

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XXV · NUMBER 1

July 2023

ARTICLES

- 1 Giving Up on Someone
Kiran Bhardwaj
- 22 Kant and the Balance of Moral Forces
Santiago Sanchez Borboa
- 59 Promises, Commitments, and the Nature of
Obligation
Crescente Molina
- 80 What Is the Point of Nondomination?
Devon Cass
- 105 On an Analytic Definition of Love
Tyler J. VanderWeele
- 136 Nonhuman Animals and Epistemic Injustice
Andrew Lopez
- 164 Crying Havoc and (Re)claiming Rights: How
the Liabilities of Revisionism and the Just War
Tradition Alter the Moral Equality of Combatants
Marcus Hedahl

DISCUSSION

- 212 It's a Fine Line between Sadism and Horror
Scott Woodcock

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	Anca Gheaus
Stephanie Collins	Hallie Liberto
Dale Dorsey	Errol Lord
James Dreier	Colleen Murphy
Julia Driver	Hille Paakkunainen
David Plunkett	

Discussion Notes Editor

Tristram McPherson

Editorial Board

Elizabeth Anderson	Stephen Perry
David Brink	Philip Pettit
John Broome	Gerald Postema
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
Susan Wolf	

Managing Editor

Rachel Keith

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

GIVING UP ON SOMEONE

Kiran Bhardwaj

IT IS A STRIKING PRACTICE how we often give up on people we find morally (or otherwise) criticizable. We judge the person as having some characteristic we are averse to and, as a result, no longer consider that person to be worth our time and concern. This behavior is not only common, it is also seen as morally admirable. There is something virtuous, it might be argued, in not being willing to engage with people one finds morally noxious. And even were the person to be only *somewhat* objectionable, we still tend to think that giving up is nonetheless acceptable. Not only can we choose our friends and associates, the argument might continue, we are entitled to decide who we want to avoid entirely or lessen our exposure to.

Yet my deep concern is that these quick justifications of our practices of giving up on morally criticizable people are flawed. In this paper, I ask: When, if ever, is giving up on someone as a person morally justified? My chief aim will be to defend the interest and importance of this question, but I also will tentatively identify several considerations that count against giving up as a moral practice. To focus our attention further, I will be discussing only cases of giving up on a person we have a direct relationship with, rather than giving up on someone (such as a public figure) we do not or cannot interact with in any meaningful way.

Section 1 of this paper will offer a conceptual analysis of the phenomenon of giving up and set out some distinct varieties of how we do so. In section 2 and section 3, I will then turn to the normative considerations involved in giving up on a person. Section 2 details the most radical manner in which one can give up on a person, and argues that doing so is morally impermissible (or, at the very least, morally inappropriate). Section 3 details the normative considerations for a less radical version of giving up on the person: I present an argument for three considerations that must be met in order to meet an obligation of due care before giving up. The last of these considerations is that we must consider the broader implications of our choice on the remainder of the community. Section 4 reviews and responds to some possible objections to the arguments of section 2 and 3.

1. WHAT IS “GIVING UP”?

We use the language of “giving up on someone” in a multivocal way. Sometimes we speak of giving up on someone for reasons that are *benign*—perhaps I simply may not connect socially or emotionally with the person, our interests or activities do not align, or geographical distance from each other has caused us to drift apart. This paper is focused on a different kind of giving up: when one gives up on someone due to an *aversion* to some aspect of a person or their behavior.¹ In this section, I will detail two features of aversive cases: the force of the reasons one has for giving up and the scope of one’s giving up.

In aversive cases of giving up, we generally respond to something we think of as a fault: a pattern of misbehavior or disliked behavior, or a single egregious instance of misbehavior or disliked behavior, which leads us to think they are worth giving up on.² For example, I might give up on a friend who is flaky, or a sibling who broke their promise to repay a loan. We also see cases in which a person gives up on someone due to their aversion to the other’s membership (or perceived membership) in a social group or some other feature of their identity—for example, giving up on someone who has political or religious views one disapproves of, or disowning a loved one because they are queer. We can further distinguish between aversive cases involving moral aversion and those involving nonmoral aversion. We might, for example, give up on someone because we find them to be socially annoying or irksome, but not because of moral wrongdoing.

The reasons one has for giving up have some degree of force. That force is measured by the severity or regularity of the behavior, or the “badness” of the characteristic in question. Cases of moral aversion often, though not always, carry more force than nonmoral aversion.³ A profound wrong is ascribed much

- 1 While I will not be focused on benign cases of giving up in this paper, there is surely more to say about ways in which we can mistakenly confuse benign and aversive cases, and whether or not there are normative constraints on giving up for benign reasons. Thank you to an anonymous reviewer for their helpful classification of benign reasons for giving up.
- 2 There is a secondary form of aversive giving up in which I judge *myself* to be the person who has the fault, and I give up on the other person because of how they cast my fault into relief. For example, consider a person who gives up on a successful friend they feel has a perfect life because they feel bad by comparison when around them (even though they think the friend has done nothing morally or otherwise wrong).
- 3 I suspect that while generally we do—and should—consider moral reasons to be qualitatively stronger than nonmoral reasons, this may not always be the case. We might be somewhat averse to finding out a person has told a one-off lie that has done some minor harm, and be even more averse to a person who is constantly and irritatingly late. I also note that we may also have difficulty in disambiguating moral from nonmoral cases, even

more weight as a reason for giving up than a behavior we would describe as profoundly unlikeable but not wrongful.

As a result, giving up on someone can be more or less radical in scope—generally, in proportion to the force of one’s reasons. I suggest that we first draw a distinction between giving up on the *relationship* versus giving up on the *person*. If I give up on the relationship, I withdraw from the person while remaining (in some sense) open to them.⁴ Giving up on the relationship does not require a certain outcome: I could end the relationship (or let it end) *or* change to a less-engaged or less-intimate relationship, or a less-involved *kind* of relationship—e.g., a former romantic partner becomes a friend, or a friend becomes an acquaintance. Giving up on a relationship involves no longer engaging with the person as much as previously, or perhaps entrusting a different or restricted set of things I care about to them.

In contrast, if I give up on the person *as a person*, I close all future possibilities of interacting with the person that would leave myself or others vulnerable to them, so far as is under my control.⁵ When I give up on a person in this sense, I think of them as hopeless or have a standing disposition of suspicion toward

though the distinction can be a useful heuristic. The latter, such as the constantly late person, are sometimes read as giving us information about one’s considerateness of other people (or other moral attitudes). Note, of course, that disambiguation is even more complicated insofar as phenomena such as punctuality are culturally relative social practices.

- 4 The kind of openness might vary, but could include such features as being open to hearing about their lives from those who we socially have in common, being willing to act on new information that might change my attitudes or behaviors toward them, or being willing to re-engage in some sort of relationship or a more engaged form of relationship with them in the future.
- 5 I use the term “vulnerability” in a broad sense. The term most obviously tracks the ways in which we can be susceptible to those who have, institutionally or relationally, some power over us. Yet it seems like I could give up on a person who does not seem to hold such power over me—perhaps (1) because their actions are constrained so they cannot directly harm me (perhaps because of imprisonment or death), (2) because they meant so little to me that nothing they say or do could bother me, or (3) because I am the person who holds the institutional or relational power. In each case, however, I suspect there is a sense in which we are vulnerable. For 1: even if I am insulated from their actions (which may not genuinely be the case), I still might be *indirectly* vulnerable to them because what they do impacts others who remain in my life, or even because they still imaginatively or emotionally affect me (I suspect this is what we mean by the idiom that a person is living rent-free in our heads). For 2: the fact that they mean so little may be an indication that I have already given up on the person, but it could also mean that I am choosing not to care about the ways in which they can harm me, or people or projects I care about—even though I am genuinely vulnerable. For 3: the person in power might describe their time, effort, or consideration as things that are vulnerable to the person they have power over. Thank you to an anonymous reviewer and Josh Kissel for pressing this point.

them. I may not think of myself as being “in community” with them and avoid engaging in joint activities with them.⁶ I might also recommend to or exhort others to do the same. However it is accomplished, giving up on a person is a radical form of shutting down engagement with the person in question.⁷

Even after having given up on the person as a person—and remaining closed to them—I may nonetheless still have to concern myself with what the person has done or will do or spend time around them. For example, I could give up on a badly behaved colleague as a person, yet know I cannot distance myself from interacting with them unless one of us leaves our job. However, I will, by one measure or another, place myself in the best position to no longer be vulnerable to that person from here on out.

There are at least two different ways in which we give up on someone as a person. One of the most philosophically interesting ways to do so—and which will be the focus of section 2—is when we no longer give a damn about them: we hold an ongoing objective attitude toward them (in Strawson’s sense).

Ordinarily, we hold reactive attitudes and intentions toward others—attitudes such as gratitude, resentment, forgiveness, love, and hurt feelings.⁸ These reactive attitudes are those that demonstrate involvement or participation in a relationship.⁹ However, I might alternatively hold an objective attitude toward the person, which is to see that person “as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided.”¹⁰

As Strawson pointed out, we characteristically suspend or modify our reactive attitudes toward those agents who are set apart in some way, as abnormal or immature. So I might suspend my annoyance toward a child because I reflect on their cognitive development and think they’re doing *exactly* what an eight-year-old characteristically does at this age. We also rely on objective attitudes as a resource for dealing with otherwise ordinary agents. I could think about

6 What we seem to mean by the phrase “not ‘in community’ with them” is that we deny affiliation with them—perhaps that we do not belong to each other in some shared sense, lack shared objectives, or do not hold special obligations toward each other.

7 Note that sometimes we say, “I give up on that person,” and mean something like, “I no longer expect this person to be capable of changing themselves or their behaviors,” or, “I will no longer try whatever I was previously trying to do when I interacted with that person.” As a result, those turns of phrase can sometimes indicate only that we are giving up on a relationship—not that we are necessarily giving up on them *as a person*.

8 Strawson, “Freedom and Resentment,” 75.

9 Strawson, “Freedom and Resentment,” 79.

10 Strawson, “Freedom and Resentment,” 79.

a person's ordinarily enraging behavior in a different light if I temporarily suspend my reactive attitudes toward them. I then might be able to gain insight into why they are behaving in such a way, or avoid the effort of being involved with the person.

I give up on a person in the first sense when I hold a *permanent* objective attitude toward them. This means that rather than engaging with them, I at most aim to manage their behavior (insofar as I interact with them at all).¹¹ I would consider them no longer worth my time and concern *as an agent*, and would treat them as such. Consider this case study:

Person *A* has behaved egregiously, but due to circumstances outside my control I will still have to interact with them regularly. I was originally hurt or angered by their misbehavior, but now I have a cool and objective attitude toward them: nothing they do really bothers me anymore, because I expect continuing bad behavior from them (and perhaps have taken other preventative measures to mitigate harm). When I do have to interact with them, I treat the experience like an anthropologist or scientist: I am observing what this creature does in their native habitat. Those who know me well would say that I no longer get angry at Person *A*, and are a bit bemused by my objectivity toward them.

The protective mechanism, in this case, is to no longer regard them *as a person*—instead, I regard them as something to be managed or controlled. My guard is up, and I am not susceptible to anger or resentment or similar reactive attitudes from their future misbehavior as a result.

The second way of giving up on someone as a person is to continue to hold the reactive attitudes toward the person but to act to protect ourselves from the person to the greatest degree that we can. Characteristically, my behavior would incorporate the defensive reactive attitudes (those “spikier” reactive attitudes that serve to keep the wrongdoer away or on their guard, like anger, disgust, or contempt). We may also guard ourselves against them by cutting off contact, refusing to engage in joint activities, and closing off the possibility of reconciliation. This kind of giving up on a person will be the focus of section 3.

11 On Strawson's analysis of the objective attitude, he says that we cannot hold an objective attitude toward an ordinary person for very long; he asserts that, at some point, we must “sever” the relationship. See Strawson, “Freedom and Resentment,” 79–80. I disagree with Strawson's point: it is possible to hold an objective attitude for an extended period of time, though, of course, doing so no longer could be described as a relationship in the ordinary social sense. It is just not possible to do so while retaining one's recognition of the person *as a person*.

Giving up on a *person* is a much more serious action than giving up on a relationship—and as I will argue in sections 2 and 3, there are moral considerations that weigh against giving up on persons as persons.

2. GIVING UP ON A PERSON, TYPE 1

The normative claims in this paper are motivated by a broadly Kantian story about what we owe to ourselves and other people. While I will offer an account of when we may (and when we may not) give up on others for aversive reasons on those grounds, I also suggest that those who are not Kantians could agree with similar (even if not as stringent) conclusions. In this section, I will lay out our obligation to regard moral agents *as* agents, and then suggest that, on those grounds, the first type of giving up on a person is impermissible (or at the very least, morally inappropriate).

The familiar Kantian claim is that we must treat every person as someone with the capacity for choice, and we should never treat them as mere “things.”¹² Their personhood gives us a duty to recognize them as an agent—that is, someone who is morally responsible for what they do. As Stephen Darwall puts it when characterizing this respect for someone as a person, one is owed such respect by virtue of *being a person*—not whether or not they are a *good* person.¹³ There are no exceptions to this obligation—even those who are blameworthy or otherwise objectionable are persons, and ought to be regarded as such. In fact, regarding someone as accountable for their wrongdoing is part of what it is to recognize them as a person—and engaging in a process of accountability demonstrates this same kind of recognition.¹⁴

12 Kant, *Groundwork to the Metaphysics of Morals*, 4:428; *Metaphysics of Morals*, 6:380.

13 The “recognition respect” Darwall discusses in his paper—the respect we owe to persons *because* they are persons—is the form of respect Kantians argue we have a duty to have for all rational agents. We must recognize that all persons should be treated as having equal moral standing. This is in contrast to what Darwall calls “appraisal respect” for a person, in which we think the person has a good or worthy character (which can be a matter of degree). Recognition respect is not a matter of degree—I either recognize a person as a person, a fact that places moral constraints on our behavior, or I do not. I can think a person is a bad person (that is, have a low degree of appraisal respect for them), but recognize that they are a person and must be treated as such. See Darwall, “Two Kinds of Respect,” 46–47.

14 Importantly, an obligation to regard a person as a person may not require that we engage in a *process of accountability* with those who are blameworthy or are otherwise objectionable. While such a process is a way of showing wrongdoers that we are engaging with them as responsible agents, engaging in such a process to hold others responsible for what they do is a defeasible obligation. Sometimes it is overridden when such a process would put us into harm’s way to a degree that it breaches an obligation that we have to ourselves. Sometimes prudential considerations may matter (e.g., if what I alone could do to hold the

Not everyone, of course, is motivated by the Kantian view that we have a duty to respect persons because they are persons. However, I would suggest those with other views should be convinced by a similar but weaker version of the claim: persons are the kinds of beings where it is *appropriate* to regard them as having agency. We owe it to people to regard them as having the capacity to make choices and be responsible for what they do. This is part of what is required to treat people with respect (or some equivalent: perhaps the emphasis is instead on how to consider their interests in a way that is responsive to their moral standing *as* a responsible agent).

With those general points in mind, let us begin by discussing the case of giving up on someone as a person in which we permanently suspend the reactive attitudes toward the person. I suspect this method is extremely tempting. We may or may not be able to avoid the person in question, but we can still protect ourselves by studying their bad behaviors, constraining the person when and how we can, and pushing whatever causal levers might cause their behavior to be less egregious.

Holding such a permanent objective attitude is a much more common way of responding to others than I suspect we like to think.¹⁵ We see this when we subject elders who harm others to this kind of treatment (due to, often, implicit or explicit ageism) where we say, “Oh, they’re stuck in their ways and are never going to change,” and then try to manage them into doing less harm. We do this as well with those who have (or we believe have) a mental illness. We may look like we are engaging the person, but we are not really doing so—it is a minimalistic pretense in which we have deliberately muted the person *as* a person.¹⁶

Yet despite this temptation, a permanent or even semi-permanent objective attitude toward someone—the first kind of giving up on a person—is morally impermissible on Kantian grounds. Even for those not motivated by Kantian claims, I argue that giving up on a person in this way is morally inappropriate.

person responsible or to change their behavior would backfire). When prudential reasons have sufficient weight, our obligation to hold others responsible can be fairly minimal—sometimes this involves simply recognizing the person as a person who is responsible for what they have done.

15 Elinor Mason writes that it “takes a lot for us to permanently give up on [other agents] as agents”—and we might think that such cases are rare and exceptional as a result (*Ways to Be Blameworthy*, 147). But I do not think that threshold is as high in practice as we like to think—it is quite tempting to dehumanize those we do not consider to be paradigmatic agents, and we similarly do so to those we disagree with or dislike. I explore Mason’s account further in section 4.

16 We also may recognize this same or a similar phenomenon in unjust social behaviors that manage members of oppressed groups rather than granting them autonomy or full engagement.

I will first lay out the Kantian argument, and then a broader argument, for these claims.

Though it may be tempting to settle into a permanent or semi-permanent objective attitude toward any person for self-protection, this kind of giving up is impermissible. The wide variety of reactive attitudes we have toward other people demonstrates that we are still involved with them as agents in full. In contrast, it would fail to treat others as autonomous moral agents if we try to merely manipulate or manage the people in question, as is characteristic of the objective attitude.¹⁷ We may still continue to criticize a wrongdoer's behavior as wrongful, and to treat them as responsible for what they have done in a variety of ways (which can include constraining their behavior). I ought to take those alternatives instead.¹⁸

I suggest that even non-Kantians should consider it to be morally inappropriate to hold a permanent objective attitude toward wrongdoers. The objective attitude is incompatible with regard for persons, no matter the normative ground for that regard. For example, we might take ourselves to have an interest in being treated as a full moral agent or treated as responsible for what we do, or that a caring regard for a person would see them in such a light. It would then be morally inappropriate to hold this permanent objective attitude toward others, on such a view, even if it is ultimately permissible to do so because of the weight of other interests, or due to normative claims at play in yet other theories. The objective attitude also dulls the ways in which we think about and respond to the badness of their wrongdoings. We might not, due to that objective stance, respond to the person in a way that communicates their fault (even when doing so is warranted and would help to change their behaviors). Holding a permanently objective attitude toward wrongdoers might sometimes have prudential or moral benefits. However, the *prima facie* moral inappropriateness of a permanently objective stance toward a person may outweigh these benefits, especially the prudential ones.

There are some cases of wrongdoing that nonetheless pose a challenge to what I argue is a duty to avoid treating others as nonagents. Some kinds of

17 Similar points are made also by Holmgren, *Forgiveness and Retribution*, 7–8; and Langton, “Sexual Solipsism,” 162–63.

18 An objector might ask if it is fair to characterize a permanent objective attitude as a “failure to treat a person as a person.” After all, what if I am moving someone like Person A into the ranks of people who I recognize exist, but who I no longer engage with? There are billions of people I consider persons but who I neither engage with nor emotionally respond to. In response, it is indeed possible to think of a person as a person, but not have a relationship with them. To be at a remove is possible and permissible (the conditions for which will be discussed in section 3), but in that case I still would not be in a state of permanent suspension of all reactive attitudes toward them altogether.

wrongdoing are damaging to *one's own sense of agency*, such as trauma or abuse.¹⁹ In such a case, the need to be able to maintain self-respect could be argued to take precedence over our obligation to refrain from giving up on others in this first sense. Suspending the reactive attitudes may be one's only or best method of ensuring that self-respect—we would want, in such cases, to defer to the recommendations of those with expertise in healthy coping mechanisms for trauma and abuse.

