzi, *Holes and Other Superficialities*, 156–58; Sorensen, *Seeing Dark Things*, chs. 10, 14, and "Hearing Silence"; Soteriou, "The Perception of Absence, Space, and Time"; Phillips, "Hearing and Hallucinating Silence"; and Simon and Garfunkel, "The Sound of Silence."

22 Gibson, "The Theory of Affordances," and *The Ecological Approach to Visual Perception*; Chemero, "An Outline of a Theory of Affordances"; Prosser, "Affordances and Phenomenal Character in Spatial Perception"; Nanay, "Action-Oriented Perception"; and Siegel, "Affordances and the Contents of Perception."

23 Gibson, *The Ecological Approach to Visual Perception*, 127.

24 See, e.g., Reed, *Encountering the World*; and Chemero, "An Outline of a Theory of Affordances."

25 Affordance properties appear, then, to come in two levels of strength—some features of objects render things possible, while others render things as appearing (practically) necessary. For discussion of this point, see Siegel, "Affordances and the Contents of Perception."

psychologists and philosophers, affordance properties are perceivable.²⁶ Insofar as this is right, it casts doubt on Strict CC, since the perception of affordance properties is incompatible with it. In short, affordance properties give us further reason to favor a less robust causal constraint on perception than Strict CC.

Finally, consider perception of the mental states of others. Many philosophers of perception and mind have recently argued that we can literally perceive the affective states of others.²⁷ Rowland Stout, for example, argues that we can “literally perceive someone’s anger” in the sense that this perception is non-inferential.²⁸ The causal efficacy of mental states is one of the thorniest issues in philosophy. But, as far as I know, no one arguing against the perception of mental states has claimed that the perception of these states hinged on this controversy. Appeals in favor of the claim that we can perceive these states are generally phenomenological and empirical (appealing to modules in the brain dedicated to “mindreading”), not to the causal efficacy of these states. So it seems as though at least many liberals should be friendly to the perception of the mental states of others, regardless of their direct causal efficacy.

If a broadly liberal view of perceptual content is correct, it seems like Strict CC is not the right way to understand the causal constraint on perception. However, given the consensus that there is some causal constraint on perception, some weaker constraint must hold. Unfortunately, without taking controversial stands on the cases above (and others), a full account cannot be explicated and defended here. However, if any of the properties discussed above are perceivable, something *at least as weak* as the following must hold:

Weak CC: Necessarily, if a property *F* is part of the contents of *S*’s perceptual experience *e*, then either (a) *F* or (b) some property (or set of properties) *G* that perceptually grounds *F* is at least partially causally responsible for *e*.²⁹

Depending on what one says about the cases above, Weak CC may remain too strong to be an accurate causal constraint on perception. And notice also that

26 This claim is far from uncontentious. But so far as I know, no one has rejected the perceivability of affordances on the grounds that they are not causal.

27 See, e.g., Green, “Perceiving Emotions;” Stout, “Seeing the Anger in Someone’s Face;” and McNeill, “On Seeing That Someone Is Angry.”

28 Stout, “Seeing the Anger in Someone’s Face,” 29.

29 By “perceptually ground” here, I mean the low-level perceptual properties upon which the high-level perceptual property is perceived. For example, a perceptual experience of a table is perceptually grounded in the perceptual representation of shades of brown, edges, etc. Of course, a complete theory of perceptual grounding would require more to be said, but this lies far outside the scope of this paper.

Weak CC is a necessary but not sufficient condition for perceptual eligibility. I hope to have established that the liberal about perceptual experience should favor something at least as weak as Weak CC, independent of any consideration of the perception of moral properties.

We have now seen that this attempt to single out moral properties using the causal constraint on perception is not so simple. The perception of causally inefficacious properties is compatible with the causal constraint on perception, properly construed, as long as those properties are related to causally efficacious properties in the right sort of way. Given some plausible causal constraints on perception, nonnatural properties will be perceivable after all.³⁰ Given that perceptual access is paradigmatically epistemic access, epistemic access to moral properties is possible if we endorse MP.