For those non-Kantians who would claim that giving up in this first sense is morally inappropriate (rather than impermissible), the way to assess the normative considerations in such cases is comparatively straightforward—there would be some threshold past which any obligations one has to the wrongdoer can be overridden by obligations one has to oneself. For those who do not think there are self-regarding duties, merely moral prerogatives, then this same normative claim holds insofar as we are entitled, morally speaking, to regularly prioritize our interests over the demands of others.

The Kantian account can make sense of how self-care is important for those who, given societal expectations and pressures, are asked to be self-effacing or to give too much of themselves in favor of others.²⁰ Yet these cases—in which a duty of self-care feels at odds with the duty to refrain from treating others as nonagents—pose a challenge for a Kantian account, as both are important kinds of obligations.

While giving up on the person under such conditions is certainly understandable because of the ways in which it allows one to maintain one's self-respect, I argue that on a Kantian account, this choice would be at most an *excusing condition* for those who give up on their wrongdoers in this way. It would be far better if, instead, the person could work toward eventually being able to see their wrongdoer again as a responsible agent. That is, perhaps one cannot right now engage with the wrongdoer as a person and must hold an objective stance toward them (and, if the situation requires, manage them). However, if granted time and removal from the dangers that the wrongdoer poses—and, even better, accountability and change on the part of the wrongdoer—one

19 This same point, I think, also applies for other kinds of trauma, including those from various kinds of oppression that are not centered on a particular actor's doings—it is instead sourced from a whole structure of oppressive institutions and norms, and also expressed in the actions of particular people who do harm. There may be a similar desire to cope with that trauma by permanently suspending the reactive attitudes toward those who behave in ways that express their approval of or commitment to that structure.

20 Carol Hay's discussion in *Kantianism, Liberalism, and Oppression* of the Kantian conception of self-respect and how it can serve as a tool to resist gendered and other kinds of oppression is particularly helpful here. See especially *Kantianism, Liberalism, and Oppression*, 68–78.

might be able to again recognize the person as a person. Nonetheless, in such cases, I suggest that the second kind of giving up would better allow the person to *both* recognize the person's agency and care for themselves.²¹ Let us turn to that alternative now.

3. GIVING UP ON A PERSON, TYPE 2

I have argued that giving up on a person via a permanent objective stance is morally impermissible (or, at the very least, morally inappropriate), because it is not consistent with respecting a person as a person. We should also ask the same question about the second, less radical kind of giving up on a person—in which I have closed all future possibilities of interacting with the person that would leave myself or others vulnerable to them, yet still maintain the reactive attitudes toward the person.

Unlike the first type of giving up on someone, this second type of giving up on someone does not outright deny their personhood. However, I will argue that it remains a morally risky practice given the gravity of such closure. I contend that for any case of giving up on a person, we have an obligation of due care before we give up on them—that is, an obligation to “check our work” and ensure we have evaluated all considerations that matter before acting.²² For example, we ought to assess whether we are making decisions based upon good information.²³ I take this epistemic component of due care for granted—and will focus this section on further features of what we must do in order to meet our obligation of due care.

- 21 Note that not all cases of taking the objective point of view constitute giving up. We might have reasons to *temporarily* take up the objective point of view. For example, an aging parent may require daily care despite holding views that are deeply objectionable (such as being racist or sexist)—and taking an objective point of view is what allows me to care for that parent in the right way. However, in such a case, if we do take up the objective point of view, we should do so in a circumscribed manner and with a specific goal in mind (a goal that the person can, at least hypothetically, share with us). Thank you to an anonymous reviewer for this example and pressing this point.
- 22 I suspect that a parallel set of due care considerations holds for how we ought to deliberate when giving up on a *relationship* as well as giving up on a *person*—however, further argument about the normative considerations involved in giving up on a relationship are outside the scope of this paper.
- 23 We are, of course, prone to error, deception, and self-deception. For example, a person could give up on someone after (without confirmation) taking hearsay about their behavior as fact. It is clear that we have an obligation of due care to confirm our facts before proceeding. In what follows, I will assume that we're proceeding as well as we can epistemically—that as far as is possible for us, we are accurately tracking the nature and severity of the disliked or wrongful behavior.

In particular, there are three further obligations of due care before giving up: (a) we must confirm that the choice to give up on someone is fitting for the situation, (b) we must confirm that we have given special obligations due to the nature of the relationship with the person their full weight, and (c) we have assessed whether we can continue to meet our obligations to the overall community if we are to give up on the person. I believe that the first two obligations are straightforward and can be quickly gestured at, and I provide an argument for the third. Again, while I would argue that it is a *duty* to engage with the due care considerations offered in sections 3.1–3.3 below, I think that those who do not share similar moral commitments should at least think that it is *appropriate* to do so.

3.1. Assuring “Fit”

The first obligation of due care for giving up on a person (in the second sense) is that we must confirm that our response is fitting for the severity of wrongdoing in question. After all, we surely can mistake our dislike or discomfort with a kind of behavior as something that warrants giving up on a person, even when doing so is wildly disproportionate. Consider a person who takes a coworker’s mildly annoying behaviors as a reason to give up on them as a person, when, in fact, they are simply being uncompromisingly impatient with the coworker. Or consider a person who goes no contact with a sibling who said something unkind or inappropriate in the context of a generally good or workable sibling relationship.²⁴ Choosing to close off the relationship permanently, for aversive reasons that have little force, would be ill-fitting.²⁵

What makes this assessment of fittingness before giving up on a person (in the second sense) so important is that giving up in this sense closes all future possibilities of interacting with the person, so far as is under my control.²⁶ If

24 I assume for this case that what is said was objectionable but would not be the kind of statement such that it—by and of itself—would warrant a correspondingly severe degree of response.

25 Some philosophers might even argue that we would be making an error of fit if I fail to accept if the person has done what they ought to deserve forgiveness or acceptance after having done something wrong. Margaret Holmgren argues, for example, that forgiveness is warranted when the wrongdoer has taken the appropriate response after their wrongdoing (*Forgiveness and Retribution*, 10). Holmgren’s is obviously a contentious claim, which I do not defend here—merely offer as a possible position.

26 Here I follow Trudy Govier, who argues that there are only select cases of the unforgivable: when the wrongdoings in question are “enormities, appallingly wrong acts that violate profoundly important moral principles,” and where they have neither acknowledged nor made restitution for what they have done (“Forgiveness and the Unforgivable,” 68–72). In many ordinary cases of giving up, the wrongdoing in question does not meet the threshold

one gives up on the *relationship*, one is open to new and relevant facts about the person—for example, that they have changed or genuinely regret what they have done. The permanent closure of a relationship demands a certain kind of fit for the wrongdoing in question in order to be proportionate to the wrongdoing.²⁷ Note, of course, that we can also make the reverse mistake: we do *not* give up on a person, even when we have aversive responses to them (or their behaviors) that have a great deal of force—enough that one *ought* to close oneself off to the person. For example, a person might choose to stay with an abusive partner because the partner asked for yet another chance, even when doing so would fail to be responsive to a duty they have to themselves (or dependents they are responsible for). That would be another failure of due care with regard to fittingness.

3.2. *Special Obligations to the Person*

There is also a second, particular obligation of due care that, crucially, is to confirm the nature and weight of the obligations we have toward particular people and that are at least partially determined by the kind of relationship we stand in with them.

Special relationships of various kinds come with a requirement to assist the other person in their choice-making or self-improvement, especially if the person requests my assistance *or* if I have a special responsibility for them that emphasizes such assistance (as with one's children).²⁸ We are all flawed people. We often have off days and wobbles, and every one of us exhibits a character with features that we and others may not think are worthwhile or good for us to possess. We make mistakes (both moral and nonmoral), many times repeatedly, and we can be exceptionally bad at picking up on our mistakes. Luckily, many of us also are surrounded by people who help us identify what we are doing poorly,

suggested by Govier: we give up on people when they have acted wrongly, yes, but are comparatively minor kinds of wrongdoing.

- 27 It may be possible, then, that a person should only be given up on *as a person* for moral reasons. Nonmoral reasons, such as finding the person's personality grating or annoying, would not qualify as a sufficient justification for giving up on the person. (They may, however, justify giving up *on the relationship*, up to and including ending the relationship.)
- 28 This is characteristically true of one's obligations to one's children and other educative relationships (students, mentees, etc.). It may also show up in other relationships—for example, many of us rely on those we are close to (friends, parents, older and adult children) or those who have special kinds of expertise (counselors, doctors) to assist us in making judgments about what we should do. We also discretionarily extend this kind of assistance to those we do not have a special relationship with, or a relationship where such assistance is not characteristic—as we do when we advise those we do not know personally or specially.

and who can help us put plans of action into place to make a better choice or to improve ourselves. Our relationships allow for assistance to each other in order to become morally better agents.

So, in such relationships, I may need to be willing to listen to the person in question, and take time to understand why they made the choices they did.²⁹ I might offer observations or suggestions (especially—but not only—when solicited) when those comments are true, necessary, and kind. It might also involve recognizing my own limitations in helping the person, and to instead help them work with a different person who might be better suited to help them improve themselves.³⁰ This is implicit in the way we raise our children, but the obligation also applies to others as well.³¹ For those I have a special relationship with—and especially if I wield some kind of power in the relationship—it seems likely that I should give them an extra benefit of the doubt, and withdraw only in cases where I need to for my own self-protection (or the protection of someone else toward whom I have more pressing special responsibilities).³² In many cases, this might include being open to reestablishing the relationship. As a result, *giving up on this specific kind of relationship* (as described in section 1) may be more appropriate than the closed nature of *giving up on the person as a person*.

There are certainly occasions in which someone with whom one has a special relationship has done something wrong with sufficient force that would justify giving up on them as a person. However, the general responsibility to assist those we have a special relationship with means we must give the decision a special kind of confirmation before doing so—we may wish to not move too quickly or conclusively in our decision to give up on them.

We also might ask if there are special relationships in which it is always morally impermissible to give up on the other person in any way—perhaps parental (or filial) relationships have this character of unconditional obligation.³³ Older and adult children, and certainly our parents, can be responsible for actions that

29 See Calhoun, “What Good is Commitment?” 619–20, and “Changing One’s Heart,” 95–96, for similar points.

30 For example, I might assist a loved one to get access to a therapist, when I realize that they need the objective positioning or extra skill set offered by such therapeutic relationships—and that our loving relationship allows only for other kinds of support or assistance.

31 These points are closely related to—and are intended to mirror—Ryan Preston-Roedder’s view of the reason we ought to have faith in humanity: having faith in people’s decency helps them to act rightly, treat them justly, and provide support for them (“Faith in Humanity,” 666).

32 This point is similar to Barrett Emerick’s, particularly insofar as he explores why disengagement is an inadequate response to those who hold objectionable positions in a loving relationship (“Love and Resistance,” 8–9).

33 Thank you to an anonymous reviewer for asking for further consideration of this point.

cause enough harm that there is sufficient reason to justify giving up on them as a person. For example, it would be deeply harmful to insist that a child who was abused by their parent has an unconditional obligation to reciprocally care for them in their old age. In such cases, I think, the unconditional-seeming filial obligation would be to give that decision to give up on the parent as a person a special weight of consideration. Due to the unique painfulness involved in giving up on a parent (or child), however, I suspect that such weight of consideration is almost always already part of the choice to give up.

3.3. *Obligations to the Community as a Whole*

Finally, someone may be right that they genuinely stand a risk of harm if they are to continue to engage with an unpleasant person, and want to close themselves off to avoid being vulnerable to that person. Yet I argue that we sometimes have a responsibility to not give up on others in this way, due to what we owe the remainder of our community. One of the faults in our conventional practices of giving up on people is that we often fail to consider the third parties who will continue to be exposed to the harms that person's unlikeable or morally wrongful behaviors might pose.³⁴ Instead, I argue we are morally required to ensure that the harm we would face is not less important than other possible harms to others in our community.³⁵

Many cases of giving up on someone fail to meet this criterion of due care—a white person who wants to be an ally might withdraw from someone they know who says something racist; people who think of themselves as upstanding members of society may shun the addicted and those who have come out of prison.³⁶ To see why this is a problem, consider first that the most pressing concern is the harm many marginalized racial groups (particularly BIPOC) face from racism. White allies are right to be indignant about the racist beliefs or actions of their community members or willed blindness to institutional racism. Yet they may not be doing the targeted communities any favors by giving up on these community members: they are well-positioned

34 Aristotle, in the *Nicomachean Ethics*, seems to exhibit an example of this insensitivity to this concern. He says that if a friend who was once virtuous (but who has fallen into vice) can be morally improved, we ought to work to rescue his character. However, “if one friend stayed the same and the other became more decent and far excelled his friend in virtue, should the better person still treat the other as a friend? Surely he cannot” (1165b:21–25).

35 I use the term “community” as having a wide scope—not merely to indicate some kind of geographical proximity, but those persons who can impact and be mutually impacted by the choices and actions of others.

36 The assumption of “having been in prison” being a proxy for being morally problematic is itself problematic, as extensively discussed by figures such as Michelle Alexander in *The New Jim Crow*.

to challenge racist narratives and less harmed by the expressed views. If they withdraw from the person, it places the burden on others—and, most often, the members of the marginalized groups themselves—to contend with that person instead.³⁷ As such, it seems incumbent on white allies, even if they are disgusted by the expressed views of racists and those steeped in racist acculturation, to engage in the burden of trying to convince their community members of the wrongness of racism (with deference, of course, to the considered preferences of the harmed groups for how allies should proceed).

Similarly, with regard to offenders, the primary concern is to offer safety and support for victims (both primary and those who face secondary or tertiary harms, like their families or loved ones, or communities), and avoid future victimization through re-offense. We also know that the use of community support networks can help decrease recidivism amongst high-risk sexual offenders, and it seems plausible that the same is true for other kinds of criminal offense.³⁸ I may then have a responsibility to continue to engage with an offender, other things being equal. This might require me to refrain from giving up on the offender, and perhaps to serve as part of a community support network to reintegrate them with the community and support their rehabilitation or supervision.

We have a strong “not in my backyard” (NIMBY) tendency not only for prudential cases, but also moral cases. And like those ordinary NIMBY cases, these tendencies can exacerbate existing patterns of injustice.³⁹ So an important

37 I am characteristically thinking of white people on Facebook who, in the wake of Ferguson (and many other cases of anti-BIPOC racism), proudly announced that they have unfriended other white people who have posted racist things, or gone no contact with such family members. They take themselves to be good allies by not listening to the racist things their peers are saying. I take this, however, to leave a heavier burden on those non-white people the racist friends or family members will continue to interact with.

38 Lobanov-Rostovsky, “Sex Offender Management Strategies.”

39 Of course, this argument by analogy to more typical NIMBY cases only goes so far. In the ordinary cases, NIMBY attitudes involve a public good that we all want, but comes with undesirable features—for example, a landfill (which is accompanied by undesirable smells) or a highway (which is accompanied by undesirable sounds). We do not want those undesirable features in *our* immediate vicinity, even if we want the public goods of a highway or landfill. What public good comes from having unpleasant people around? If the cases are not similar in the relevant ways, the attempt to pull an analogy will fail—and perhaps, so will my argument.

I agree that it is not easily possible to identify a public good that corresponds to the highway or landfill in the conventional case. Yet I resist the move to think about people in precisely the same way we should public goods and their accompanying local harms—after all, these are *people* we are thinking of. We can build a road or utility but cannot choose to bring other rational agents into or out of the world, or move them around, with the same kind of purely prudential reasoning. Nonetheless, I believe the analogy helps up to a certain point. The attitude remains deeply similar to the conventional NIMBY attitudes

component of the obligation of due care is that a person must ask if they *genuinely* face a special risk of harm (or other excusing conditions) that would allow them to appropriately close themselves off from the target person—and, correspondingly, whether giving up on that person would shift the burden of engagement onto others in ways that replicate overall patterns of injustice.⁴⁰ To fail to give sufficient care to this concern would be, perhaps, to promote injustice: after all, those who do not have the ability to or cannot afford to insulate themselves from morally (or otherwise) criticizable persons are those who are more vulnerable and with fewer resources.

One's own aversion to the person or their behaviors may not be more important than assessing what one can do while maintaining engagement with the offending person. If we are not a special target for that person's behaviors, and have the ability to intervene and address their behaviors while only taking on small harms, or the risk of small harms, then we ought to do so for the sake of the full community. However, we do not have to face the potentially threatening person alone: we may also choose to work together with other members in the community to limit the amount of harm that each still-engaged person faces.⁴¹

4. FURTHER OBJECTIONS AND RESPONSES

There are surely objections to the arguments made in the previous sections. In what follows, I set out four objections and their replies. The first and second apply generally to all forms of giving up on the person, and the third and fourth objections pertain to the second type of giving up.

(insofar as it is directed to unpleasantness in our locality), and the procedures for solving the problem seem to be the same (to try to move that unpleasantness as far away from *me* as possible). Similarly, it matches in how those who have undue power or privilege are better able to secure this effect.

- 40 Again, I am speaking of cases where the *same* threat is posed. This is why I may be able to give up on a romantic partner with a certain unpleasant or morally objectionable habit without falling afoul of being a “not in my backyard” objection in the same way I may not be able to give up on a coworker with the same bad habit. For similar reasons, a person with a history of trauma (e.g., of a particular kind of violence; from being harmed by racism) might be able to withdraw from a person exhibiting such behaviors, because their retraumatization is different from the harm the remainder of the moral community faces.
- 41 We may be even more obligated to consider the broader implications of our choices for common *nonmoral* reasons to give up on a person (for example, we find them annoying or boring). When many members of the community choose to give up on the person for such reasons, it is felt as unkindness and exclusion—the kinds of behaviors that if we saw our children exhibit, we would intervene. For this reason, the corresponding obligation of due care would be to determine whether the person is lonely or has friends; in the case of the former, we have more of an obligation to continue to engage with them.

First, an objector might ask whether giving up on a person is something we genuinely have control over. Sometimes, I realize I have given up on someone almost without meaning to—it's not a choice; I was just so averse to their behaviors that I *find* myself closed off from them in one or the other sense.

In such cases, I would suggest that we still want to do a careful assessment of the normative considerations that may hold, and perhaps take action accordingly. After all, we may not have *direct* control over our first-pass reactions to objectionable behavior, but we certainly have control over our deliberations and actions that follow. I can reflect on those reactions, and decide whether there are indeed reasons to be open to the person in the future—or at least to gather more information, at minimum. For the first kind of giving up on a person, we have *indirect* control of our anger or resentment and other reactive attitudes (and, I suspect, the lack thereof). I can use various tactics to help me decide whether my objective attitude toward the person is unwarranted, and recall that they are a person—not just something to be managed or constrained.

A second kind of objector might ask whether it is too stringent a claim that giving up on a person (in either sense) is impermissible. After all, it seems like such an action does us a certain kind of service, particularly for those who think that our interests in autonomy or self-determination allow us to generally choose who we interact with and how we do so.⁴²

In response: I do think we have obligations to ourselves, but they do not go so far as allowing us complete flexibility to choose who to permanently close ourselves off from (though I note they *can* justify or excuse doing so in the special cases discussed in sections 2 and 3). We have an obligation to treat even objectionable persons as *persons*, and we have obligations to others that, in some cases, require us to stay open to the person or even continue to engage with them. I will note, however, that this obligation to not give up on others (in the manners and contexts described in sections 2 and 3) is not the kind of duty that is enforceable.⁴³ However, when we are personally considering what we ought to do, the considerations against giving up in those manners and contexts should shape how we proceed.