4. MORAL PERCEPTUALISM AND EXPLANATORY CONNECTIONS TO THE FACTS

4.1. *How to Think about Explanatory Connections*

The second potential condition on nonnatural justification or knowledge that I want to consider has to do with another kind of connection between our moral beliefs and the moral facts. According to this principle, if our moral beliefs are to constitute knowledge, the (nonnatural) moral facts must play a role in an explanation about why we have them. Compare this principle to another epistemic principle raised against nonnaturalism in this context:

Explanation of Reliability (ER): If we have no explanation of the reliability of our beliefs about some domain *D*, our justification for beliefs about *D* is defeated.³¹

For reasons that others have raised, I think this understanding of the challenge is misguided.³² In any case, if this is the strongest plausible challenge that can

30 There is a deep non-epistemological problem lurking in the background here: not only does the moral realist need to establish that we can in principle perceptually represent noncausal properties, but she will also need to provide a theory of the fixing of perceptual content that does not require a causal connection between the representation and the property represented. Defending a moral realist friendly theory of content fixing is a nontrivial task, and I cannot hope to achieve that task here. But for some approaches that seem promising, see Werner, "Getting a Moral Thing into a Thought"; and Schroeter and Schroeter, "The Generalized Integration Challenge in Metaethics." Thanks to Bar Luzon for helping me see how important this issue is.

31 For readings about the epistemological objections to nonnaturalism along these lines, see Vavova, "Debunking Evolutionary Debunking"; Crow, "The Mystery of Moral Perception," 19–21; and Schecter, "Is there a reliability problem for logic?," Section 6.

32 Jonas, "Access Problems and Explanatory Overkill."

be raised against nonnaturalist epistemology, then nonnaturalists have nothing to fear from the debunker.³³ However, I raise this principle here merely to distinguish *Explanatory Connections* (EC) from it. EC is in one sense easier and in one sense more difficult to meet than ER. EC is easier to meet because a belief can in principle be explanatorily connected to a fact without being reliable.³⁴ On the other hand, EC is harder to meet because reliability alone, even explained reliability, does not guarantee an explanatory connection. This is because someone could have a belief forming method that is coincidentally and robustly reliable without having anything to do with the domain itself. Consider some method that reliably results in a belief that *P*, where *P* is some necessary truth. The method could be totally arbitrary and have nothing to do with *P*, but we would still have an explanation of reliability. The belief would not, though, meet EC.

So how can EC be met, if not merely by reliability? Something like this principle has been most recently and powerfully defended by Matthew Lutz.³⁵ After pointing out that a causal constraint on knowledge is subject to several counterexamples, Lutz explains:

This is why [EC] does not refer to *causal* connections but rather to *explanatory* connections. . . . If we reject the notion of a “final cause” as being genuinely explanatory—as is common, post-Darwin—we can identify three different kinds of explanatory relations: formal explanation, material explanation, and causal explanation. . . . The statue exists *because* the lump exists, in the form of a statue. The window breaks *because* I threw the rock.³⁶

As Lutz here points out, EC can be met by noncausal factors. For example, constitutive explanations can connect two facts, such as the connection between the statue’s existence being explained by the lump’s existing in a particular form. The strength of the stone can be explained by its material composition. And of course there may be other explanatory connections as well. What is important here is that the EC condition is not a causal condition in disguise.

Elsewhere, Lutz provides a more general theory as to the kind of EC that fits best with a set of beliefs about some domain. I do not want to take a stand here

33 Clarke-Doane, “Debunking and Dispensability,” ch. 6. See also Enoch, *Taking Morality Seriously*.

34 If this seems unacceptably weak, not to worry; see the discussion of “Well-Explained Belief” below. My aim in section 4.2 is to show that MP can meet even the stronger condition of Well-Explained Belief.