In order to understand the third kind of objection, we must first borrow some insights from Elinor Mason's description of "detached blame." In chapter 5 of her book *Ways to Be Blameworthy*, Mason describes a distinction between "ordinary" and "detached" blame.⁴⁴ Ordinary blame has the purpose of communicating

42 Thanks to an objector for suggesting this framing.

43 Thank you to an anonymous reviewer for suggesting this point.

44 Mason, *Ways to Be Blameworthy*, 102. Thanks to an anonymous reviewer for the reference to Mason.

that someone has failed on a shared value system, and that the wrongdoer can come to recognize the problem with their behavior. Yet there are agents from whom we cannot expect a response—they have so dissimilar a value system, or no value system at all—and, as a result, she says, the communicative act of ordinary blame would not be apt.⁴⁵ In such cases, detached blame can be of use to the blamer insofar as it allows them to “let off steam, to signal disapproval to her peers, to manage and manoeuvre around the offending agent.”⁴⁶

Detached blame is not quite the same thing as giving up in the first sense—Mason claims that it is a position in between ordinary communicative blame and a truly objective stance.⁴⁷ (The latter is what I mean by the first type of giving up.) She believes that detached blame might involve reactive attitudes of a different kind—contempt or disdain, perhaps—and, as such, detached blame may be better described as an example of giving up in the *second* sense (if we remain closed to the person).⁴⁸ Mason gives us reasons to think it is felicitous to deploy detached blame in the cases she specifies. Detached blame can indeed serve to protect oneself and communicate with others who share our own values, and motivates the blamer to withdraw from or avoid what she cannot change.⁴⁹ As she puts it, “There is no point in trying to communicate with such people, but we are bound to react to their trampling on the values we hold dear.”⁵⁰ Mason argues that those who are deeply morally ignorant are apt for detached blame.⁵¹ So perhaps we should think that similar considerations do favor giving up on the person (in the second sense) in the cases in which we cannot expect communicative uptake from the blamee, due to their possessing too dissimilar or no value system.⁵²

Yet in response to a Mason-style objection, I would make two points. First, I worry about the ways in which detached blame is susceptible to the kinds of concerns discussed in 3.3. We may fail to meet our obligations to other members of our community if we stop ourselves *at* detached blame (despite the goods it can secure) without due care, particularly if we do not consider what else might need to be done to protect those who might be otherwise harmed by the person. Second, she is right that we cannot expect communicative uptake from

45 Mason, *Ways to Be Blameworthy*, 103, 113.

46 Mason, *Ways to Be Blameworthy*, 116.

47 Mason, *Ways to Be Blameworthy*, 147, 121.

48 Mason, *Ways to Be Blameworthy*, 118.

49 Mason, *Ways to Be Blameworthy*, 142.

50 Mason, *Ways to Be Blameworthy*, 122.

51 Mason, *Ways to Be Blameworthy*, 149.

52 Thank you to an anonymous reviewer for pressing me on my framing here.

all people we might encounter—particularly those who have an exceedingly dissimilar value system. At the same time, as C. Thi Nguyen discusses in “Echo Chambers and Epistemic Bubbles,” perhaps there are other kinds of goals we may have in mind for such persons who not only have a dissimilar value system, but also systematically discredit reasons to believe other than they do—that is, they are in an echo chamber.⁵³ (Nguyen gives an extended discussion of a person with neo-Nazi beliefs as one such case.) Even though we cannot directly communicate blame or otherwise engage in a process of moral accountability or moral improvement with such a person, perhaps the strategy is different—we should cultivate trust that could help the person reboot their thinking about any flaws in their commitments.⁵⁴ Nguyen recognizes that this kind of transformation is difficult to secure and surely taxing, but it is one method that has realistically been able to change those with such commitments.⁵⁵

One final objection—in a strange result of the argument of section 3—is that it seems like it is more acceptable for me to give up on people if I am particularly bad at intervening with wrongdoers or eliciting moral improvement. After all, it would seem that I am not well-positioned to secure the desired good for the rest of the community, or to elicit moral improvement from the person in question. This point, however, is fairly straightforward to respond to: even if I am remarkably bad at helping others in their self-diagnoses or self-correction, I have a different kind of responsibility to *improve my skills*. A lack of skill at intervention, social protection, or moral education (depending on the features of the case) is not an excusing condition for not shouldering the obligations we have; it is simply an obligation to do the work involved in acquiring or improving those skills.

5. CONCLUSION

In conclusion, I would like to reaffirm that giving up on a person is not a practice that should proceed merely on the basis of personal preference. One form of giving up on a person—in which we hold a permanent objective attitude toward the target person—is morally impermissible, or at the very least morally inappropriate, on the grounds that doing so would fail to treat the person in question *as a person*. Even for the second kind of giving up on the person—in which we continue to hold the reactive attitudes toward the target person, but use other methods to protect ourselves from them—such a choice should be

53 Nguyen, “Echo Chambers and Epistemic Bubbles,” 146.

54 Nguyen, “Echo Chambers and Epistemic Bubbles,” 158–59.

55 Nguyen, “Echo Chambers and Epistemic Bubbles,” 158–59.

given due care, including considering (a) the fittingness of doing so, (b) any special obligations we have to the person, and (c) the broader implications of our choice for the community around us. In many common cases, we impermissibly weigh the harms to ourselves over the harms to the entire community. If I am capable of refraining from giving up on the person without significant harm to myself, and can do so in a way that will aid the other persons in the community, I may have an obligation to do so.

So what do we do when facing morally (and otherwise) criticizable persons? Many things, I think. We can be angry, we can shame them, we can tell them off, we can try to change their hearts, or we can ignore them until we have a chance to cool down. But what we should not do so easily—at least, not until we know we have sufficient *justification*—is give up on them as a person.⁵⁶

Phillips Academy, Andover
kbhardwaj@andover.edu

REFERENCES

- Alexander, Michelle. *The New Jim Crow*. New York: The New Press, 2012.
- Aristotle. *Nicomachean Ethics*. 2nd ed. Translated by Terence Irwin. Indianapolis, IN: Hackett, 1999.
- Calhoun, Cheshire. “Changing One’s Heart.” *Ethics* 103, no. 1 (October 1992): 76–96.
- . “What Good is Commitment?” *Ethics* 119, no. 4 (July 2009): 613–41.
- Darwall, Stephen L. “Two Kinds of Respect.” *Ethics* 88, No. 1 (October 1977):

⁵⁶ Work in philosophy is best recognized as a community endeavor, and I am grateful for the philosophers and other individuals who have allowed me to spend a great deal of time thinking with them about when it is permissible to give up on someone. A special thanks goes to two anonymous reviewers for the *Journal of Ethics and Social Philosophy*, whose excellent comments on the penultimate draft prompted me to deepen and sharpen the argument, and especially to Jordan MacKenzie, whose friendship and incisive advice helped me in how to do so. Other thanks go to Jen Kling, Josh Kissel, Rachel Murree, Ryan Ravanpak, Ryan Preston-Roedder, Doug MacLean, Tom Hill, Barry Maguire, Joshua Blanchard, Tamara Fakhoury, Lauren Townsend, Caleb Harrison, Larisa Svirsky, Roger Crisp, Carla Merino-Rajme, Russ Shafer-Landau, Camil Golub, and Eileen John, and audiences at the APA Eastern, University of Iowa, and University of North Carolina (and the organizing members of those conferences where I presented early versions of this paper). Thanks as well to many family members and friends who have been thoughtful interlocutors as well—particularly Meera Bhardwaj, Mary Bhardwaj, and Michael Boyle. A final thanks to Charles Mills, whose belief in this project will always be appreciated and remembered.

- 36–49.
- Emerick, Barrett. “Love and Resistance: Moral Solidarity in the Face of Perceptual Failure.” *Feminist Philosophy Quarterly* 2, no. 2 (2016): 1–21.
- Govier, Trudy. “Forgiveness and the Unforgivable.” *American Philosophical Quarterly* 36, no. 1 (January 1999): 59–75.
- Hay, Carol. *Kantianism, Liberalism, and Feminism: Resisting Oppression*. Basingstoke, UK: Palgrave Macmillan, 2013.
- Hill, Jr., Thomas E. “Servility and Self-Respect.” *The Monist*, 57, no. 1 (January 1973): 87–104.
- Holmgren, Margaret R. *Forgiveness and Retribution: Responding to Wrongdoing*. New York: Cambridge University Press, 2012.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals*. Translated by Arnulf Zweig. Edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- . *Metaphysics of Morals*. Translated by Mary Gregor. New York: Cambridge University Press, 1996.
- Langton, Rae. “Sexual Solipsism.” *Philosophical Topics* 23, no. 2 (Fall 1995): 149–87.
- Lobanov-Rostovsky, Christopher. “Sex Offender Management Strategies.” In *the Sex Offender Management Assessment and Planning Initiative*. Department of Justice. http://www.smart.gov/somapi/sec1/ch8_strategies.html.
- Mason, Elinor. *Ways to Be Blameworthy: Rightness, Wrongness, and Responsibility*. New York: Oxford University Press, 2019.
- Nguyen, C. Thi. “Echo Chambers and Epistemic Bubbles.” *Episteme* 17, no. 2 (June 2020): 141–61.
- Norlock, Kathryn J. “Giving Up, Expecting Hope, and Moral Transformation.” In *Reasonable Responses: The Thought of Trudy Govier*, edited by Catherine E. Hundleby, 174–90. Windsor, ON: University of Windsor, 2017.
- Preston-Roedder, Ryan. “Faith in Humanity.” *Philosophy and Phenomenological Research* 87, no. 3 (November 2013): 664–87.
- Strawson, Peter. “Freedom and Resentment.” In *Free Will*, 2nd ed., edited by Gary Watson, 72–93. New York: Oxford University Press, 2003.

KANT AND THE BALANCE OF MORAL FORCES

Santiago Sanchez Borboa

IN A MEMORABLE PASSAGE in the *Doctrine of Virtue*, Kant writes, “If we talk of laws of duty (not laws of nature) and indeed in outer relation of humans to each other, then we consider ourselves in a moral (intelligible) world, in which according to the analogy with the physical, the connection of rational beings (on Earth) is effected through *attraction* and *repulsion*.” He continues:

By means of the principle of **mutual love**, they are admonished constantly to *come closer* to one another, through that of the **respect** they owe one another, to keep themselves *at a distance* from one another; and should either of these great moral/ethical forces fail [*sittlichen Kräfte sinken*], “then nothingness (immorality) with gaping throat would swallow the whole realm of (moral) beings like a drop of water” (if I may here make use of Haller’s words in a different relation).¹

Here, Kant draws an intriguing analogy between the moral world of humans and the physical world of bodies. Unfortunately, it is unclear what exactly this moral-physical analogy amounts to or how Kant can justify it and the bold conclusion he draws from it, viz., that having two powers of moral repulsion (respect) and attraction (love) is a condition for the possibility of a moral world of humans.² Few commentators discuss Kant’s moral-physical analogy,

- 1 Kant, *Doctrine of Virtue*, 6:449. All citations to Kant except those to the *Critique of Pure Reason* are to the Akademie numbers listed in the margins of most editions. For the English translations of Kant’s cited texts, I have used the Cambridge Editions as a starting point, though I modify where appropriate. Kant’s talk of a “gaping throat” “swallowing the whole realm of moral beings like a drop of water” is a reference to Albrecht von Haller’s “Imperfect Poem about Eternity.” Marco Santi helpfully pointed out to me that von Haller was Kant’s favorite poet and that, in this poem, God is considered the ground of everything and God’s powers are considered as grounding time and eternity together in a way analogous to how the balance of moral forces in this passage grounds the realm of moral beings.
- 2 Here, Kant’s focus is practical rather than aesthetic respect and love. In this sense, respect and love are primarily maxims rather than feelings, respectively, “the maxims of limiting our self-esteem by the dignity in another person” and “the maxim of *benevolence*” (Kant, *Doctrine of Virtue*, 6:449). It is worth noting that Kant discusses other analogies between

but even fewer find it helpful. Those who do have yet to develop a reading of the analogy that makes sense of it. Some complain that Kant gets the nature of love wrong.³ Marcia Baron has a more nuanced and sympathetic take on the analogy but ultimately argues that it is untenable insofar as love and respect do not plausibly pull us in opposite directions, concluding that it “takes a good point too far” and that “love and respect are less different and less opposed than Kant suggests.”⁴ Commentators who have defended this analogy, such as Christine Swanton and Eleni Filippaki, have tended to do so by focusing on the moral side, arguing that it provides a fruitful ethical framework for thinking about moral relations such as friendship.⁵ Filippaki gives the most sustained treatment of this analogy in the literature. She argues that Kant’s dynamical theory of matter supports a fruitful view of moral relations, attending to how the tension between attractive and repulsive forces is key to Kant’s theory of friendship and his vision of moral relations in the *Metaphysics of Morals*. However, even Filippaki focuses mostly on the upshots of this analogy for Kant’s

the natural and moral in other practical works. As I discuss below, Kant puts different analogies to different uses for different purposes. My focus throughout is on the *Doctrine of Virtue’s* analogy between the physical world of bodies and the moral world of humans. I use the term “moral-physical analogy” or “the analogy” to refer to this particular analogy between the physical world of bodies and the moral world of humans.

- 3 Robert Johnson argues that what Kant calls love is really just a form of respect (“Love in Vain”). David Velleman argues against love being a form of coming closer, instead holding it is “like a state of attentive suspension, similar to wonder or amazement or awe” (“Love as a Moral Emotion,” 360).
- 4 Baron, “Love and Respect in the Doctrine of Virtue,” 41–42.
- 5 Swanton, *Virtue Ethics* and “Kant’s Impartial Virtues of Love”; and Filippaki, “Kant on Love, Respect, and Friendship.” Swanton follows Kant in claiming that respect and love are modes of moral response that involve keeping one’s distance and coming closer (*Virtue Ethics*, 99–110). She defends Kant’s distinction between respect and love but employs this distinction for her own pluralistic virtue ethical theory. In doing this, Swanton focuses on connecting this Kantian distinction with profiles of the virtues and her own view of virtues as dispositions to respond or acknowledge in an excellent way people, objects, situations, inner states, or actions falling in the field or scope of different virtues (*Virtue Ethics*, 1–2, 19–30). Swanton’s focus then is not Kant interpretation but the articulation of some of Kant’s resources for the purposes of developing her own rich view. Swanton does explicitly defend Kant’s conception of love as a moral force, but her defense is of this view of love against certain virtue ethical criticisms (“Kant’s Impartial Virtues of Love”). Filippaki argues that the “polarity” and “tension” between respect and love underlies all moral relations. According to her, the “picture of Kantian morality that emerges against this background is thus one where individuals constantly strive both to retain their agency and open up to others by acknowledging and embracing ends other than one’s own, hence finding themselves in a constant struggle for balance” (“Kant on Love, Respect, and Friendship,” 23). Swanton’s and Filippaki’s takes on this strike me as insightful and correct, so far as they go, but we can go further in understanding the analogy itself and its significance.

moral philosophy rather than on making sense of the analogy itself, admitting that a “deeper, comparative investigation of the role of attraction and repulsion as metaphysical principles in the *Metaphysics of Natural Science* and the *Metaphysics of Morals* . . . goes beyond the scope of [her] paper.”⁶

In this paper, I undertake this “deeper investigation,” giving a systematic account of this moral-physical analogy, one that explains the use to which Kant puts it in his philosophy and that spells out how the different elements of the physical and moral worlds are analogous to each other. My reading interprets the moral-physical analogy as an instance of Kant’s use of analogies to determine the content of ideas (whose objects we cannot sense) in a certain way, viz., by giving us a sufficient grasp of their formal character.⁷ By “sufficient grasp of the formal character of an idea,” I mean a grasp of the relations the supersensible object of an idea constitutively bears that allows reason to put it to purposive use, i.e., to use for its characteristic ends.⁸ Kant holds that, in order to gain such a sufficient formal grasp, reason must associate intuitions of sensible objects with ideas indirectly, in a way that allows us to think of the supersensible object of the idea as bearing similar relations to their symbol objects despite having different intrinsic properties.

In section 1, I discuss the context for this moral-physical analogy and the role analogies like this one play in Kant’s philosophy. In section 2, I sketch Kant’s view of the physical world in the *Metaphysical Foundations of Natural Science*. In section 3, I apply the account of analogies from section 1 to the view of the physical world developed in section 2 in order to give a systematic account

6 Filipkaki, “Kant on Love, Respect, and Friendship,” 42.

7 I thank Jim Conant for very helpfully suggesting the terminology of the “formal character” of an idea to me.

8 My view of the role of symbolization/analogy in Kant thus differs from Andrew Chignell’s. Chignell holds that for Kant symbols/analogies allow us to gain a fragmentary grasp of what it would be like for the idea to be instantiated in an individual and so of what it would be for its supersensible object to exist, giving us a limited sense of whether the symbolized object is really possible (“Are Supersensibles Really Possible?” “Beauty as a Symbol of Natural Systematicity,” “The Devil, the Virgin, the Envoy,” and “Real Repugnance and Belief about Things-in-Themselves”). On my view, symbolization/analogy achieves something more precise and less ambitious: it gives some determinate content to ideas, a formal content or character that consists of thinking of their supersensible objects as bearing certain relations. This does not yet settle the question of whether the objects of ideas are really possible. Settling that question requires that one successfully put the symbolized ideas to purposive use. In a companion paper, I spell out how symbolization/analogy of metaphysical ideas like substance and ground allows us to gain a formal grasp of them that then allows us to make purposive use of them in theoretical philosophy, in the service of reason’s constitutive ends of seeking unifying explanations and, ultimately, the unconditioned (Sanchez Borboa, “Making the Supersensible Intuitive”).

of the moral-physical analogy and of what it achieves in the *Doctrine of Virtue*. On the basis of this application, I argue that in this passage Kant endorses a moral analog of the balancing argument from his natural philosophy. There, Kant argues that purely repulsive bodies face a problem of total dispersion while purely attractive bodies face a problem of total collapse. In doing so, he offers an *a priori* argument that explains the possibility of bodies (as genuine objects of outer experience) as grounded in the balancing opposition of bodies' physical repulsion (expansion) and attraction (gravitation). I argue that, analogously, purely respectful humans face a problem of total moral dispersion while purely loving humans face a problem of total moral collapse. That is, I interpret this passage as offering an *a priori* argument that explains the possibility of humans (as genuine embodied subjects of moral duties) as grounded in the balancing opposition of humans' moral repulsion (respect) and attraction (love). I discuss some interpretive payoffs of reading this analogy this way. Finally, I conclude by drawing some implications for contemporary theorizing about moral interactions, relationships, and communities.

1. CONTEXT FOR THE ANALOGY

1.1. *Kant's Twofold Metaphysics and the Privileged Place of <Matter/Body> and <Human>*⁹

Kant discusses the analogy between the physical world of bodies and moral world of humans within the *Doctrine of Virtue*, a work where Kant gives his normative ethical theory. Here, Kant applies the principles of morality (established and defended in the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*) to the concept of a human as a finite rational being to derive a system of ethical duties that apply to all humans. As such, this work occupies a position in Kant's practical philosophy that is parallel to that of the *Metaphysical Foundations of Natural Science* in his theoretical philosophy. There, Kant applies the principles of a general metaphysics (developed in the *Critique*

9 Throughout the paper, my focus is the empirical concept <matter> as the movable in space insofar as it fills space. Note that Kant allows for a nonempirical concept of matter (*Critique of Pure Reason*, A359). All citations to the *Critique of Pure Reason* are given by the standard A and B pagination of the first (1781) and second (1787) editions, respectively). Despite the analogous roles the concepts of <matter> and <human> play within Kant's theoretical and practical systems, this should not be taken to imply that there are not important differences between these two concepts. In particular, Kant's treatment of <matter> is more complicated, since for Kant (given his Aristotelian intellectual heritage) we arguably only have one empirical concept of <human> that is relevant for moral philosophy, that of <rational animal>.

of *Pure Reason*) to the concept of matter as movable in space in order to derive a system of metaphysical principles that applies to corporeal nature as such.¹⁰ In these two works, then, the *a priori* principles of pure understanding and pure practical reason are applied to empirical concepts in order to derive principles that we can apply as empirically determinable subjects engaging, respectively, in natural science and ethical deliberation. This application of intellectual principles to empirical concepts requires that these empirical concepts meet certain *a priori* requirements. If the concept <matter> is to be the concept of the relevant object of outer experience, it must meet certain *a priori* requirements that allow its object to fill a region of space determinately (viz., that bodies are endowed fundamental physical powers that ground the possibility of filling space). Similarly, if the concept <human> is to be the concept of the relevant rational/sensible subject of morality—i.e., the concept employed in ethics—it must meet certain *a priori* requirements that allow its objects to form moral communities, viz., that humans are endowed with fundamental moral powers that ground the possibility of moral community.

The fact that Kant employs these central concepts of <human> and <matter> in making the analogy is reason to take this analogy seriously as a claim that is central to his twofold metaphysics of nature and morals, rather than as a mere metaphor from Newtonian physics that “takes a good point too far.”¹¹ That Kant takes the analogy to have broader significance for his philosophical system is supported by the fact that in the conclusion to the “Elements of Ethics,” he characterizes friendship (considered in its perfection) as “the union of two persons through equal mutual love and respect,” noting that “love can be regarded as attraction and respect as repulsion, and if the principle of love bids friends to draw closer, the principle of respect requires them to stay at a proper distance from each other.”¹² The fact that Kant uses the moral-physical analogy to explain the moral phenomenon of friendship suggests that Kant takes this analogy to be more than a stylistic metaphor.

With this context for the analogy on the table, I turn to discuss the role that analogies such as this one play in Kant’s critical philosophy.

1.2. Analogies and Symbols as a Way of Thinking about Determinate Supersensible Relations (and Concepts)

Kant uses symbols and analogies in many places in different ways and for different purposes in his philosophy. Some symbols/analogies seem indispensable to

10 Cf. Kant, *Metaphysical Foundations of Natural Science*, 4:469–77.

11 Baron, “Love and Respect in the Doctrine of Virtue,” 41, 42.

12 Kant, *Doctrine of Virtue*, 6:469, 6:470.

the nature of our cognitive capacities (as when reason must think of a being that is responsible for metaphysical explanations as a “ground”).¹³ Others seem to be used as pedagogical tools for better understanding a particular philosophical project (as when juridical analogies are used to describe the project of critiquing pure reason or when chemical analogies are used to describe proper moral-philosophical methodology).¹⁴ Still other analogies have the purpose of inspiring us to live up to our moral vocation and so aid us in moral motivation, or, as Kant puts it, to “bring an idea of reason [the categorical imperative] closer to intuition and thereby to feeling” and to effect in us a “lively interest in the moral law.”¹⁵ Heiner Bielefeldt (who has arguably undertaken the most ambitious study of the role symbols play in Kant’s philosophy to date) contends that Kant “does not provide a comprehensive typology of symbols, nor does he develop a system enabling us to gain an overview of how precisely the various forms of symbolic representation are interrelated.”¹⁶ I agree with Bielefeldt that there is no single general account of symbols or analogies in Kant. However, I think we can give a systematic account of a key species of use of symbols/analogies.

The relevant species of symbol/analogy consists of the use of a sensible object as a symbol for thinking of a supersensible/intelligible object represented by an idea in a certain way. Kant employs this use of symbols to solve a problem arising from the nature of reason: that reason can never cognize things in themselves, yet it must nevertheless think of them as grounding things as they constitute objects of our experience. By its nature then, reason must think of certain supersensible things it cannot cognize. It cannot cognize these things because it has no intuition of them. Because reason has no such intuitions, it cannot determinately think of these concepts of reason, for it has no guarantee that it is not merely saying empty words with no real content in attempting to think of things using these concepts. In order to successfully grasp a concept through which we can think of an object, we need to associate some intuition with the concept, but we seem unable to do so for ideas, as we can have no intuition of supersensible objects. The problem then is that it seems impossible

13 Kant, *Critique of the Power of Judgment*, 5:352.

14 Kant, *Critique of Pure Reason* Axi–xii, A501/B529, A787/B81, and *Critique of Practical Reason*, 5:92, 5:163.

15 Kant, *Groundwork of the Metaphysics of Morals*, 4:436, 4:463. Other uses of symbols include the one Alexander Rueger and Şahan Evren highlight that connects natural beauty and the systematicity of nature through the form of reflective judgments of taste (“The Role of Symbolic Presentation in Kant’s Theory of Taste”). Chignell notes that Kant uses the forms of both judgments of taste and of beautiful objects as symbols for the systematicity of nature (“Beauty as a Symbol of Natural Systematicity”).

16 Bielefeldt, *Symbolic Representation in Kant’s Practical Philosophy*, 180.

for reason genuinely to think something that it must: supersensible objects of ideas of reason.

Kant solves this problem by means of analogies between sensible and supersensible objects in which we use sensible objects as symbols for thinking about the supersensible ones.¹⁷ This method of symbolization or analogy allows reason to borrow the intuitions of sensible objects (which it can experience) and use them as a guide for thinking about supersensible objects of ideas (which it otherwise could not do). Such analogies determine the contents of ideas “for us,” though without determining them unconditionally and “in themselves,” i.e., in a way that helps us make certain uses of ideas but not in a way that determines the intrinsic properties of their objects.¹⁸ These analogies determine the content of ideas in this way by giving us a grasp of the formal character of an idea, which we get by thinking of its supersensible object as analogous to a sensible symbol object. This formal grasp is gained by thinking of the supersensible object of the idea as bearing similar relations as the sensible, symbol objects.¹⁹ As Kant puts it, cognition “according to analogy” signifies not “an imperfect similarity between two things, but rather a perfect similarity between two relations in wholly dissimilar things.”²⁰ This kind of use of symbols/analogies therefore allows us to grasp the formal character of ideas, allowing us to think of their supersensible objects without ascribing any sensible properties to these objects, instead merely thinking of these as bearing similar relations to their symbol objects.²¹ By means of such analogies then, we “can ... provide a concept of a relation of things that are absolutely unknown [*unbekannt*] to [us],” allowing us to think of

17 In a companion paper, I spell out in detail what is at issue in this species of use of symbols and analogies and discuss the key role it plays across Kant’s critical philosophy (Sanchez Borboa, “Making the Supersensible Intuitive”).

18 Kant, *Prolegomena to Any Future Metaphysics*, 4:358.

19 Kant uses the German, *Ähnlichkeit* and *ähnlich*, rather than equality, *Gleichheit* and *gleich*, to describe these relations. As such, the relation that supersensible objects bear need not be strictly the same as their symbol objects.

20 In section 90 of the third *Critique* there is a parallel passage confirming the idea that in analogies we think there is a similarity that is not a similarity in properties or marks: “One can, of course, think of one of two dissimilar things, even on the very point of their dissimilarity, by means of an *analogy* with the other; but from that respect in which they are dissimilar we cannot *draw an inference* by means of the analogy, i.e., transfer this mark of the specific difference from the one to the other” (Kant, *Critique of the Power of Judgment*, 5:464).

21 Insofar as the relations thought in analogies hold for both sensible and supersensible objects, they must be understood as extrinsic relations, which are not determined by the intrinsic properties of their relata (cf. Moore, “External and Internal Relations”; Lewis, *On the Plurality of Worlds*).

these supersensible objects.²² Once reason gains a grasp of the formal character of these ideas (i.e., determines a rational way of thinking of these objects as bearing certain relations) through such analogies, it can think of the objects of ideas as determinately bearing certain relations. The formal grasp of these ideas acquired through these analogies thereby allows reason to make use of these ideas (thought of in this determinate way, i.e., with their supersensible objects bearing similar relations to their sensible objects) in the service of the ends of our cognitive capacities, especially reason and pure understanding.²³

We can elucidate this use of symbols/analogies using examples like the *Prolegomena's* example of thinking of God as a clockmaker.²⁴ Here, Kant uses a sensible object (a clockmaker) as a symbol for thinking about a supersensible object (God). By drawing an analogy between God and a clockmaker in this way, reason comes to think of God and its relation to nature (the sensible world) as similar to the relation between a clockmaker and a clock: a relation of an intelligent causality (be it God or the clockmaker) that produces a whole constituting a dynamical order (be it nature or the clock). Through this analogy, we do not attribute to God any sensible properties of the clockmaker. We instead think of the sensible world and its relation to God by thinking of it as bearing a similar relation to God as the clock bears to the clockmaker. This analogy makes the concept of God “sufficiently determinate *for us*,” i.e., for our reason and its ends, without determining the idea of God “unconditionally and *in itself*,” thus without making claims about what the supersensible object of this idea is in itself.²⁵ In this way, the analogy allows us to determinately think of God as bearing the relation of being an intelligent creator of a dynamical order. This symbolization thereby allows reason to grasp the formal character of the idea of God and to make use of it (thought of in this determinate way, i.e., with its object bearing a relation of being an intelligent causality of a dynamical order) for its theoretical and practical ends, e.g., explaining the systematicity of nature or the possibility of the highest good.²⁶

22 Kant, *Prolegomena to Any Future Metaphysics*, 4:357n. I thank Angela Breitenbach for pointing out this passage and its significance for this paper.

23 Note that the actual success condition for this kind of analogy is giving enough determinate content/formal grasp to an idea that you can put it to use for the ends of a cognitive capacity. Thus, by making this kind of analogy, you strictly speaking get a candidate for the determinate content/formal grasp of an idea that might be put to such an end. However, to see whether it is the right determinate content/formal grasp, you need to actually succeed in making use of the idea to achieve ends of our cognitive capacities.

24 Kant, *Prolegomena to Any Future Metaphysics*, 4:357–60.

25 Kant, *Prolegomena to Any Future Metaphysics*, 4:358.

26 Kant, *Critique of Pure Reason*, A697–702/B725–30, and *Critique of Practical Reason*, 5:124–29.

Another example of the species of symbolization/analogy that I focus on is the *Groundwork's* analogy between the realm of nature and the realm of ends.²⁷ Here the sensible object of a realm of nature is used as a symbol for reflecting on the supersensible object that is the realm of ends.²⁸ By symbolizing the idea of the realm of ends in this way, reason comes to think of the relations among members of the realm of ends as similar to the relations among members of the realm of nature: relations that constitute a universal dynamical order connected according to common universal laws. This analogy gives us a formal grasp of the idea of a realm of ends by allowing us to think of the realm of ends as a realm in which rational beings bear similar relations to the relations different natures bear in the realm of nature. This symbolization of the idea of the realm of ends thereby allows reason to make use of it (thought of in this determinate way, i.e., with members of this realm bearing relations that constitute a universal dynamical order connected according to common laws) for its practical ends. These purposive uses include (a) explaining the requirements of the categorical imperative for finite rational beings sharing a world and (b) serving as a moral ideal that enlivens the moral interest in us and thereby inspires us to live up to our moral vocation.²⁹

On my view, the moral-physical analogy I began with is a more determinate version of the one between the realm of nature and the realm of ends in the

27 Kant, *Groundwork of the Metaphysics of Morals*, 4:438. As Ralf Bader helpfully pointed out to me, one might worry that the realm of nature (and the physical world) cannot be used as symbols in the way I propose because these are not objects of intuition. In response, I note, following Houston Smit, that we should distinguish between Kant's use of "intuition" as a singular representation and as a species of cognition consisting in part of intuitions in the former sense (call the latter intuitive cognition) (Smit, "Intuition"). Given this distinction, one can make a corresponding distinction between objects of sensible intuition and objects of intuitive cognitions. Objects of intuitive cognitions are objects that we cognize by means of a concept of them and an intuition of them (as a whole). Objects of mere (sensible) intuition are objects of which we can have some intuitions (perhaps only of their parts), though we are not able to have intuitive cognition of them. On my view, the realm of nature and the physical world are not objects of intuitive cognitions, but they are objects of mere intuition.

28 I should note that in interpreting the realm of nature in this way in this passage, I am giving what we might call a *dynamical* as opposed to *teleological* reading of the realm of nature in this part of the *Groundwork* (like the one given by, e.g., Allison (*Kant's Groundwork for the Metaphysics of Morals*, 249, 257–58)). According to this interpretation, the realm of nature that is analogous to the realm of ends is thought of not teleologically (as a teleological system of natural ends) but rather merely dynamically (as a dynamical system of things insofar as their existence is determined in accordance with universal laws). Though I cannot make the full case for this here, my interpretation is supported by Kant's point that in this analogy "nature as whole" "is regarded as machine" (Kant, *Groundwork of the Metaphysics of Morals*, 4:438).

29 Cf. Kant, *Groundwork of the Metaphysics of Morals*, 4:436, 4:662.

Groundwork. According to this first analogy, we should think of the moral world of humans as a universal dynamical order of beings connected according to particular kinds of universal common laws, viz., governing attractive and repulsive forces between the beings that make up this dynamical order. I argue that in drawing the analogy between the moral and physical worlds, we gain a formal grasp of the idea of a moral world of humans. This formal grasp allows reason to make practical, purposive use of this idea to give an *a priori* moral-balancing argument that is analogous to the physical-balancing argument in his natural philosophy. I shall also argue that this analogy (and its determination of the idea of a moral world) allows reason to put this idea to use as a moral ideal of community and interactions between humans that can guide our conduct. I now turn to sketch Kant's view of the physical world as grounded on the interaction of fundamental attractive/repulsive forces that binds bodies together.

2. BALANCING PHYSICAL REPULSION AND ATTRACTION

In the "Dynamics" chapter of the *Metaphysical Foundations of Natural Science*, Kant explains the possibility of matter's essential property of filling space. He claims that "matter is the *moveable* in space, insofar as it *fills space*" where "to *fill* a space is to resist every moveable that strives through its motion to penetrate into a certain space."³⁰ In order to fill space, Kant argues, matter must be endowed with two fundamental forces or powers (*Grundkräfte*): repulsion and attraction. He characterizes this repulsion as that moving force (*bewegende Kraft*) "by which a matter can be the cause of others removing themselves from it" and attraction as "that moving force by which a matter can be the cause of the approach of others to it."³¹ Employing these conceptions of repulsive and attractive force, Kant gives a balancing argument comprising two sub-arguments that together establish that, in order for matter to determinately fill space, it must jointly exercise physical repulsion and attraction.³²

30 Kant, *Metaphysical Foundations of Natural Science*, 4:496.

31 Kant, *Metaphysical Foundations of Natural Science*, 4:498, 4:498. Note that, for Kant, bodies can be compressed to infinity "but can *never* be *penetrated*" by another matter because their repulsive force becomes infinite "when compressed into an infinitely small space" (*Metaphysical Foundations of Natural Science*, 4:501). Thus, physical repulsion grounds bodies' physical impenetrability.

32 As recent scholarship (e.g., Edwards, *Substance, Force, and the Possibility of Knowledge*; Watkins, "Kant on Extension and Force"; Friedman, "Metaphysical Foundations of Natural Science") emphasizes, Kant's natural philosophy (and his dynamical theory of matter) enjoyed a rich development from his precritical *Monadology* to the critical *Metaphysical Foundations of Natural Science*, to the *Opus Postumum*. However, as Filipkaki notes, despite the significant changes Kant's natural philosophy undergoes (especially regarding

The first half of the argument, as Daniel Warren notes, “moves from the supposition of repulsive force to the need for attractive force.”³³ Kant here asks us to consider purely repulsive matter. Such matter would expand without any force to limit or curb this expansive repulsion and so would “disperse itself to infinity” such that “no specified quantity of matter would be found in any specified space.”³⁴ “Therefore, with merely repulsive forces of matter, all spaces would be empty, and thus, properly speaking, no matter would exist at all,” for matter essentially fills space.³⁵ Thus, if matter is endowed with a repulsive force that grants it a capacity for expansion, it must also be endowed with a counteracting attractive force that grants it a capacity for compression.

The second half of the argument “closely parallels the first, but in reverse.”³⁶ Kant here asks us to consider purely attractive matter. Such matter would be such that “all parts of the matter would approach one another without hindrance and diminish the space that matter occupies.”³⁷ This would make it the case that the parts of matter “coalesce in a mathematical point.”³⁸ Therefore, if there were only attractive forces, “space would be empty, and thus without any matter” for matter essentially fills space.³⁹ Thus, if matter is endowed with an attractive force that grants it a capacity for compression, it must also be endowed with a counteracting repulsive force that grants it a capacity for expansion.⁴⁰

questions of the divisibility of matter and space), one constant is “the postulate of attractive and repulsive forces” (“Kant on Love, Respect, and Friendship,” 28). She is right to claim that Kant remains committed to the idea that if matter is to have the essential property of filling space, this must be grounded in matter’s essentially exercising both attraction and repulsion. However, this is not so much a postulate as the conclusion of a certain kind of argument: a balancing argument.

33 Warren, “Kant on Attractive and Repulsive Force,” 195.

34 Kant, *Metaphysical Foundations of Natural Science*, 4:508. Kant’s reasoning here seems to depend on the idea that a finite quantity of matter distributed in an infinite space would result in there being effectively zero quantity of matter in any finite subregion of that infinite space. This raises complications concerning Kant’s conception of infinity that lie outside the scope of this paper.

35 Kant, *Metaphysical Foundations of Natural Science*, 4:508.

36 Warren, “Kant on Attractive and Repulsive Force,” 196.

37 Kant, *Metaphysical Foundations of Natural Science*, 4:511.

38 Kant, *Metaphysical Foundations of Natural Science*, 4:511.

39 Kant, *Metaphysical Foundations of Natural Science*, 4:511.

40 As Sheldon Smith notes, a heuristic that Kant sometimes employs for seeing these problems is to imagine a stable configuration of matter exercising both forces and then imagine turning one or the other off and seeing what occurs to the configuration with only one force (“Does Kant Have a Pre-Newtonian Picture of Force in the Balancing Argument?” 476). This is how Kant explicitly thinks of the balance argument in, e.g., the *Opus Postumum*: “If the attraction of the internal cohesion in matter were suddenly to cease

By means of this argument, Kant infers that, without two forces of attraction and repulsion to balance each other, matter cannot possess the essential property of filling space. Matter endowed solely with physical repulsion faces a problem of total dispersion. Matter endowed solely with physical attraction faces a problem of total collapse. In either case, matter cannot possess the essential property of filling space through the exercise of its powers, so, strictly speaking, either kind of unbalanced matter is impossible. Kant concludes that in order for matter to be genuine matter (a movable in space that fills space through the exercise of its powers) it must jointly exercise expansion and gravitation. Without such joint repulsion and attraction, matters cannot relate to each other as they essentially must if they are to constitute objects of outer experience, and so there can be no matter or physical world at all.⁴¹

It is worth highlighting the argumentative role that balancing arguments play. Watkins and Stan give a helpful general characterization of such arguments as

an existence proof for a type of force. Its premises are (1) an accepted universal fact, viz. that a certain stable configuration obtains; and (2) a type of force independently known to exist. The argument seeks to prove that the stability in question is impossible unless a second kind of force exists to balance the first kind.⁴²

As such, balancing arguments generally comprise two sub-arguments, each of which (a) assumes unbalanced beings possessing a single force for the purposes of a *reductio ad absurdum* argument and (b) aims to show that, on the supposition of a single force, the stable configuration is not possible. This implies that the way the *reductio* is established for each unbalanced force can be importantly different. Indeed, in this physical case, there is a way in which the impossibilities at issue in the two *reductio* cases are different. For in the purely repulsive case, bodies fail to even have a specifiable physical location. At least in the purely attractive case, bodies all occupy the same point in space (though they fail to fill it).

completely, matter would extend itself infinitely, and, if repulsion ceased, matter would coalesce into one point" (*Opus Postumum*, 21:409).