35 Lutz, “The Reliability Challenge in Moral Epistemology.”

36 Lutz, “The Reliability Challenge in Moral Epistemology,” 303.

on whether this is the precisely correct understanding of EC, but it will be useful to have the concept of a *Well-Explained* belief for what follows. As Lutz argues,

S's belief that *P* is Well-Explained if and only if S's belief that *P* is the product of a reliable belief-forming method, *M*, and there is an explanatory connection between the fact that S is using *M* and the fact that *M* is reliable.³⁷

Is showing that a set of beliefs about some domain are Well-Explained necessary for meeting EC, and thus necessary for responding to the skeptic about nonnaturalism? One complication here is that being Well-Explained could be a condition on knowledge, but nonetheless the burden rests on the skeptic to show that this condition *could not* be met, rather than on the non-skeptic to show that it in fact is. Consider raising a skeptical worry about vision against a philosopher living in a time before vision was well-understood. It would be unfair to demand that she must give a story about why our visual beliefs are Well-Explained in order for her to go on trusting her vision. It just needs to be the case that there *is* an explanation that connects our method *M* and facts in the domain in question; we do not need to understand or grasp that explanatory connection, even as theorists. Put another way, a belief's being Well-Explained is an externalist condition that needs to be met for a domain to be non-skeptical, not an internalist one. It can be met without our grasp of an explanation as to how, compatible with the rejection of skepticism.

However, the concept of a belief's being Well-Explained can still be useful for assessing EC. If we have reason to think that our beliefs about some domain could not, even in principle, be Well-Explained, that would cast serious doubt on a non-skeptical account of that domain. A plausible story about our (non-natural) moral beliefs being Well-Explained would defang an epistemological argument against nonnaturalism based in EC. To reiterate, I am not claiming that beliefs about some domain need to be Well-Explained in order to meet EC; instead, I am claiming that being Well-Explained, since it is a particular way of providing an explanatory connection between beliefs and the facts they are about, is sufficient for meeting EC.

In any case, my purpose here also is not to defend EC as a necessary (or sufficient) condition on knowledge. So even if *a priori* moral epistemologies cannot meet the condition, they could defend their epistemic credentials by arguing that the view is false. But let us set that possibility aside and from here forward assume that something like EC is true. If it is at least plausible, and MP is in better standing than its *a priori* rivals, that would provide some reason to

37 Lutz, "The Case for Moral Skepticism," 68.

favor MP, other things being equal. So let us turn to a consideration of how MP can meet this condition.

4.2. Explanatory Connection and Moral Perception

Consider first the story about how EC will be met for ordinary, boring perceptual beliefs. Norma has the belief that there is a book on the desk. Her belief was caused by her visual experience of a book being on the desk. Norma's belief is well-explained: her belief is the product of a reliable belief-forming method (visual experience) and there is an explanatory connection between her use of visual experience and the holding of the facts that she is experiencing. She accepts her visual experience in this case (implicitly) because it has a long history of getting things right. Put another way, if she were in conditions in which she had often found her visual experience had gotten things wrong in the past, then she would have been more hesitant to form a belief on the basis of her visual experience. So there is a correlation between Norma's willingness to form beliefs on the basis of her visual experience and the facts that she is visually experiencing. She believes *because* of the facts in question. EC is met.

Now turn to the moral case. Suppose Norma has the belief that *the cat's suffering is bad*. And suppose furthermore, in accordance with MP, that her belief is based off of a visual experience of a cat on fire. The relevant question here is whether there is an explanatory connection between the fact that Norma trusts her visual experience and her visual experience's reliability. Initially, at least, it appears that the answer is yes, for the same reason as above. As long as Norma is a responsible moral agent, she will not trust her visual experience in poor visual conditions, or conditions in which her perception of moral properties may be unreliable. So, as with above, there may be a correlation between Norma's willingness to form (moral) beliefs on the basis of her visual experience and the accuracy of her visual experience. It looks like EC is met.