41 I agree with Smith that Kant typically starts his thinking about the balancing argument by "thinking in terms of two-body problems" ("Does Kant Have a Pre-Newtonian Picture of Force in the Balancing Argument?" 478n24). Smith notes that "Kant tries to learn what he can from thinking of the two-body case with the hopes of being able to derive the information about the more complex cases from there" ("Does Kant Have a Pre-Newtonian Picture of Force in the Balancing Argument?" 477).

42 Watkins and Stan, "Kant's Philosophy of Science."

In articulating the structure of the physical world, Kant attributes the following relations to material substances: R-rep(x_n), a relation of repulsion, of keeping their distance; R-att(x_n), a relation of attraction, of coming closer; and R-bal(x_n), a relation of balancing opposition between these forces.⁴³ Moreover, he attributes to the two kinds of unbalanced matter either R-col(x_n), a relation of collapse (in the case of matter bearing only R-att(x_n)), or R-dis(x_n), a relation of dispersion (in the case of matter bearing only R-rep(x_n)). As I spell out in the next section, by symbolizing the moral world using the physical world, the moral-physical analogy allows us to think of humans as bearing similar relations and so as making up a moral world whose possibility is grounded in a balancing opposition of repulsion and attraction.

3. BALANCING MORAL REPULSION AND ATTRACTION

3.1. *Symbolizing the Moral World by Analogy to the Physical World*

In the passage I began with, Kant claims that we consider ourselves “in a moral (intelligible) world where, by analogy with the physical world, the connection of rational beings (on Earth) is effected through *attraction* and *repulsion*.”⁴⁴ I propose that Kant here uses the physical world (and its constitutive balancing opposition of repulsion and attraction) as a symbol for thinking about the moral world. Recall that an analogy between a sensible and a supersensible object enables us to think of the supersensible object determinately (and gives us a sufficient formal grasp of this idea) by thinking of it as bearing similar relations to the sensible object. In the moral-physical analogy, the relevant sensible object consists of bodies as they make up a physical world whose possibility is grounded in a balancing opposition of fundamental forces of (physical) repulsion and attraction. By means of this analogy, we come to think of humans as making up a moral world whose possibility is grounded in a balancing opposition of fundamental forces of (moral) repulsion and attraction. By symbolizing the idea of the moral world then, we think determinately of a moral world of humans as constituted by a balancing opposition of repulsion (respect) and attraction (love) between them, just as the physical world is constituted by a balancing opposition of repulsion (expansion) and attraction (gravitation)

43 The variables in these relations range over bodies and humans, where n = the number of bodies/humans in the world. Note that there are determinate aspects of these relations in their physical realization that will not transfer to their moral realization, e.g., love is a form of attraction like the physical force of gravitation, but it need not be describable by an inverse-square law, as gravitation is.

44 Kant, *Doctrine of Virtue*, 6:449.

between bodies. We therefore think of humans in a moral world as bearing similar relations to each other as bodies in a physical world do: R-rep(x_n), R-at(x_n), and R-bal(x_n).

I now unpack this sketch to give a systematic account of this analogy and the moral-balancing argument. This unpacking proceeds in stages. I first explain how respect and love are forces of repulsion and attraction. I then discuss how they are united in morally ideal interactions between humans. Finally, I give the moral-balancing argument.

3.2. *Respect as Repulsion and Love as Attraction*

First, note that Kant's characterizations of moral forces, like those of physical forces, rely on some notion of distance. He talks of love making us "come closer" and respect making us "keep at a distance." However, the notion of Euclidean distance cannot be applied to the moral side of the analogy, as it is inherently spatial and so cannot span both sides of the analogy. Instead, Kant seems to have some analogous notion of distance in mind, which seems to be a function of the way different humans' ends relate. Exercises of respect and love determine humans' wills to set and pursue ends in a way that, respectively, maintains and diminishes moral distance.

Moral repulsion or respect is the motivational force by which humans are motivated by the thought of duty to comply with one of the two kinds of fundamental duty humans have to one another as such: duties of respect "not to degrade any other to a mere means to my ends (not to demand that another throw himself away in order to slave for my end)" and to "keep oneself within one's own bounds."⁴⁵ This respect is a form of moral motivation that treats others as "self-standing [*selbständigen*]" ends "which must never be acted against," i.e., as sources of inviolable moral worth.⁴⁶ This moral power of respect keeps humans at a distance insofar as it grounds humans' moral impenetrability (and the keeping of moral distance between distinct humans) as bodies' expansive physical repulsion grounds their physical impenetrability (and the keeping of physical distance between distinct bodies). In both cases, the power of repulsion grounds the integrity and impenetrability of the beings of the world.

Moral attraction or love is the motivational force by which humans are moved by the thought of duty to comply with the other kind of fundamental duty humans have to each other as such: duties of love, "duties to make others' ends my own (provided only that these are not immoral)" and that put another

45 Kant, *Doctrine of Virtue*, 6:450. I thank Houston Smit for helping me see the need to clarify that exercising these moral forces consists of being motivated by the thought of duty to comply with duties of respect and love.

46 Kant, *Groundwork of the Metaphysics of Morals*, 4:437.

under obligation by carrying them out.⁴⁷ This love/benevolence is a form of moral motivation that treats the fact that something is another's (morally permissible) end as itself a (normative and motivational) reason to adopt that end, as *their* end. That is, practical love/benevolence consists of making it your project to help another achieve their project (as they conceive of it) from the motive of duty. Practical love brings humans morally closer because it grounds humans' coming to share ends in this distinctly moral way as bodies' gravitational physical attraction grounds bodies' coming physically closer. In both cases, the power of attraction grounds a closer dynamical connection between the beings of the world.⁴⁸

Two things are worth noting about moral forces and their relation to moral distance. First, although my discussion of the power and hence duties of love has focused on their being duties to adopt another's ends, Kant in other passages frames duties of love/benevolence in terms of others' happiness.⁴⁹ Human happiness for Kant includes more than simply the achievement of humans' ends. It consists of inclinations, desires, and subjective feelings of agreeableness with one's existence satisfaction.⁵⁰ But a human's ends and their happiness are intimately connected, for Kant holds that finite rational beings, by the very

47 Kant, *Doctrine of Virtue*, 6:450.

48 As a reviewer helpfully pointed out, one might wonder what the relationship is between practical love and what we might call the positive side of respect discussed in the *Groundwork*. Here, the second formulation of the Categorical Imperative claims that we should act so that "you use the humanity in your own person as well as in that in the person of any other, always at the same time as an end, never merely as a means" (*Groundwork of the Metaphysics of Morals*, 4:429). Thus, there seems to be a negative side of respect (never treating the humanity merely as a means) as well as a positive side of respect (always treating the humanity in others merely as a means). This positive side of respect seems to motivate us to treat their humanity, i.e., the rational nature of other humans, as an end in itself and so their ends as our ends. As Kant notes, for the representation of humanity as an end in itself "is to have its *full* effect in me, the ends of a subject that is an end in itself must, as much as possible, also be *my* ends" (*Groundwork of the Metaphysics of Morals*, 4:430). The practical love in the *Doctrine of Virtue* is a more determinate version of this positive respect, which motivates us to share in other ends by considering that humans are a particular species of finite rational beings on earth who require each other's help to achieve their ends. This is in keeping with the place of these two works in Kant's philosophical system. The *Groundwork* establishes and legitimizes the supreme principle of morality (*Critique of Practical Reason*, 5:8). The *Doctrine of Virtue* then applies this principle to the empirical concept of a human (as a particular finite rational being) to derive a system of duties that apply to humans as such.

49 For example, Kant, *Doctrine of Virtue*, 6:387–88. I thank an anonymous reviewer for pointing out to me the need to address this.

50 Kant, *Groundwork of the Metaphysics of Morals*, 4:393, 4:395; see also *Critique of Practical Reason*, 5:22.

nature of their embodied practical reason, set the end of happiness as a “commission from the side of sensibility.”⁵¹ Each human has as an end their own happiness, but happiness is itself an indeterminate concept.⁵² As such, each human has their own determinate conception of their own happiness that they actually try to bring about. This is entirely dependent on individual humans’ likes and dislikes. As Kant puts it, “in what each of us has to put his happiness comes down to the particular feeling of pleasure and displeasure in each.”⁵³ By seeing that all finite rational beings, including humans, set their own particular conception of happiness as their end, we can see that duties of love, to make another’s morally permissible ends our own, essentially include duties to adopt the morally permissible happiness of another, with its own particular shape.

Second, the particular degree of moral distance between humans varies and depends on how closely knit the moral community they form is, i.e., on how many of their ends they share in this distinctly moral way. The more of another human’s morally permissible ends one human adopts as their ends simply because they are that other human’s ends, the closer this human is morally to the other. And because being motivated to comply with duties of love, i.e., exercising practical love, puts the beneficiary under obligation, the closer this latter human ought to be morally to the first.⁵⁴ That is, those whom we love more ought to love us more, and we ought to love more those who love us more.

With these characterizations of the moral forces of repulsion and attraction and of how they maintain/diminish moral distance, I turn to explain how these two forces are at play in morally ideal interactions between humans.

3.3. *Uniting Respect and Love in Morally Ideal Interactions between Humans*

In the discussion following the moral-physical analogy, Kant claims that the moral law “essentially connects” these two kinds of duties in morally ideal interactions between humans. He writes that the two kinds of duties “are essentially always connected together with one another in one duty according to the law [*im Grunde dem Gesetze nach jederzeit mit einander in einer Pflicht zusammen verbunden*], only in such a way that now one duty and now the other is the subject’s principle with the other joined to it as accessory [*accessorisch geknüpft*].”⁵⁵ Kant here claims that the moral law unites both kinds of duties to other

51 Kant, *Critique of Practical Reason*, 5:61; see also *Groundwork of the Metaphysics of Morals*, 4:415, 4:418.

52 Kant, *Groundwork of the Metaphysics of Morals*, 4:418.

53 Kant, *Critique of Practical Reason*, 5:25.

54 Cf. Kant, *Doctrine of Virtue*, 6:450.

55 Kant, *Doctrine of Virtue*, 6:448.

humans as such. This seems to imply that morally ideal interactions between humans (where they fully treat humans as ends in themselves) constitutively require complying with both duties of respect and of love and so jointly exercising these moral powers.

The example Kant gives supports this reading. It is an example of how benevolent action, if it is to be genuinely moral, must also be respectful. He notes that we ought to be benevolent to someone poor in a respectful way: “it is our duty to behave as if our help is either merely what is due him or but a slight service of love, and to spare him humiliation and maintain his respect for himself.”⁵⁶ Kant here claims that there is a duty of love (to be benevolent toward the impoverished person), i.e., a duty to exercise our moral power of love by adopting the ends of the impoverished person and helping them from the motive of duty. However, the exercise of love (benevolence) humbles the beneficiary. That is, this exercise of love in isolation inherently injures the self-respect of the recipient human. Given this, the exercise of love toward a human inherently generates a corresponding duty of respect to be beneficent in a way that allows the recipient to maintain self-respect (i.e., a duty to exercise the moral power of respect).

If what I have said is correct, then the general structure of how duties of respect and duties of love are “essentially always connected together with one another in one duty according to the law” seems to be that, given certain features of human nature (that we are humbled when we depend on the benevolence of others), exercising one kind of moral power in isolation (and so being moved by the thought of duty to comply with only one kind of duty [love] in our interactions with other humans) has certain inherently morally pernicious effects on them (their losing self-respect). This generates a duty to exercise the other corresponding moral power (respect) when exercising the one in interactions between humans in order to counteract the morally pernicious effects of the exercise of a single moral force on the human recipient. In order for the exercise of the moral power of love to generate an interaction in which humans fully do morally right by each other, it must be tempered by respect so as to minimize the loss of self-respect of those whom we help.

56 Kant, *Doctrine of Virtue*, 6:448–49. A reviewer helpfully noted that, in this case, one might argue that it is simply more beneficent to spare another bad feelings when we aim to help and so respect is not necessary in this case. I agree that it can be more beneficent to spare another bad feelings when we aim to help. However, I stress that this point is compatible with a different point that Kant is making here. Kant’s concern here is clearly about how the moral law requires the joint exercise of respect and love between humans, for the passage deals with how the moral law unites duties of respect and love, not with how to be as beneficent as possible in exercising benevolence.

Though Kant himself only gives an example of how morality for humans requires that love be tempered by respect, we can give parallel examples of how it requires that respect be enriched by love. A complication arises given that Kant characterizes duties of respect merely negatively, as duties not to succumb to three kinds of vices: arrogance, defamation, and ridicule.⁵⁷ These are duties of respect *not to* interact with other humans (a) in a way that demands that others think little of themselves in comparison with us (arrogance), (b) in a way that brings into the open something “unfavorable to respect to others,” i.e., something that is liable to make others lose respect for someone (defamation), or (c) in a way that exposes the faults of others in order to maliciously take pleasure in these faults (ridicule). Such vicious conduct clearly violates the dignity of those to whom we are arrogant, those we defame, and those we ridicule. For all such conduct fails to treat others as equal sources of inviolable moral worth. Compliance with duties of respect (and so the exercise of the moral power of respect) is thus a constitutive aspect of humans’ fully treating one another as ends in themselves.

Given the negative characterization of these duties, any given exercise of the moral power of respect is an instance of conduct that we abstain from doing something in our moral interactions with others. Thus, the exercise of the moral power of respect is something that inherently maintains a distance between humans by having them not interact in a way that they otherwise would. This distance keeping is a plausible candidate for a morally pernicious effect inherent to the isolated exercise of respect among humans. After all, it seems humans can be moved by the thought of duty to comply with essentially negative duties of respect with regard to one another simply by staying clear of one another. Such pure distance keeping between humans, however, does not constitute fully treating another as an end in itself. As Kant puts it in the *Groundwork*, if the representation of humanity as an end in itself “is to have its *full* effect in me, the ends of a subject that is an end in itself must, as much as possible, also be *my* ends.”⁵⁸ Thus, it seems plausible that the exercise of respect generates a corresponding duty to exercise a moral power of love toward a person in order to counteract pernicious pure distance keeping between humans.⁵⁹ That is, just as we should be benevolent toward an impoverished person in a respectful

57 Kant, *Doctrine of Virtue*, 6:464–65.

58 *Groundwork of the Metaphysics of Morals*, 4:430.

59 Though I do not have the space to go into details here, it seems clear that the other duties of love (gratitude and sympathetic feeling) similarly enrich exercises of respect so that they constitute full treatment of other humans as ends in themselves (*Doctrine of Virtue*, 6:454–56). After all, complying with these duties from the motive of duty manifests a benevolent appreciation of a benefit rendered to us or of another’s joy or pain, both of which counteract respectful distance keeping by bringing humans morally closer in being

way that considers the effect of our helping them achieve their ends on their self-respect, we should not be arrogant toward others (not defame them, and not maliciously ridicule them) in a loving/benevolent way that considers the effect of our distance keeping on their self-love, i.e., their happiness.

Here, I focus on the duty of non-arrogance for reasons of space, but similar things can be said about the other duties of respect. In order for the exercise of respect to comply with the duty of non-arrogance to constitute our treating other humans as ends in themselves, we must not simply prevent ourselves from demanding that others think that we are better than them from the motive of duty. We must also do so while appreciating that our achievements and victories, even if motivated by good will, can affect their self-love, their happiness, i.e., while exercising love. Take the example of sharing the news that you got an attractive job with a colleague who is also looking for a job in a competitive academic job market. A clearly immoral way of sharing the news that violates the duty of non-arrogance is to flaunt getting the job, demanding that others recognize you got it because you are better than them. A better but still morally deficient way of interacting is simply by not telling your colleague and have them find out from someone else. This is an exercise of respect that complies with the duty of non-arrogance and inherently keeps your distance from your colleague in not sharing the news with them. You can enrich your exercise of (non-arrogant) respect by sharing the news in a way that is considerate and emphasizes how much of a crapshoot the job market is, making it clear that you think of them as peers that are also deserving of such a position. By enriching respect with love in this way, you can considerably share the news while making it clear that you are not demanding that others think less of themselves.

Note that I discuss these examples of the lived moral tension in human life (of respectful benevolence and considerate non-arrogance) because Kant himself brings up how the moral law unites these duties (hence moral powers) and to clarify the distinction between these two moral powers (and duties to other humans), as well as give a sense of how they are both in play in human moral life. However, we should take care to distinguish between (a) this tension (and uniting of respect and love) and (b) the balancing opposition between respect and love that makes moral interactions possible and is at issue in the moral-balancing argument. Here, I am not trying to get argumentative mileage out of conflicts between respect and love for the sake of giving a moral-balancing argument.

At this point, one might worry, following Baron, that even if there is a sense in which respect has to be enriched by love and love has to be tempered by

moved by the thought of duty to share in another's ends (through either gratitude or sympathy).

respect, respect does not depend on love in the way love depends on respect. Baron argues that respect (in the usual, intuitive sense) either does not need love at all or needs it in a highly attenuated sense, but that love (benevolence) is only benevolent if it is respectful.⁶⁰ I grant the point that the mutual dependence of respect and love is not fully symmetric, i.e., that there is a sense in which love (benevolence) more strongly depends on respect (insofar as benevolence is only benevolence if it is respectful). But this asymmetry is compatible with morality between humans requiring that respect be enriched by benevolence. Indeed, Baron herself ultimately admits that “love does need respect, lest love be paternalistic or otherwise invasive, and respect may need love to check its tendency to aloofness.”⁶¹ That is, despite the asymmetry in dependence she rightly argues for, Baron seems to admit that respect needs to be enriched by love. And this is all Kant needs for his purposes.

What Baron finds objectionable about Kant’s view that respect and love are opposing forces is that, as she sees it, Kant exaggerates “the degree to which keeping a distance from others is needed in order to respect others’ freedom and self-direction and to preserve self-respect.”⁶² Baron is right that humans do not need to keep a large distance from one another to respect each other’s self-direction and self-respect, but I do not think this is the point Kant is making in this passage. Rather, I think Kant is claiming that exercising the moral force of respect is a matter of not breaching humans’ inviolable moral boundaries. For Kant, exercising respect does not increase the degree of moral distance from others’ moral boundaries. It simply maintains the required minimal moral distance, i.e., it ensures that we do not cross fundamental moral boundaries. Given this, Kant is right to claim that for the thought of duty to motivate humans fully to treat one another as ends in themselves, (a) their love must be tempered by respect, and (b) their respect must be enriched by love. That is, the thought of duty must move humans (a) to keep themselves within their moral boundaries as they help each other achieve their ends, and (b) to treat each other’s morally permissible ends as ends as they keep their moral boundaries and integrity. Baron is right to find the idea that respect and love pull us in opposite directions problematic, but Kant’s general view of how repulsion and attraction are balanced is not that they pull in opposite directions.⁶³ Rather, it is that they jointly ground a stable configuration of beings

60 Baron, “Love and Respect in the Doctrine of Virtue,” 33.

61 Baron, “Love and Respect in the Doctrine of Virtue,” 42.

62 Baron, “Love and Respect in the Doctrine of Virtue,” 41.

63 Their pulling in opposite directions would require that the increase/decrease of one force causes a decrease/increase effect of the opposing force and vice versa. This is also not the case for physical repulsion and attraction. For example, the attraction of two bodies

in a way that explains their essential properties (of bodies filling space and of humans interacting morally).

3.4. *Balancing Opposition of Respect and Love in Relationships*

I suggest that we interpret Kant as holding that a constitutive requirement of moral interaction between humans is that it consists of a joint exercise of respect and love/benevolence. That is, in order for humans to interact morally with one another and form moral communities, they must be moved by the thought of duty to treat each other both (a) as self-determining and having dignity, i.e., “unconditional, incomparable worth” (such that they exercise the moral power of respect and comply with duties of respect toward each other), and (b) as needy and vulnerable (such that they exercise the moral power of love and comply with duties of love toward each other).⁶⁴ Interactions between humans that do not meet these two requirements fail to constitute moral interactions and communities.⁶⁵ We can illustrate this point by seeing how idealized purely respectful/loving cases of two-person relationships fail to constitute moral interactions between humans.⁶⁶

An example of an idealized purely respectful relationship wholly lacking in love would be one between two respectful egoists.⁶⁷ The thought of duty could move these unbalanced humans to comply with duties of respect and maintain their moral boundaries but not to comply with duties of love and be benevolent to each other. Given their egoistic psychology and lack of love, they would each (respectfully) pursue their own interests and would lack the power to be motivated by the thought of duty to overcome their selfish inclinations and adopt each other’s ends as their ends from the motive of duty. They may incidentally come to share ends (e.g., putting out the same large fire threatening their two homes). But in their interactions, they could not be motivated by the thought

increases as they come closer together, but this does not cause them to repel each other less. Similarly, the repulsion of a body increases as we compress it, but this does not cause the body or its parts to attract each other less.

64 Kant, *Groundwork of the Metaphysics of Morals*, 4:436.

65 Kant thinks of community and interaction as interchangeable (e.g., Kant, *Critique of Pure Reason*, B258).

66 This use of examples is different from the one above. Here the examples of purely respectful and loving relationships are assumed as cases for *reductio* arguments that establish that genuine moral interaction is not possible between humans lacking either love or respect. The strategy of using these two-person cases as a guide to the general case of moral dispersion and collapse is similar to how, as we have seen Smith point out, Kant uses two-body problems as a guide for thinking about the general case of physical total collapse and dispersion.

67 I thank Mark Timmons for the helpful suggestion to use two respectful egoists as a case here, given the isolating dynamics to which their psychology leads.

of duty to overcome their selfish inclinations and adopt the other's (morally permissible) ends and so treat each other fully as ends in themselves. As such, purely respectful egoists could not be moved by the thought of duty in a way that grounds moral interactions (and community) between them even if they lack the desire to do so. They would be two isolated moral units but not a moral community. These purely respectful humans lack the power to be motivated by the thought of duty to overcome selfish desires and do justice to the finite, sensible aspect of human nature, the aspect that requires that we rely on each other in the pursuit of our (morally permissible) happiness.

An example of an idealized purely loving relationship would be one between two purely selfless lovers. The thought of duty could move these unbalanced humans to comply with duties of love and be benevolent to each other but not to comply with duties of respect and treat each other (or themselves) as sources of inviolable moral worth.⁶⁸ Given their selfless psychology and lack of respect, they would each (benevolently) adopt and pursue the other's ends and would lack the power to be motivated by the thought of duty to overcome their selfless inclinations and keep at a moral distance, maintaining their moral boundaries.⁶⁹ In their interactions, they could not be moved by the thought of duty to overcome their selfless inclinations and never treat each other or themselves as mere means. As such, purely selfless lovers could not be motivated by the thought of duty in a way that grounds moral interactions (and community) even if they lack the desire to do so. They would form a loving community of selfless

68 Plausibly, if one lacks the power to respect another as an end in itself, then one also lacks the power to respect oneself. After all, this power consists of being motivated by the thought of duty not to violate the dignity common to all humans, whether it is exercised to respect oneself or another.

69 One might object that, given how Kant characterizes practical love, it constitutively includes respect and keeping of our moral distance. For, after all, exercising the moral power of love involves sharing in another's ends as *their* ends from the motive of duty. As a reviewer helpfully pointed out, this seems to include some keeping of distance between humans and respecting the moral worth of the beneficiary. Moreover, it is natural to suppose that, in adopting another's *morally permissible* ends (as opposed to ends in general) we already exercise respect and keeping our distance insofar as we refrain from adopting another's *morally impermissible* ends (i.e., those that fail to respect human dignity). However, we must distinguish between this more minimal sense of one's moral distance keeping involved in exercising practical love and the full-blooded sense of moral distance that is constitutive of exercising respect and that maintains the *moral boundaries and integrity of both individuals*. For this richer sense of keeping one's moral distance, more is required. In particular, what is required is not just that they not be benevolent in ways that the beneficiary conceives as part of their happiness or that wrong someone, but that each maintains their moral integrity as separate, morally impenetrable agents with distinct boundaries as they adopt the other's ends.

happiness seekers but one in which the individuals lose their moral integrity as they fail to maintain moral boundaries. These purely loving humans lack the power to be motivated by the thought of duty to overcome selfless sensible desires and do justice to the rational, self-determining aspect of human nature, the aspect that requires that each of us choose our own way of life.

Given humans' dual nature as rational yet sensible beings, moral interactions between them require that they jointly exercise both respect and love. Humans need to be moved by the thought of duty both to maintain their moral boundaries with each other and to share ends morally. Only if the thought of duty moves them in these dual ways do humans interact morally with one another. With this view of how moral interaction/community between humans constitutively requires both respect and love in hand, we can give the moral-balancing argument.

3.5. *The Moral-Balancing Argument*

This argument is an *a priori* argument that explains the possibility of human moral subjects (as embodied subject of moral duties) as grounded on the joint exercise of moral repulsion (respect) and attraction (love), as the physical-balancing argument explains the possibility of bodies (as objects of outer experience) as grounded on the joint exercise of physical repulsion (expansion) and attraction (gravitation).

First, a sub-argument that moves from the supposition of humans with (repulsive) respect to the need for (attractive) love to avoid total dispersion. A world of purely respectful humans (e.g., respectful egoists), would be a world where the thought of duty could move humans to comply with duties of respect and maintain their moral boundaries but not to comply with duties of love and be benevolent to each other. Given their egoistic psychology and lack of love, they would all (respectfully) pursue their own interest and would lack the power to be motivated by the thought of duty to overcome their selfish inclinations and adopt each other's (morally permissible) ends as their ends from the motive of duty. They may incidentally come to share ends but not from the motive of duty. Indeed, given their egoistic psychology, they would tend to steer clear of each other. For in their interactions they could not be motivated by the thought of duty to overcome their egoistic inclinations and come morally closer by treating each other fully as ends in themselves. As such, this would be a world where humans could not be moved by the thought of duty to have moral interactions between each other even if they lack the desire to do so. A world of such humans thus faces a problem of total moral dispersion. This moral dispersion consists in the fact that purely respectful humans lack the power to be moved by the thought of duty to come morally closer by treating

each other fully as ends in themselves even if they lack the sensible desire to do so. This yields a condition of moral nothingness for such unbalanced human beings, which is equivalent to the emptiness of space for unbalanced bodies. This moral nothingness plausibly consists in the fact that the thought of duty cannot motivate such humans to interact morally and form a moral world, which (given ought implies can) implies that they are not under obligation to interact morally/form a moral world.⁷⁰

Now, a second half that closely parallels the first in reverse. A world of purely loving humans (of, e.g., selfless lovers) would be a world where the thought of duty could move humans to comply with duties of love and be benevolent but not to comply with duties of respect and treat each as a source of inviolable moral worth. Given their selfless psychology and lack of respect, they would all (benevolently) help each other achieve their ends but would lack the power to be motivated by the thought of duty to overcome their selfless inclinations and treat each as a source of inviolable moral worth. In their interactions they could not be motivated by the thought of duty to overcome their selfless inclinations and keep their moral distance by never treating each other or themselves as mere means.⁷¹ As such, this would be a world where humans could not be motivated by the thought of duty in a way that grounds moral interactions (and community) between them even if they lack the desire to do so. A world of such humans thus faces a problem of total moral collapse. This moral collapse consists in the fact that purely loving humans lack the power to be motivated by the thought of duty to keep their moral distance by maintaining the separateness of persons even if they lack the desire to do so.⁷² This collapse leads to a

70 One might call this the *impossibility interpretation* of moral nothingness. I do not pretend to have given a convincing case for this interpretation of moral nothingness. Indeed, as Bianca Ancillotti and Houston Smit have pointed out to me, the fact that Kant characterizes moral nothingness as “immorality” rather than “amorality” suggests a different, *immorality interpretation* of moral nothingness. According to this interpretation, the nothingness consists of the permissibility of intuitively immoral relations between unbalanced humans (of disrespectful benevolence/callous respect) due to their only being able to be moved by the thought of duty to comply with one kind of duty to one another. For my purposes, it does not matter whether we adopt the impossibility or immorality interpretation of moral nothingness. On either interpretation, moral nothingness constitutes a condition where genuinely moral relations between humans are impossible.

71 In this more general case, the thought of duty cannot motivate purely loving humans to prevent themselves from treating themselves as mere means, but also from treating selfless, willing third parties as mere means for the sake of another.

72 Jacob Barrett helpfully pointed out to me that Rawls makes a similar point to Kant’s here about classical utilitarianism not respecting the separateness of persons in *A Theory of Justice*: “The most natural way . . . of arriving at utilitarianism . . . is to adopt for society as a whole the principle of rational choice for one man . . . , this [impartial] spectator who is

condition of moral nothingness, which plausibly consists of the fact that the thought of duty cannot motivate such humans to interact morally and form a moral world, which (given ought implies can) implies that they are not under obligation to do so.⁷³

Note that, as is to be expected of balancing arguments, moral interaction between humans is shown to be impossible in different ways in the two *reductio* sub-arguments. This is highlighted by the fact that these worlds fail the universalizability test of the categorical imperative in different ways.⁷⁴ A purely respectful world is one in which the maxim of nonbeneficence would be a universal law. Kant explicitly notes that such a world is one we can conceive of but cannot rationally will.⁷⁵ By contrast, a purely loving world is one in which the maxim of generally treating each other as ends in themselves but sometimes treating some humans (perhaps themselves) as mere means for the sake of others' ends would be a universal law. This latter maxim is like that of making a lying promise when in need: both maxims that hold that one can sometimes treat humans as mere means.⁷⁶ They are thus nonuniversalizable maxims that violate perfect duties to others and yield a contradiction in conception when universalized. Consequently, although moral interactions between humans are impossible in both a purely respectful and a purely loving world, these unbalanced worlds are not symmetrical. A purely loving world is actually impossible (as it yields a contradiction in conception) whereas a purely respectful world is merely morally

conceived as carrying out the required organization of the desires of all persons into one coherent system of this desire; it is by this construction that many persons are fused into one" (*A Theory of Justice*, 23–24). Rawls is explicitly referring to this as "a sense in which classical utilitarianism fails to take seriously the distinction between persons" (*A Theory of Justice*, 163; see also 24). Rawls notes, "This results in impersonality, in the conflation of all desires into one system of desire" and amounts to "mistak[ing] impersonality for impartiality" (*A Theory of Justice*, 164, 166).

- 73 As Sasha Newton and Marie Newhouse have pointed out to me, one related way of understanding this moral collapse is that, without respect, humans cannot be moved by the thought of duty to respect each other's rights. Elaborating the connection between rights and moral collapse would require interpreting Kant's "construction of the concept of right" by analogy with presenting the possibility of bodies moving freely under the law of the *equality of action and reaction* (*Doctrine of Right*, 6:233). This lies beyond the scope of this paper.
- 74 I thank Houston Smit for helping me see the need to connect the moral-balancing argument's sub-cases with the categorical imperative test.
- 75 This is made clear by, e.g., Kant's claim in the second section of the *Groundwork* that a universal law according to the maxim of nonbeneficence could subsist but cannot be rationally willed as a universal law of nature (*Groundwork of the Metaphysics of Morals*, 4:423).
- 76 Cf. Kant, *Groundwork of the Metaphysics of Morals*, 4:422; see also 4:429–30.

undesirable (as it yields a mere contradiction in the will). Still, both are worlds in which genuine moral interactions between humans are impossible.⁷⁷

By means of this argument, we can infer (as Kant did in the physical case) that without two forces of repulsion and attraction to balance each other, humans cannot possess the essential property of having genuine moral interactions. Humans endowed solely with moral repulsion face a problem of total moral dispersion, while humans endowed solely with moral attraction face a problem of total moral collapse. In either case, humans fail to possess the essential property of having genuine moral interactions through the exercise of their moral powers (i.e., through being motivated by the thought of duty). We can conclude that in order for humans to be genuine humans (subjects of morality who have moral interactions in being motivated by the thought of duty), they must jointly exercise respect and love (i.e., the thought of duty must move them to be respectful and benevolent to each other). Without such joint repulsion and attraction, humans cannot relate to each other as they essentially must if they are to constitute embodied subjects of a moral community, so there can be no moral world at all. And if there can be no moral interactions/communities, i.e., no morality *between* humans, there is a sense in which there is no morality at all.⁷⁸

The moral-balancing argument's conclusion—that the thought of duty needs to motivate humans to respect and love one another for moral interactions and so morality between humans to be possible—fits nicely with Kant's claim in section 12 of the introduction to the *Doctrine of Virtue* that certain forms of affective receptivity to concepts of duty “lie at the basis of morality, as *subjective* conditions of receptivity to the concept of duty, not as objective

77 This is not unlike the physical case, for a purely repulsive world is one where bodies fail to even have a specifiable physical location, whereas a purely attractive world is one where bodies occupy the same point in space (though they fail to fill it). Note that a purely repulsive *physical* world seems more incoherent than a purely attractive one, but a purely *attractive* moral world seems more incoherent than its counterpart. This shows that we should not expect this kind of analogy to grant too much symmetry to Kant's theoretical and practical philosophies. This is made clear by considering how the different sides of this analogy interact with other aspects of Kant's system: the moral worlds interact with the categorical imperative test while the physical worlds interact with infinity and the relationship between geometric and physical objects.

78 This is in keeping with Kant's discussion of the autonomy formula in the second section of the *Groundwork* as essentially dealing with a community of autonomous legislators. For here, he considers an autonomous will as a “universally legislating will” (*Groundwork of the Metaphysics of Morals*, 4:432), i.e., a will that legislates not just for a single “I” but for all of “us” fellow rational agents. This way of reading Kant also fits nicely with the fact that, as Jens Timmermann notes, Kant seems to regard the autonomy and realm of ends formulations as statements of the same third formula of the categorical imperative (Timmermann, *Kant's Groundwork of the Metaphysics of Morals*, 105).

conditions of morality.”⁷⁹ That is, Kant claims that it is a condition of the possibility of morality for certain subjects (humans) that the thought of duty be able to move them in certain ways. Kant thinks that every human by nature must be able to be moved by the thought of duty to feel moral feeling, consciousness, love of neighbor, and respect for oneself and that “it is by virtue of [these forms of receptivity to the thought of duty] that he can be put under obligation.”⁸⁰ I think that with the moral-physical analogy, Kant means to add that the capacity for joint exercise of respect and love is among these subjective conditions of receptiveness to the thought of duty. Indeed, in discussing moral feeling, Kant makes similar claims to the conclusion of the moral-balancing argument but about moral feeling rather than respect and love: that if a human were “completely lacking in receptivity to [moral feeling] he would be morally dead,” and that “if (to speak in medical terms) the moral vital force could no longer excite this feeling, then humanity would dissolve (by chemical laws, as it were) into mere animality and be mixed irretrievably with the mass of other natural beings.”⁸¹ That is, Kant claims that if the thought of duty could not move humans to feel moral feeling, then humans could not constitute genuine moral subjects and would be mere natural beings that are moved by desires according to natural laws (but not by the thought of duty, by representations of laws).⁸² Kant thus seems to think that for humans to be genuine moral subjects and so genuine humans (rational animals under obligations to interact morally that can be motivated by the thought of duty and not mere brute animals without such obligations that can only be motivated by desires), the thought of duty must be able to move them to feel moral feeling (and conscience, love, and self-respect) and to jointly exercise respect and benevolence.

Both the physical- and moral-balancing arguments constitute *a priori* arguments that link the possibility of certain kinds of individuals (bodies as objects of outer experience/humans as moral subjects) and the possibility of their existing in a system or community (bodies filling space/humans morally interacting). They both seek to show that the individuals in question are ultimately only possible if the system of individuals is possible (if bodies can interact so as to fill space/if humans can interact so as to form moral communities). This moral-balancing argument thus yields insight into the possibility of humans as moral subjects that have moral interactions as grounded in a

79 Kant, *Doctrine of Virtue*, 6:399. I thank an anonymous reviewer for pointing out to me the relevance of Kant’s discussion of these moral endowments to this project.

80 Kant, *Doctrine of Virtue*, 6:399.

81 Kant, *Doctrine of Virtue*, 6:400.

82 Cf. Kant, *Groundwork of the Metaphysics of Morals*, 4:412.

balancing opposition of moral repulsion and attraction. This is just as the physical-balancing argument yields insight into the possibility of bodies as movables in spaces that fill space as grounded in a balancing opposition of physical attraction and repulsion. Without this balancing opposition of repulsion and attraction, there would not be bodies as such objects of outer experience or humans as such moral subjects at all. I give a systematic account of the different aspects of the moral-physical analogy as I have fleshed it out here in the appendix, but now I turn to discuss some upshots and payoffs of my reading of this analogy.

3.6. *Uphots and Interpretive Payoffs*

Apart from making sense of an intriguing yet puzzling passage, my reading offers interesting interpretive payoffs. By drawing this analogy, we think of a moral world of humans as constituted by an essential balancing opposition of repulsion and attraction. By thinking of the moral world of humans in this way, we gain a formal grasp of this moral idea (i.e., a grasp of its formal character), which makes the idea determinate enough for reason to make purposive use of it. Since this idea is that of a world of humans whose wills are fully determined according to the moral law, symbolizing the moral world of humans through this analogy to the physical world allows reason to put this idea to purposive use in practically cognizing that we ought to bring about such a world by interacting with one another in a dynamical harmony of respect and love.⁸³ One purposive use of the idea of the moral world that this analogy enables is thus the practical use that Kant attributes to ideals in the “Ideal of Pure Reason” chapter of the “Transcendental Dialectic.”⁸⁴ Here, Kant notes that ideals have “*practical power* (as regulative principles) grounding the possibility of the perfection of certain *actions*.”⁸⁵ Ideals provide “an indispensable standard for reason, which needs the concept of that which is entirely complete in its kind, in order to assess and measure the degree and the defects of what is incomplete.”⁸⁶ For Kant, ideals play an important role in guiding our moral improvement by providing a standard by means of which we can evaluate our conduct. Kant tells us that the

83 Kant characterizes practical cognition in the first *Critique* as the cognition “through which I represent *what ought to exist*,” and the practical use of reason as that “through which it is cognized *a priori* what ought to happen” (*Critique of Pure Reason*, A633/B661; emphasis added).

84 Kant, *Critique of Pure Reason*, A568–71/B596–99. Note that the claim that the concept of the moral world is an ideal that fleshes out the realm of ends fits well with Kant’s calling the realm of ends a “splendid ideal” in the third section of the *Groundwork* (*Groundwork of the Metaphysics of Morals*, 4:462).