This is too quick. It is too simplistic to think of a belief-forming process such as visual experience as reliable or unreliable *simpliciter*. The reliability or unreliability of a particular belief-forming process depends not just on the process, but on the process relative to the domain in question. For example, an electromagnetic field (EMF) meter is reliable with respect to the detection of an EMF, but it would be silly to infer from this that ghost hunters are forming reliable beliefs when they take EMF meters to convey information about the presence of ghosts. An EMF meter is a reliable method for the domain of EMF information, but unreliable for the domain of ghost information.³⁸ Similarly,

38 The example comes from Gibilisco, "Theories of Properties and Ontological Theory-Choice," 107–8. Thanks to Christopher Gibilisco for pressing me on this point. For an early statement of this idea, see Nozick, *Philosophical Explanations*, sec. 3.1.

the debunker can claim that visual experience is reliable with respect to notebook information, but unreliable with respect to moral information. Or, to state this more carefully, since the debunker wants to remain neutral on the question of reliability: there is an explanatory connection between perceptual beliefs about notebooks and the reliability of perceptual experience of notebooks, but no such explanatory connection between perceptual beliefs about moral properties and the reliability of perceptual experience of moral properties.³⁹

Before responding to this worry, let us get clear about exactly what the defender of an anti-skeptical MP owes the debunker. The debunker cannot demand process-independent proof of the reliability of perceptual experience for detecting moral properties. Such a requirement would lead to a near-universal skepticism, not just about morality, but about all perceptual beliefs. So here the debunker must be making a more restricted claim, that, *even assuming* the reliability of perceptual experience with respect to detecting moral properties, it will still be the case that there is no explanatory connection between perception's reliability and our tendency to trust it on moral matters. In the ordinary boring case of perceptual belief, we have a long evolutionary story about why human beings and other animals' trust of perceptual experiences of ordinary objects selects for accuracy. Not so for moral perceptual experiences. Even if such experiences are reliable, they were not selected for their accuracy. So there is a deep explanatory connection in the ordinary object case between the reliability of the method and our use of it. No such connection exists in the moral case.

To address this, it will help to make use of a distinction first incorporated in the metaethics literature by Andreas Mogensen—the distinction between *proximate* and *ultimate* explanations.⁴⁰ A proximate explanation is an explanation of why some particular individual has some trait by way of appealing to their particular life history, while an ultimate explanation appeals to a species' evolutionary history. As Mogensen stresses, these explanations are not competing, but complementary:

Imagine that insects in one species, S_1 , have a certain pattern of colouration that serves as camouflage: it resembles the surrounding foliage. Natural selection has favoured this pattern of colouration because it allows the insects to avoid predators. Suppose the pattern of colouration arises because juveniles eat a certain kind of moss during a critical development period. However, the fact that the juveniles have this diet is irrelevant in explaining why having this kind of colouration confers greater

39 Vavova, "Evolutionary Debunking of Moral Realism."

40 The distinction is originally from Mayr, "Darwin's Biological Work."

relative fitness: the colouration would be equally advantageous if it came about as a result of a different set of developmental factors.⁴¹

Notice that the proximate/ultimate distinction here illustrates that explanatory connections can hold even when there is no deep evolutionary story about why a particular process is reliable. There is an explanatory connection between the food that a juvenile *S1* eats and their pattern of coloration, despite the fact that, from an evolutionary standpoint, the fact that this particular mechanism of generating the coloration rather than some other is a coincidence. And it should be flagged that this proximate conception of an explanation is intuitively enough to meet the EC constraint as well. Even though evolution does not select for agents who can engage in chemistry, for example, this does not undermine the claim that there is an explanatory connection between a chemist and their chemistry beliefs.⁴² Requiring a deep evolutionary explanatory connection, or at least a direct one, between any set of beliefs about a domain and the reliability of the process that underwrites those beliefs would commit the debunkers to an overgeneralization of their arguments to any domain of beliefs that lack a cognitive mechanism directly evolutionarily selected for.⁴³

With all of this said, what matters for the proponent of perceptual moral knowledge is that there is at least a proximate explanatory connection between the reliability of moral perceptual experience and its use in forming moral beliefs. MP claims that perceptual experience can represent moral properties. There are relevant and vexed questions here about how representational content gets fixed. I cannot hope to even begin to scratch the surface here.⁴⁴ But what can be said is this: on at least many plausible theories of how representational content gets fixed, the content-fixing relation will guarantee an explanatory connection between a property *F* and perceptual representations of *F*. And that will in turn provide an explanatory connection between the fact that a subject is forming beliefs on the basis of perceptual representations of *F* and the reliability of the method:

Badness ⇒ *Representation of badness* ⇒ *Reliability of belief-forming method*

41 Mogensen, "Evolutionary Debunking Arguments and the Proximate/Ulimate Distinction," 198.

42 Street made this point in her very influential paper on evolutionary debunking ("A Darwinian Dilemma for Realist Theories of Value," sec. 8).

43 For similar (and more developed) thoughts here, see FitzPatrick, "Debunking Evolutionary Debunking of Ethical Realism."

44 For discussion, see Suikkanen, "Non-Naturalism and Reference"; Dunaway, Reality and Morality; and Werner, "Getting a Moral Thing into a Thought."

The arrows here represent explanatory connections. The idea is this: suppose that forming moral beliefs on the basis of perceptual moral experiences is reliable. Furthermore, suppose that our representations of *badness* are explained in terms of their content-fixing connection to the property of *badness*. The method's use is then explanatorily connected to its reliability in virtue of the fact that, without an explanatory connection to *badness*, the method would not be carried out in the first place.

Of course, even this response on behalf of the proponent of MP is contentious. It depends on deep and difficult questions about metasemantics that I cannot hope to answer here.⁴⁵ Such an approach may turn out to fail once an adequate metasemantics for moral content is developed. I think proponents of MP should be honest about this—this may be the best hope that nonnaturalists have of meeting the EC constraint. And it seems, at least initially, to be more amenable to a perceptual, *a posteriori* moral epistemology than an *a priori* one.

4.3. A Conceptual Competence–Based Explanatory Connection?

It is worth saying a bit about why the structure of providing an explanatory connection just given is not available to one recently influential *a priori* theory of moral epistemology—what I call the *conceptual competence* strategy.⁴⁶ This strategy manifests in different ways, but they all share a common commitment to the idea that (a) normal human individuals have a competent grasp of normative concepts, and (b) this grasp entails at least some moral knowledge. Some authors also appeal to self-evidence as having a role to play in explaining how conceptual competence guarantees moral knowledge.⁴⁷ Some proponents of this strategy assume a psychological theory of concepts, while others a Fregean

45 In the interest of intellectual honesty, I will note that I have attempted to give a metasemantic picture for nonnaturalists (Werner, “Getting a Moral Thing into a Thought”). Because that view attempts to partially reduce the metasemantic story for ethical concepts to an epistemic relation, it is unclear whether it is compatible with the solution given here. Things will get complicated here, but I hope to provide a resolution to the seeming paradox in future work. In any case, anyone who rejects the account given in that paper can accept the account given here (or vice versa).

46 Perhaps the most popular attempt to rebut epistemological objections to moral realism is to appeal to third-factor explanations (Enoch, “Taking Morality Seriously”; Wielenberg, “On the Evolutionary Debunking of Morality”; Skarsaune, “Darwin and Moral Realism”). Whatever other advantages and disadvantages such an approach may have, it does not even attempt to meet EC. Instead, proponents of third-factor explanations should argue directly against EC as a legitimate epistemic constraint. For discussion of related points, see Korman and Locke, “Against Minimalist Responses to Moral Debunking Arguments”; and Killoren, “An Occasionalist Response to Korman and Locke.”

47 See, e.g., Shafer-Landau, *Moral Realism*, 15; and Audi, “Intuition, Inference, and Rational Disagreement in Ethics.”

view.⁴⁸ For present purposes, these (important!) differences between distinct versions of the view can be set aside.

The relevant difference here between all versions of the conceptual competence view and MP is that, according to MP, but not the conceptual competence view, the representation of *badness* is (proximally) explained by *badness* itself. On the conceptual competence view, the representation of *badness* comes first, and through reflection on that (conceptual) representation, it latches onto the stance-independent property of *badness*. So it appears, at least initially, as though providing an explanatory connection between *badness* and the representation of *badness* must have a particular explanatory direction for the structure of the solution given above to work, and this direction is not available to the conceptual competence theorist. It is hard to see how an *a priori* epistemology could do this, without endorsing a Gödelian intuitionism, according to which we are directly acquainted with abstracta.⁴⁹ I fully admit that such a view could meet the EC constraint—at least insofar as my MP-based proposal does—but I worry that such views have other problems.⁵⁰