85 Kant, *Critique of Pure Reason*, A569/B597.

86 Kant, *Critique of Pure Reason*, A570/B598.

ideal of the Stoic sage guides our moral improvement as individual humans.⁸⁷ I suggest that the ideal of the moral world guides our moral improvement as a community, admonishing us to bring about a dynamical harmony of respect and love in our interactions with one another. The moral world is a realm of ends among humans, considered in thoroughgoing determination according to the moral law, and it allows us to assess and measure the degree and defects of what is incomplete in our empirically realized moral communities. According to this interpretation, the moral-physical analogy, in determining the idea of a moral world of humans in this way, constitutes a cognition of the moral requirements that follow from the application of the categorical imperative to humans. This yields cognition of a duty of humans as such to form a moral world with one another by jointly exercising respect and love in their interactions. The cognitive gain yielded by the moral-physical analogy therefore forms part and parcel of the articulation of the system of duties that constitutes the *Doctrine of Virtue*.

Seeing the moral world as a realm of ends among humans allows us to see further interpretive payoffs for Kant's normative ethics. For example, the dynamical harmony of love and respect that Kant thinks we humans ought to bring about in our interactions helps correct a misleading toy picture of Kant's ethics one might get by focusing solely on the *Groundwork*. Such a narrow focus might lead one to think that Kant's ethics is inherently cold and heartlessly concerned only with acting on universalizable maxims and doing our duty for the sake of duty. But the idea of a moral world of humans with its constitutive dynamical harmony of love and respect allows us to see that acting on morally permissible maxims for the sake of duty is by no means a cold, merely intellectual activity. It is a richly social and emotionally laden activity that seeks to do justice both to the rational and to the finite aspects of human nature. The moral-physical analogy (and the resulting dynamical harmony of love and respect) helps us see that Kant's ethics has a rich heart, making room for emotional and social aspects of our moral lives as humans.

Considering this moral world as an ideal that realizes a realm of ends and so a world in which all humans comply with the categorical imperative (and the moral law) can further inform our interpretation of Kant's moral philosophy if we use it to work backward toward more fundamental aspects of his moral philosophy. This is because Kant's claims about morality and rational agents in the *Groundwork* and the second *Critique* ultimately have to apply to humans who ought to bring about a dynamical harmony of love and respect. In other words, the idea of a moral world of humans spells out the material conditions for a

87 Kant, *Critique of Pure Reason*, A569/B597. The idea of the son of God in the *Religion* serves a similar purpose as a moral ideal or exemplar in a religious context, as Chignell discusses in "The Devil, the Virgin, the Envoy," esp. 120–22.

realm of ends of humans in particular.⁸⁸ I suggest then that the moral-physical analogy and its idea of the moral world can inform our interpretation of Kant's treatment of the categorical imperative and moral law more generally. It can do this by specifying that although these fundamental practical principles may seem (and indeed be) sparse, they need to be able to ground the kind of rich moral community we find in the idea of a moral world of humans.

Further interpretive payoffs that I do not have the space to fully develop here concern how the moral-physical analogy can help us understand why Kant makes certain intriguing but somewhat puzzling claims about the moral law and its contemplation in different writings. For example, he claims in the *Doctrine of Virtue* that we can cultivate virtue by contemplating the dignity of the moral law.⁸⁹ But this seems puzzling if we just consider the moral law as a practical principle.⁹⁰ What is it that we are contemplating when we contemplate this principle? Do we simply contemplate its constitutive concepts and the way they are combined? That does not seem to inspire one to be virtuous. I suggest that part of what we can contemplate in the moral law is a realm of ends of humans in a dynamical harmony of love and respect. This is supported by Kant's claims that the realm of ends is "intended to bring an idea of reason closer to intuition (by a certain analogy) and thereby to feeling."⁹¹ Kant thinks

88 I thank Houston Smit for helping me see the need to make this point explicit.

89 Kant, *Doctrine of Virtue*, 6:397.

90 To the extent that the moral world of humans is a world of virtuous humans, the moral-physical analogy can also help us understand why Kant claims in the *Religion* that virtue is "strictly ... beneficent in its consequences, more so than anything that nature or art might afford in the world" and that "if we consider the gracious consequences that virtue would spread throughout the world, should it gain entry everywhere, then morally oriented reason (through the imagination) calls sensibility into play" (*Religion within the Boundaries of Reason Alone*, 6:23n). Virtue is so beneficent because it has as a consequence a dynamical harmony of love and respect for all others, i.e., the pursuit of one's own self-determined happiness as one is helped by others and helps all others pursue theirs. And "should virtue gain entry everywhere," we would have an ideal moral world of humans living in a dynamical harmony of love and respect, of humans that all stay within their moral boundaries as they help each other achieve their own authentic autonomous happiness. In contemplating this ideal moral world, practical ("morally oriented") reason guides the imagination and enlivens sensibility, inspiring us to live up to our moral vocation and cultivate the virtue we ought to develop in ourselves.

91 Kant, *Groundwork of the Metaphysics of Morals*, 4:436. Timmons discusses how bringing the concept of the realm of ends closer to intuition and to feeling in part consists of conceiving of how one might realize a community whose end is the practice of virtue ("Closer to Intuition (According to a Certain Analogy) and Thereby to Feeling"). Timmons helpfully relates this to Kant's concept of a "visible church" whose constitution satisfies the characteristics of an ethical community and has for its content moral laws, arguing that this visible church is a way of making intuitive a communal duty to institute a religious

that, in bringing the requirements of the moral law closer to intuition, “the splendid ideal of a universal realm of *ends in themselves* (rational beings)” helps produce in us “a lively interest in the moral law.”⁹² Going a step further, the moral world of humans (as an ideal world of humans whose wills are fully determined by the moral law and its constitutive dynamical harmony of love and respect) helps bring the requirements of the moral law even closer to intuition and to feeling. By helping us see that this wonderful dynamical harmony of love and respect is what the moral law requires of humans, the moral-physical analogy and its resulting idea of a moral world inspire us to do our duty, become virtuous, and live up to our moral vocation so that we might contribute to the realization of such a harmonious community.

If what I am suggesting is correct, then we can contemplate the moral law by contemplating the dynamical moral order that it prescribes humanity to realize on earth (in the sensible world). This, in turn, can help enrich our understanding of why not just the “starry heavens above me” but also the “moral law within me” “fill the mind with ever new and increasing admiration and reverence the more often and more steadily one reflects on [it].”⁹³ That is, the moral-physical analogy helps us see that contemplating the moral law, just like the starry heavens, consists of contemplation of a universal dynamical order among beings wrought by the use of reason. The starry heavens constitute a universal dynamical order of bodies that is thought by understanding and theoretical reason. The moral law gives a universal dynamical order of rational beings and humans that is thought by practical reason.

If my interpretation of the moral-physical analogy is correct, this suggests that we can apply the same account of the use of symbols and analogies in my interpretation to other key analogies that Kant uses to give determinate content to ideas in his practical philosophy (in the form of a grasp of their formal character). These include the idea of the realm of ends (which is “possible only by analogy with a realm of nature”) that I mentioned above as well as those of right and legal relations of human actions (the former of which is presented “in pure intuition *a priori*, by analogy with presenting the possibility of bodies moving freely under the law of the *equality of action and reaction*” and the latter of which

community, i.e., a church whose realization would make humanity worthy of happiness. I find Timmons’s suggestion compelling and add that we should consider the visible church to be an even more determinate version of an ideal moral community between humans than the moral world, one that adds details about instituting a community with certain key ethical features and whose end is the practice of virtue.

92 Kant, *Groundwork of the Metaphysics of Morals*, 4:462, 4:463.

93 Kant, *Critique of Practical Reason*, 5:161.

is analogous to “the mechanical relations” of bodies’ “moving powers”).⁹⁴ I propose that these different analogies provide a framework for interpreting Kant’s philosophical project of a metaphysics of morals as a whole. Through these different layered and systematically related analogies, Kant offers a philosophical framework for thinking about the moral and political relations between humans as importantly analogous to the geometric and dynamical relations between bodies. This is an intriguing, majestic philosophical picture of the ethical relations between humans, one that we can only properly appreciate if we take seriously these analogies as more than mere metaphors and are willing to explore in detail how Kant richly relates theoretical and practical aspects of his philosophy through them. It is worth highlighting that nothing I have discussed here concerns how the moral-physical analogy illuminates the foundations of Kant’s moral philosophy directly. My aim here has been to give a reading of this analogy itself and spell out some interpretive payoffs for Kant’s ethics. Still, if my reading of the moral-physical analogy is correct, it suggests an interpretive approach that articulates how certain analogies relate to key aspects of Kant’s practical and theoretical philosophies.

4. CONCLUSION

In this essay, I have given the first systematic account of the analogy Kant draws in the *Doctrine of Virtue* between the physical world of bodies and the moral world of humans of which I am aware in the literature. In doing so, I have provided a philosophically attractive interpretation of Kant’s intriguing albeit obscure conclusion that the moral forces of respect and love are conditions for the possibility of a moral world of humans. I argued that we should interpret Kant as endorsing a moral-balancing argument and offered a reconstruction of this argument, explaining how it justifies Kant’s conclusion.⁹⁵ I also argued that this analogy allows the idea of a moral world to play the role of a compelling ideal for moral interactions between humans. I hope to have shown that by taking this analogy seriously as a key part of Kant’s systematic philosophy, we can better understand an intriguing yet puzzling and ultimately central aspect of his practical philosophy.

94 Kant, *Groundwork of the Metaphysics of Nature*, 4:438, *Doctrine of Right*, 6:232, and *Prolegomena to any Future Metaphysics*, 4:357n.

95 In giving the moral-balancing argument, I have assumed Kant’s normative premises concerning how a community of moral beings ought to be structured as well as metaphysical assumptions concerning the nature of forces. To the extent that one finds these assumptions implausible, one will find the justification for the claim lacking. I thank Robert Clewis and Janis Schaab for making me see the need to make this explicit.

In addition to helping us understand Kant's own philosophy, there is an insight in Kant's moral-balancing argument that we should take seriously in contemporary moral theorizing. Even if we do not accept all of Kant's philosophical system or of views about love and respect, there is something deeply right about his balancing argument, viz., that proper moral relations between humans require a balance of respect that allows each to grow as an individual, and love that allows all to grow closer. Kant's take on the balancing of moral forces and his corresponding argument show that, given our human constitution as rational but finite beings, moral interactions require that we balance respect and love in our moral relations with each other. Kant's moral-balancing argument thus sets forth certain *a priori* requirements for morally ideal interactions and relations: they must consist of a balancing of these two moral concerns. If a given interaction between humans fails to manifest either, then it fails to do justice either to the rational, self-determining aspect or to the sensible, finite aspect of human nature. If a relationship between humans is wholly lacking in either respect or love, then it is a morally pernicious relationship, one in which at least one human's dignity or happiness is compromised.⁹⁶ It is an intuitive and attractive thought that respect and love must be balanced in correct moral relations. Kant's moral-balancing argument explains why this is so. Without this balance, we cannot build the right kinds of relationships or communities, ones that allow each of us to grow and develop autonomously as we help contribute to the (morally permissible) happiness of all.⁹⁷

University of Arizona
 ssanchezborboa@email.arizona.edu

APPENDIX: SYSTEMATIC ACCOUNT OF THE MORAL-PHYSICAL ANALOGY

Space: Physical space of points——Moral space of ends.

Occupants/Bearers of Powers/Forces: Bodies——Humans.⁹⁸

⁹⁶ Robert Gooding-Williams helpfully pointed out to me that we can also connect Kant's idea of striking the right balance of love and respect in a contemporary context with the idea of setting and negotiating the right personal boundaries in one's different relationships.

⁹⁷ Throughout its development, this paper has greatly benefited from the helpful feedback of many thoughtful persons who have engaged with previous drafts at a wide variety of talks. For invaluable comments, objections, and discussion, I am especially grateful to Houston Smit, Mark Timmons, Ritwik Agrawal, and two anonymous reviewers.

⁹⁸ Note that, as Martin Sticker helpfully pointed out to me, this moral-physical analogy is compatible with there being important disanalogies between the bearers of moral and

Distance: Function of locations of bodies——Function of ends of humans.

R-rep(x_n), *Force of Repulsion, Keeps at Distance*: Expansion (grounds physical impenetrability)——Respect (grounds moral impenetrability).

R-att(x_n), *Force of Attraction, Closes Distance*: Gravitation (brings bodies physically closer)——Love (brings humans morally closer).⁹⁹

R-dis(x_n), *Total Dispersion*: All bodies disperse into infinity——The thought of duty cannot move humans to come morally closer by treating each other fully as ends in themselves and adopting each other's morally permissible ends even if they lack the sensible desire to do so.

R-col(x_n), *Total Collapse*: All bodies coalesce into a mathematical point——The thought of duty cannot move humans to keep their moral distance by treating each as an inviolable source of moral worth with dignity even if they lack the sensible desire to do so.

R-bal(x_n), *Balancing Opposition*: Bodies fill space determinately at every point with different density grounded in a dynamical harmony of expansion and gravitation——Humans interact morally with one another, helping each other pursue their own distinct happiness, forming a moral community grounded in a dynamical harmony of respect and love.

Nothingness: Bodies fail to possess the essential physical property of filling space——Humans fail to possess the essential moral property of having moral interactions.

physical forces. Indeed, one important disanalogy between these in the Critical period is that the primary bearer of physical forces is a material continuum, whereas the primary bearers of moral forces are discrete humans.

99 One apparent disanalogy between these two forces of attraction is that, although gravitation is symmetrically exercised between all bodies, it is not exercised to the same *degree* between them. In accordance with Newton's inverse-square law, the closer bodies are to each other, the more strongly they exercise their attractive forces. However, we can note that love can and ought to be exercised to different degrees between agents, for we ought to love practically those with whom we share special relations to a higher degree. Furthermore, exercising practical love involves fulfilling duties of love, which Kant notes, "put another under obligation" (*Doctrine of Virtue*, 6:450). Thus, those whom we love more ought to love us more, and we ought to love more those who love us more.

REFERENCES

- Allison, Henry Edward. *Kant's Groundwork for The Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press, 2011.
- Baron, Marcia. "Love and Respect in the Doctrine of Virtue." *Southern Journal of Philosophy* 36, no. S1 (Spring 1998): 29–44.
- Bielefeldt, Heiner. *Symbolic Representation in Kant's Practical Philosophy*. Cambridge: Cambridge University Press, 2003.
- Chignell, Andrew. "Are Supersensibles Really Possible? The Evidential Role of Symbols." In *Recht und Frieden in der Philosophie Kants*, edited by Valerio Rohden, Ricardo R. Terra, Guido Antonio de Almeida, and Margit Ruffing, 99–110. Berlin: Walter De Gruyter, 2008.
- . "Beauty as a Symbol of Natural Systematicity." *British Journal of Aesthetics* 46, no. 4 (October 2006): 406–15.
- . "The Devil, the Virgin, and the Envoy: Symbols of Moral Struggle in Religion II.2." In *Klassiker Auslegen: Religion innerhalb der Grenzen der blossen Vernunft*, edited by Otfried Höffe, 111–29. Berlin: Akademie Verlag, 2011.
- . "Real Repugnance and Belief about Things-in-Themselves: A Problem and Kant's Three Solutions." In *Kant's Moral Metaphysics: God, Freedom, and Immortality*, edited by Benjamin J. Bruxvoort Lipscomb and James Krueger, 177–209. Berlin: Walter De Gruyter, 2010.
- Edwards, Jeffrey. *Substance, Force, and the Possibility of Knowledge: On Kant's Philosophy of Material Nature*. Berkeley: University of California Press, 2000.
- Filippaki, Eleni. "Kant on Love, Respect, and Friendship." *Kant Yearbook* 4, no. 1 (2012): 23–48.
- Friedman, Michael. "Metaphysical Foundations of Natural Science." In *A Companion to Kant*, edited by Graham Bird, 236–48. Oxford: Blackwell, 2006.
- Johnson, Robert. "Love in Vain." *Southern Journal of Philosophy* 36, no. S1 (Spring 1998): 45–50.
- Kant, *Critique of the Power of Judgment*. Translated and edited by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press, 2002.
- . *Critique of Practical Reason*. In *Practical Philosophy*, 137–271.
- . *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1998.
- . *Doctrine of Right*. In *Practical Philosophy*, 363–506.
- . *Doctrine of Virtue*. In *Practical Philosophy*, 507–603.
- . *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary J. Gregor and Jens Timmermann. Cambridge: Cambridge University Press, 2011.
- . *Metaphysical Foundations of Natural Science*. Translated and edited by

- Michael Friedman. Cambridge: Cambridge University Press, 2004.
- . *Practical Philosophy*. Translated and edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1996.
- . *Prolegomena to Any Future Metaphysics*. Translated and edited by Gary Hatfield. Cambridge: Cambridge University Press, 2004.
- . *Religion within the Boundaries of Reason Alone*. Translated and edited by Allen Wood and George Di Giovanni. Cambridge: Cambridge University Press, 1998.
- Lewis, David. *On the Plurality of Worlds*. Oxford: Blackwell, 1986.
- Moore, G. E. “External and Internal Relations.” In *Proceedings of the Aristotelian Society* 20, no. 1 (June 2020): 40–62.
- Rawls, John. *A Theory of Justice*. Rev. ed. Cambridge, MA: Harvard University Press, 1999.
- Rueger, Alexander, and Şahan Evren. “The Role of Symbolic Presentation in Kant’s Theory of Taste.” *British Journal of Aesthetics* 45, no. 3 (July 2005): 229–47.
- Sanchez Borboa, Santiago. “Making the Supersensible Intuitive.” Unpublished manuscript, June 25, 2022.
- Smit, Albert Houston. “Intuition.” In *The Cambridge Kant Lexicon*, edited by Julian Wuerth, 256–58. Cambridge: Cambridge University Press, 2021.
- Smith, Sheldon R. “Does Kant Have a Pre-Newtonian Picture of Force in the Balance Argument? An Account of How the Balance Argument Works.” *Studies in History and Philosophy of Science, Part A* 44, no. 3 (2013): 470–80.
- Swanton, Christine. “Kant’s Impartial Virtues of Love.” In *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*, edited by Lawrence Jost and Julian Wuerth, 241–60. Cambridge: Cambridge University Press, 2011.
- . *Virtue Ethics: A Pluralistic View*. Oxford: Oxford University Press, 2003.
- Velleman, J. David. “Love as a Moral Emotion.” *Ethics* 109, no. 2 (January 1999): 338–74.
- Von Haller, Albrecht. “Imperfect Poem about Eternity [Unvollkommenes Gedicht über die Ewigkeit].” In *Versuch Schweizerischer Gedichte*, edited by Michael Holzinger, 92–95. Berlin: Berliner, 2013.
- Timmermann, Jens. *Kant’s Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge: Cambridge University Press, 2007.
- Timmons, Mark. “‘Closer to Intuition (According to a Certain Analogy) and Thereby to Feeling’: Making Kant’s Kingdom of Ends Intuitive.” Unpublished manuscript, August 10, 2020.
- Warren, Daniel. “Kant on Attractive and Repulsive Force: The Balancing Argument.” In *Discourse on a New Method: Reinventing the Marriage of History and Philosophy of Science*, edited by Mary Domski and Michael Dickson,

193–241. Chicago: Open Court Publishing, 2010.

Watkins, Eric. "Kant on Extension and Force: Critical Appropriations of Leibniz and Newton." In *Between Leibniz, Newton, and Kant: Philosophy and Science in the Eighteenth Century*, edited by Wolfgang Lefèvre, 111–27, Boston: Kluwer. 2001.

Watkins, Eric, and Marius Stan. "Kant's Philosophy of Science." *Stanford Encyclopedia of Philosophy* (Fall 2014). <http://plato.stanford.edu/archives/fall2014/entries/kant-science/>.