5. TAKING STOCK

I have focused on the two possible ways of understanding the epistemological condition on knowledge that nonnaturalists are thought to be unable to meet. I focused on these two because I think they are the strongest non-question-begging ways of understanding this influential objection to nonnaturalism. Of course, as always, nonnaturalists can (and have) argued directly against these epistemic constraints. On the other hand, insofar as these constraints have an intuitive pull, it would be nice to provide a nonnaturalist moral epistemology that can fulfill these conditions on knowledge as well.

MP, I have argued, is uniquely placed to do so. Perceptual experience can provide an epistemic connection if anything can; so as long as we can perceptually experience moral properties, this condition will be met. The challenge for the proponent of MP, then, is to show that perceptual experience of moral properties is possible. I have attempted to meet this challenge above. Finally, I have argued that MP is better placed to meet the fourth condition, EC, than traditional *a priori* theories. However, even though MP is better placed, it is not a trivial

48 For the former view, see Huemer, *Ethical Intuitionism*; and Schroeter and Shroeter, “The Generalized Integration Challenge in Metaethics.” For the latter, see Cuneo and Shafer-Landau, “The Moral Fixed Points.”

49 See, for example, Gödel, “What Is Cantor’s Continuum Problem?” For a recent defense of a similar sort of view, see Chudnoff, *Intuition*; and Bengson, “Grasping the Third Realm.”

50 See Luzon and Werner, “Losing Grip on the Third Realm.”

matter whether it can be met, even by a proponent of MP, because it depends on contentious issues about content fixing. Nonnaturalists, even nonnaturalist proponents of MP, are not wholly out of the woods. But important progress can be made on these entrenched epistemological objections to nonnaturalism, so long as we endorse a perceptualist model of moral knowledge.⁵¹

Hebrew University of Jerusalem
pjwerner1@gmail.com

REFERENCES

- Audi, Robert. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, NJ: Princeton University Press, 2004.
- Audi, Robert. "Intuition, Inference, and Rational Disagreement in Ethics." *Ethical Theory and Moral Practice* 11, no. 5 (November 2008): 475–92.
- . "Skepticism about A Priori Justification: Self-Evidence, Defeasibility, and Cogito Propositions." In *The Oxford Handbook of Skepticism*, edited by John Greco. Oxford: Oxford University Press, 2008.
- Baras, Dan. "Our Reliability Is in Principle Explainable." *Episteme* 14, no. 2 (June 2017): 197–211.
- Benacerraf, Paul. "Mathematical Truth." *Journal of Philosophy* 70, no. 19 (November 1973): 661–79.
- Bengson, John. "Grasping the Third Realm." In *Oxford Studies in Epistemology*, vol. 5, edited by Tamar Szabo Gendler and John Hawthorne, 1–38. Oxford: Oxford University Press, 2015.
- Casati, Roberto, and Achille C. Varzi. *Holes and Other Superficialities*. Cambridge, MA: MIT Press, 1994.
- Chemero, Anthony. "An Outline of a Theory of Affordances." *Ecological Psychology* 15, no. 2 (2003): 181–95.
- Chudnoff, Elijah. *Intuition*. Oxford: Oxford University Press, 2013.
- Clarke-Doane, Justin. "Debunking and Dispensability." In *Explanation in Ethics and Mathematics: Debunking and Dispensability*, edited by Uri D. Liebowitz and Neil Sinclair, 23–36. Oxford: Oxford University Press, 2016.
- Crow, Daniel. "The Mystery of Moral Perception." *Journal of Moral Philosophy* 13, no. 2 (March 2016): 187–210.

51 For helpful feedback on previous versions of this paper, I would like to thank Aaron Elliott, David Enoch, Nikki Fortier, Avi Kenan, Adam Patterson, Russ Shafer-Landau, and Byron Simmons. I would especially like to thank Hille Paakkunainen, David Sobel, and two anonymous referees for detailed comments which greatly improved the paper.

- Cuneo, Terence, and Russ Shafer-Landau. "The Moral Fixed Points: New Directions for Moral Nonnaturalism." *Philosophical Studies* 171, no. 3 (December 2014): 399–443.
- Dunaway, Billy. *Reality and Morality*. Oxford: Oxford University Press, 2020.
- Enoch, David. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2011.
- Farennikova, Anna. "Perception of Absence and Penetration from Expectation." *Review of Philosophy and Psychology* 6, no. 4 (December 2015): 621–40.
- FitzPatrick, William J. "Debunking Evolutionary Debunking of Ethical Realism." *Philosophical Studies* 172, no. 4 (April 2015): 883–904.
- Gibilisco, Christopher. "Theories of Properties and Ontological Theory-Choice: An Essay in Metaontology." PhD diss., University of Nebraska-Lincoln, 2016. <https://digitalcommons.unl.edu/philosophydiss/11>.
- Gibson, James. J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.
- . "The Theory of Affordances." In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, edited by Robert Shaw and John Bransford, 127–37. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.
- Gödel, Kurt. "What Is Cantor's Continuum Problem?" In *Philosophy of Mathematics: Selected Readings*, edited by Paul Benacerraf and Hilary Putnam, 116–17. Hoboken, NJ: Prentice Hall, 1964.
- Green, Mitchell. "Perceiving Emotions." *Aristotelian Society Supplementary Volume* 84, no. 1 (June 2010): 45–61.
- Huemer, Michael. *Ethical Intuitionism*. Houndmills, UK: Palgrave Macmillan, 2005.
- Jackson, Frank. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press, 1998.
- . "Locke-ing onto Content." In *Evolution, Naturalism and Mind*, edited by Denis M. Walsh, 127–44. Cambridge: Cambridge University Press, 2001.
- Jenkins, Carrie. *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*. Oxford: Oxford University Press, 2008.
- Jonas, Silvia. "Access Problems and Explanatory Overkill." *Philosophical Studies* 174, no. 11 (November 2017): 2731–42.
- Joyce, Richard. *The Evolution of Morality*. Cambridge, MA: MIT Press, 2006.
- . *The Myth of Morality*. Cambridge: Cambridge University Press, 2001.
- Killoren, David. "An Occasionalist Response to Korman and Locke." *Journal of Ethics and Social Philosophy* 19, no. 3 (March 2021): 311–23.
- Korman, Daniel Z. and Dustin Locke. "Against Minimalist Responses to Moral Debunking Arguments." In *Oxford Studies in Metaethics*, vol. 15, edited by Russ Shafer-Landau, 309–32. Oxford: Oxford University Press, 2020.

- Lutz, Matthew. "The Case for Moral Skepticism." PhD diss, University of Southern California, 2015.
- . "The Reliability Challenge in Moral Epistemology." In *Oxford Studies in Metaethics*, vol. 15, edited by Russ Shafer-Landau, 284–308. Oxford: Oxford University Press, 2020.
- Luzon, Bar, and Preston J. Werner. "Losing Grip on the Third Realm." Unpublished manuscript.
- Mayr, Ernst. "Darwin's Biological Work: Some Aspects Reconsidered by P. R. Bell." *Isis* 52 (1961): 433–35.
- McGrath, Sarah. "Causation by Omission: A Dilemma." *Philosophical Studies* 123, nos. 1–2 (March 2004): 125–48.
- . "Moral Perception and Its Rivals." In *Evaluative Perception*, edited by Robert Cowan and Anna Bergqvist. Oxford: Oxford University Press, 2018.
- McNeill, William E. S. "On Seeing That Someone Is Angry." *European Journal of Philosophy* 20, no. 4 (December 2012): 575–97.
- Mogensen, Andreas. "Evolutionary Debunking Arguments and the Proximate/Ultimate Distinction." *Analysis* 75, no. 2 (April 2015): 196–203.
- Nanay, Bence. "Action-Oriented Perception." *European Journal of Philosophy* 20, no. 3 (September 2012): 430–46.
- Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.
- Oddie, Graham. *Value, Reality, and Desire*. Oxford: Clarendon Press, 2005.
- Phillips, Ian. "Hearing and Hallucinating Silence." In *Hallucination: Philosophy and Psychology*, edited by Fiona Macpherson and Dimitris Platchias, 333–60. Cambridge, MA: MIT Press, 2013.
- Prosser, Simon. "Affordances and Phenomenal Character in Spatial Perception." *Philosophical Review* 120, no. 4 (October 2011): 475–513.
- Reed, Edward S. *Encountering the World: Toward an Ecological Psychology*. Oxford: Oxford University Press, 1996.
- Russell, Gillian. *Truth in Virtue of Meaning: A Defence of the Analytic/Synthetic Distinction*. Oxford: Oxford University Press, 2011.
- Schechter, Joshua. "Is There a Reliability Challenge for Logic?" *Philosophical Issues* 28, no. 1 (October 2018): 325–47.
- Schroeder, Mark. "How Does the Good Appear to Us?" *Social Theory and Practice* 34, no. 1 (January 2008): 119–30.
- Schroeter, Laura, and François Schroeter. "The Generalized Integration Challenge in Metaethics." *Noûs* 53, no. 1 (March 2019): 192–223.
- Schwitzgebel, Eric. "Introspection." *Stanford Encyclopedia of Philosophy* (Winter 2019). <https://plato.stanford.edu/entries/introspection/>.
- Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford: Oxford University

- Press, 1990.
- Siegel, Susanna. "Affordances and the Contents of Perception." In *Does Perception Have Content?* edited by Berit Brogaard, 51–75. Oxford: Oxford University Press, 2014.
- Simon, Paul, and Art Garfunkel, vocalists. "The Sound of Silence," by Paul Simon. *Wednesday Morning, AM*. Columbia Records. Originally released in 1964.
- Skarsaune, Knut Olav. "Darwin and Moral Realism: Survival of the Fittest." *Philosophical Studies* 152, no. 2 (January 2011): 229–43.
- Snowdon, Paul F. "Perception, Vision, and Causation." *Proceedings of the Aristotelian Society* 81 (1980): 175–92.
- Sorensen, Ray. "Hearing Silence: The Perception and Introspection of Absences." In *Sounds and Perception*, edited by Matthew Nudds and Casey O'Callaghan, 126–45. Oxford: Oxford University Press, 2009.
- . *Seeing Dark Things: The Philosophy of Shadows*. Oxford: Oxford University Press, 2008.
- Soteriou, Matthew. "The Perception of Absence, Space, and Time." In *Perception, Causation, and Objectivity*, edited by Johannes Roessler, Hemdat Lerman, and Naomi Eilan, 181–206. Oxford: Oxford University Press, 2011.
- Stout, Rowland. "Seeing the Anger in Someone's Face." *Aristotelian Society Supplementary Volume* 84, no. 1 (June 2010): 29–43.
- Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127, no. 1 (January 2006): 109–66.
- Suikkanen, Jussi. "Non-Naturalism and Reference." *Journal of Ethics and Social Philosophy* 11, no. 2 (February 2017): 1–24.
- Timmons, Mark. "On the Epistemic Status of Considered Moral Judgments." *Southern Journal of Philosophy* 29, no. S1 (Spring 1991): 97–129.
- Vavova, Katia. "Debunking Evolutionary Debunking." In *Oxford Studies in Metaethics*, vol. 9, edited by Russ Shafer-Laundau, 76–101. Oxford: Oxford University Press, 2014.
- . "Evolutionary Debunking of Moral Realism." *Philosophy Compass* 10, no. 2 (February 2015): 104–16.
- Werner, Preston J. "Getting a Moral Thing into a Thought: Metasemantics for Non-Naturalists." In *Oxford Studies in Metaethics*, vol. 15, edited by Russ Shafer-Landau, 140–69. Oxford: Oxford University Press, 2020.
- Wielenberg, Erik J. "On the Evolutionary Debunking of Morality." *Ethics* 120, no. 3 (April 2010): 441–464.
- Woods, Jack. "Mathematics, Morality, and Self-Effacement." *Noûs* 52, no. 1 (March 2018): 47–68.