PROMISES, COMMITMENTS, AND THE NATURE OF OBLIGATION

Crescente Molina

UNDER a widespread understanding of the nature of moral obligation, one cannot be under an obligation to perform or omit an act and have a moral power to release oneself from one's obligation. According to this view, being under an obligation *necessarily* entails relinquishing one's sovereignty over the obligatory matter—that is, one's capacity to control one's own obligational world. This essay argues against such a view.

I shall argue that by making what I will call a *commitment*, a person assumes an obligation toward another on a given matter while preserving the power to release herself from her obligation. I contrast relationships of commitment with the obligations and claims constituted by promises. Like promissory relationships, relationships of commitment are relationships of directed obligation and claim: commitments, in our sense, are owed to another person. Yet, unlike promissory relationships, relationships of commitment are not what I shall call relationships of *personal authority* or *exclusive deontic control*. Commitments constitute morally obligatory acts over whose obligatoriness those obliged always retain control. They constitute obligations over which obligors preserve moral sovereignty. Thus understood, the philosophical literature has ignored the existence of commitments.¹ This is surprising given that, as I will contend, often very important normative arrangements of our lives are or entail commitments.

Here is my plan. Section 1 explores the structure of a commitment and contrasts it with that of a promise. Section 2 shows how commitments shed light on the functioning of important moral phenomena. I consider some cases of loyalty obligations in romantic partnerships and in voluntary associations, and show how distinguishing promises from commitments helps us understand the role and scope of the law of contracts. Finally, section 3 shows that the existence

1 Distinguish my notion of commitment from other interesting, yet different, ones available in the literature. See, e.g., Chang, "Commitment, Reasons, and the Will," 76–79; or Margaret Gilbert's notion of "joint commitment" in *Rights and Demands*, chs. 9–10. See also below notes 2, 7, and 9.

of commitments invites us to revise some widespread, traditional philosophical views about the nature of moral obligation.

1. THE STRUCTURE OF A COMMITMENT

It is common for philosophers to hold that validly made promises put promisors under an obligation to act as promised and grant promisees a claim to performance against their promisors.² Yet there is a tradition within the philosophy of promising that has given a distinctive interpretation to the idea that promises constitute directed obligations and claims. Under this view, the fact that promises constitute directed obligations and claims entails the creation of what I will call a relationship of *exclusive deontic control* or *personal authority* over the promised matter. This means not only that the promisee is exclusively entitled to demand performance or blame the promisee in case of breach but also that she is the only one who is exclusively entitled to release the promisor from his duty. The promisee's power over the promisor's duty may of course be regulated by different sorts of normative considerations. Perhaps, for example, the fact that performance has become unfairly burdensome to the promisor constitutes a reason for the promisee to release the promisor. Yet considerations of this sort can be seen as merely regulating the reasonableness or justifiability of the exercising of a power that exclusively belongs to the promisee. The promisee can use her power in a wise or unwise way, but she is always exclusively sovereign over the promised matter. Using a proprietorial metaphor, we may say that, according to this view of the nature of promises, the promisee *owns* the promised performance.

Alongside early modern and modern philosophers such as Grotius or Kant, leading contemporary writers endorse the idea that promises are a kind of authority-creating/sovereignty-restricting mechanism.³ Hart, for example, claims that

- 2 There are promises that are arguably not *deontically* significant at all, even if they may have other sorts of rational or normative effects (they may, e.g., affect only our epistemic reasons or put the promisor under a normative reason to act as promised yet not under an obligation to do so, etc.). Furthermore, according to some, there are promises whose deontic effects concern only the promisor's self (i.e., so-called self-promises). Here I will focus only on promises that have interpersonal bipolar deontic or obligational impact (i.e., on those that create or constitute obligations and claims directed to others). For some important contemporary work on self-promises, see, e.g., Hills, "Duties and Duties to the Self"; Rosati, "The Importance of Self-Promises"; Dannenberg, "Promising Ourselves, Promising Others"; and Fruh, "The Power to Promise Oneself"
- 3 See, e.g., Grotius, *The Rights of War and Peace*, 3:703–5; Hobbes, *Leviathan*, 91–93, 97; Pufendorf, *On the Duty of Man and Citizen according to Natural Law*, 69–70, 106; and Kant, *The Metaphysics of Morals*, 6:274.

by promising to do or not to do something, we voluntarily incur obligations and create or confer rights on those to whom we promise; we alter the existing moral independence of the parties' freedom of choice in relation to some action and create a new moral relationship between them, so that it becomes morally legitimate for the person to whom the promise is given to determine how the promisor shall act. The promisee has a temporary authority or sovereignty in relation to some specific matter over the other's will.⁴

More recently, and in the same vein, Owens argues that

promising is designed to serve our authority interests. To predict that I shall give you a lift home, or to express the present intention of so doing, is only to give you some information about what I shall do. In promising you a lift, I grant you the authority to require me to give you a lift: it is now up to you whether I must give you a lift home. . . . I may make this promise for all sorts of reasons deriving from your needs and interests (and my own). But once the promise is made, it is a matter for you whether its fulfilment is required of me—you can demand a lift home for any reason, good or bad.⁵

However, the characterization of promising as an authority-creating/sovereignty-restricting mechanism proposed by these writers should not lead us to a mistaken universal generalization—namely, that our powers to bind ourselves by fiat are *only* powers to form relationships of authority or exclusive deontic control, or that *all* our directed obligations and claims entail such relationships. I shall argue that our capacities to make commitments and to make promises belong to the same genus of normative powers. They are powers to bind ourselves by declaration or fiat. Yet they are different. We have powers to institute relationships of directed obligation that amount to relationships of personal authority or exclusive deontic control, and we may think of our power to promise as belonging to this kind. But we have a normative power that is different from promising in this sense, or so I will argue. I refer to our power to constitute relationships of directed obligation

4 Hart, "Are There Any Natural Rights?" 183–84.

5 Owens, "A Simple Theory of Promising," 71. For another influential recent account of promising along similar lines, see Shiffrin, "Promising, Intimate Relationships, and Conventionalism." On Shiffrin's view, by "promising to ϕ , the promisor transfers his or her right to act otherwise to the promisee. To not ϕ , then, is to act in a way the promisor has no right to do, and to ϕ is to act in a way the promisee has a right that she (the promisor) do" (517).

and claim that do *not* amount to relationships of personal authority as our *power to make commitments*.⁶

By making a commitment, a person assumes an obligation toward another on the committed matter. Unlike mere personal resolutions, which may under certain conditions give us reasons for acting as resolved, commitments in my sense institute relations of directed obligation.⁷ The obligor's obligation in a commitment is directed toward the commitment's recipient (the "moral creditor"), who in turn acquires a claim to performance. As long as the commitment relationship is in place, the moral creditor is *entitled* to the committed performance. The marks of relationships of *deontic entitlement*, as I will understand these relationships here, include at least: (1) the *fittingness of demand* (i.e., it is *pro tanto* fitting or appropriate for those who are deontically entitled to another's acting or refraining from acting in a certain way to demand it of their obligor); (2) the *fittingness of blame* (i.e., it is *pro tanto* fitting or appropriate for those who are deontically entitled to another's acting or refraining from acting in a certain way to blame their obligor in case of breach); and (3) the *exclusive right to forgive* (i.e., only those who are deontically entitled to another's acting or refraining from acting in a certain way have the right to forgive their obligor in case of breach). All the marks of deontic entitlement are present in relationships of commitment. What moral creditors of commitments lack, however, is personal authority over the committed matter. They do not have exclusive control over the obligor's obligation: both moral creditors *and* obligors in a commitment have an independent power to release the obligor.

Hence, in contrast to promissory relationships, we may aptly say that those obliged by commitments share *equal sovereignty* with their moral creditors over

6 The possibility of commitments is ignored by Hart who, at the end of the passage quoted above, maintains that "the promisee has a temporary authority or sovereignty in relation to some specific matter over the other's will *which we express by saying that the promisor is under an obligation to the promisee to do what he has promised*" ("Are There Any Natural Rights?" 183–84, emphasis added). As we will see below, the existence of commitments shows that not all directed obligations entail an authority relationship as Hart seems to assume.

7 Cf. Ruth Chang's notion of commitment, which perhaps in capturing part of the ordinary use of the term understands a commitment roughly as a unilateral act of the will by which agents give themselves normative reasons ("Commitment, Reasons, and the Will"). My notion of commitments is different, for the point of commitments in my sense is to constitute bipolar (interpersonal) deontic constraints and not simply to give us reasons for the committed act. Sam Shpall has offered a comprehensive theory of commitments that purports to account for different senses in which commitments can have rational or normative (and deontic) effects ("Moral and Rational Commitment"). In this regard, my notion of commitment is narrower and less theoretically ambitious, for again it refers exclusively to what I take to be a distinctive kind of bipolar, interpersonal deontic relationship.

the acts that their commitments render obligatory. Put in other terms, in relationships of exclusive deontic control or personal authority, such as promissory ones, there is only *one* key to the obligational lock, and it is held by the moral creditor (i.e., the promisee). By contrast, in relationships of commitment there are *two* keys to the obligational lock, and one is held by the moral creditor and the other by the obligor herself.

To be sure, as I understand them, commitments are *bilateral* in the same sense that some philosophers have argued promises are.⁸ The obligations that commitments constitute are a product of both the obligor's and the moral creditor's choices. The binding force of commitments is conditional on the moral creditor's *uptake* of the claim offered by the obligor, and therefore both parties involved in the formation of a commitment agree to form an obligational relationship that does not amount to a relationship of personal authority or exclusive deontic control over the obligatory matter. Thus, the difference between promises and commitments should not be understood to lie on an alleged unilaterality of commitments when contrasted with promises.⁹

In the next section, I will illustrate these points with some examples. I purport to show that important arrangements in our lives have the structure of commitments. Of course, demonstrating that commitments are part of our moral world may not be sufficient. One may hold that a complete defense of the claim that we have the power to make commitments must also show that there is value in our possessing such a power.¹⁰ My main aim here is to show that we have the power to make commitments and that we must distinguish this power from our power to make promises (i.e., from our power to form relationships of obligation that amount to relationships of exclusive deontic control). However, let me briefly elaborate on the kind of value that I believe our power to make commitments serves.

A life deprived of at least the possibility of owing and being owed special obligations lacks an important aspect of what makes it go well. Our obligations

8 See, e.g., Kant, *The Metaphysics of Morals*, 6:272–73; Owens, *Shaping the Normative Landscape*, 223–26.

9 Cf. Chang, for whom commitments (in her sense of the term) are fundamentally unilateral in nature (“Commitment, Reasons, and the Will”). Yet also cf. Shpall, for whom moral commitments are, like promises, bilateral in nature and yet (unlike my notion of commitments and like promises) unilaterally inescapable for the obligor (“Moral and Rational Commitment,” 162–63).

10 See, e.g., Raz, who maintains that “a person's act is an exercise of a normative power if it brings about or prevents a normative change because it is, all things considered, desirable that that person should be able to bring the change about or prevent it by performing that act. Those who can exercise a normative power have a normative power to do so” (*The Roots of Normativity*, 162–63). See also Raz, *Practical Reason and Norms*, 69–73.

and claims as members of families, or as friends, colleagues, fellow citizens, and the like, are often constitutive aspects of what makes these relationships valuable. They are in themselves elements of our well-being.¹¹ Yet, an autonomous life is not only one in which we engage in valuable projects, activities, and relationships but also one in which we have some control over which of these projects, activities, and relationships we pursue. Thus, if deontic entities belong to the list of goods that constitute a good life, in principle, an autonomous life is one in which we have some degree of control over the deontic aspects of our life. There is no reason to maintain, however, that our interests in being able to mold our deontic world are limited to our interests in forming relationships of authority.

One element within our interests in deontic control may be what Owens calls our “authority interest”—that is, an interest in forming relationships of personal authority with others.¹² Yet our deontic-control interests are not only authority interests. Besides an authority interest, I maintain that we have what I call an *allegiance interest*. The allegiance interest is precisely a human need for developing voluntary obligations toward others without thereby relinquishing our sovereignty over the obligatory matter. It is an interest in being able to form voluntary relationships of directed obligation over which we, as obligors, retain and not relinquish control. Whereas our authority interest is fostered by our power to make promises and by promissory obligations and claims, the allegiance interest is served by our power to make commitments and by the obligations and claims that commitments constitute. Let us now consider some examples of commitments.

2. COMMITMENTS: SOME APPLICATIONS

2.1. Loyalty Obligations

Some of the special obligations that constitute some of our personal relationships are not voluntary in nature—that is, they do not have an exercise of normative power as their source or ground. As your friend, I owe you special attentiveness, respect, and care. And I may owe you these special loyalties without ever choosing (or declaring a choice) to be bound to these things. Friends, though perhaps bad ones, may be under such loyalty obligations without even being aware of it. Friendship loyalty obligations of this sort are simply grounded in *our friendships*—that is, in the network of habits, shared

11 Different versions of this idea may be found in, e.g., Raz, “Liberating Duties,” 18–21; Schefler, “Relationships and Responsibilities”; and Owens, “The Value of Duty.”

12 See Owens, *Shaping the Normative Landscape*, 3–21.

expectations, dispositions, beliefs, and emotions that constitute our friendships. They thus need not have as their source an exercise of normative power by any of the involved parties.¹³

However, not all loyalty obligations are like this. Some are a product of the parties' normative powers. Consider the case of sexual fidelity obligations in romantic partnerships. Many of the loyalty obligations we acquire toward our romantic partners are not voluntary in nature. Like friendship obligations, they are grounded in our romantic relationships in themselves—that is, in the habits, expectations, emotions, beliefs, and dispositions we share with our romantic partners. Other loyalty obligations, however, may arise when agreed or voluntarily assumed by the involved parties. This is at times the case of sexual fidelity obligations. Perhaps one may under certain circumstances acquire a sexual fidelity obligation or claim toward one person without ever really choosing or agreeing to assume such an obligation or claim (e.g., by leading them to form sufficiently strong expectations regarding one's sexual intentions and by inducing them to rely on such intentions). Yet at least some romantic partners fix the existence, content, and scope of their sexual fidelity obligations by fiat or agreement—that is, by exercising their normative powers. This is where the central question that concerns us here arises: when they take a voluntary form, as they often do, what is the nature of our sexual fidelity obligations?

According to a common view, they are simply promises. Accordingly, under this view, sexual fidelity obligations are relationships of personal sexual authority that the involved parties have voluntarily created. Thus, when Paula voluntarily assumes a sexual fidelity obligation toward her partner John, she gives him an exclusive entitlement to determine whether she is under a duty to refrain from having sex with other people.

Assuming that voluntary arrangements regarding sexual fidelity are promises, philosopher Hallie Liberto arrives at the interesting, thought-provoking revisionist conclusion that we ought not to make this kind of promise in the first place and that, in case we do, promisees are under a duty to release their promisors.¹⁴ According to Liberto, it is morally objectionable that agents acquire genuine personal authority over other autonomous agents' sexual lives. In her view, sexual fidelity promises would be akin to an act of sexual servitude. The problem with this view, however, is that it simply assumes that *this* is what is going on in the case of sexual fidelity arrangements. One may or may not share with Liberto the intuition that some promises in the sexual

13 For a similar contrast between the ways in which we acquire promissory and friendship obligations, see Raz, *The Authority of Law*, 257.

14 See, e.g., Liberto, "The Problem with Sexual Promises," 403–13.

realm are morally problematic. However, we do not need to solve this issue to understand the nature of ordinary sexual fidelity arrangements because, in contrast to what Liberto assumes, these arrangements are not promises but commitments. Liberto's puzzle disappears once we characterize sexual fidelity arrangements as commitments.

Here is what I take to be a plausible theoretical reconstruction of the phenomenology of our voluntary sexual fidelity arrangements. By creating these arrangements, romantic partners voluntarily assume obligations regarding their sexual lives, usually obligations to refrain from having sex with other people. When undertaking these obligations, however, they also grant each other *equal, independent powers to release* themselves from their obligations. They assume sexual fidelity obligations toward each other while *preserving sovereignty* over their sexual lives. The fact that sexual fidelity arrangements are commitments and not promises explains why those under them can liberate themselves from their obligations by merely communicating their choice to their partners. To take our example from before: for the duration of Paula's sexual fidelity commitment—that is, before she or John releases her from the obligation or before any other canceling factor obtains (e.g., John's death)—Paula owes it to John to refrain from having sex with other people. John therefore has a claim that she perform her obligation, Paula's practical deliberation is presumptively constrained toward performance, and it is *pro tanto* appropriate for him to blame her and for her to feel guilty in case of breach. However, John has no authority over the committed matter. Paula remains sovereign over her sexual life and can at any time release herself from her duty by communicating her choice to John.

Some may wish to reject the idea that in sexual fidelity arrangements those obliged have a normative power to release themselves from their obligations. They may insist that sexual fidelity arrangements are just promises and that only the moral creditor therefore has the power to release the obligor. According to this view, what explains the obligor's merely apparent power to release herself is that the obligor's declaration that she no longer wishes to remain obliged gives the moral creditor strong reasons to release the obligor.¹⁵ Since the value of sexual fidelity presupposes that the parties involved are both willing to preserve their sexual fidelity arrangement, the fact that one of them no longer wishes to remain in it counts as an extremely powerful, almost nondefeasible reason for the moral creditor to release the obligor. As a result, the moral creditor's claim to performance would not actually terminate after the obligor's alleged self-release. He now only ought not to exercise this claim and must therefore release the obligor.

15 See, e.g., Kimel, "Personal Autonomy and Change of Mind in Promise and in Contract."

This view strikes me as misguided. When Paula communicates to John her decision to terminate her sexual fidelity obligation, she does not see herself as simply giving him powerful reasons for releasing her. She sees herself, in this moment, as terminating her sexual exclusivity obligation toward John. And she is correct on this point. Her capacity to release herself is not subordinated to John's reasonable exercise of his power to release her: it is in her own sovereignty to liberate herself from her commitment. It may be true that at least part of the value of the practice of sexual fidelity derives from the fact that it is voluntarily carried out by those involved. But this is not an argument against characterizing our sexual fidelity arrangements as commitments. On the contrary, it helps to explain why when choosing the deontic structure of our sexual fidelity arrangements, we use commitments and not promises. We use a tool by which we assume bonds toward each other on a certain matter while maintaining the power to release ourselves from such bonds precisely because we believe that the committed matter (i.e., sexual fidelity) merits *this* deontic structure and not the one of a relationship of personal authority.

Another way in which someone may insist on characterizing voluntary fidelity obligations as promissory in kind is by holding that fidelity promises are made under the condition that the parties remain in their relationship. Thus, under this view, Paula indeed grants personal authority to John over her sexual life by making a sexual fidelity promise, yet such promissory relationship lasts only while she remains in a relationship with John. If she breaks up with him, she is no longer bound by her promise. Yet is this view really proposing something different from maintaining that these obligations are commitments? I do not believe so.

Many different things are usually involved in the complex set of acts and declarations comprised in the process of breaking up or terminating a romantic relationship. For instance, our romantic terminations often entail our communicating to our partner a change in our feelings and dispositions toward them and our intention to change the way we behave toward them (e.g., our decision not to live together any longer or to stop being physically intimate with them, etc.). But what is it that terminates about the involved parties' relationship when we say that one of the parties (or both) have ended their relationship? It seems clear to me that the only direct, constitutive, or essential effects of such terminations are *deontic* in kind: romantic terminations cancel the claims and obligations we owe to each other *qua* romantic partners. By breaking up a romantic relationship, one of the involved parties, or both of them, terminates the claims and obligations that constitute their relationship by fiat; that is, they end them by communicating a choice to thereby terminate such obligations and claims. Romantic terminations are thus exercises of a deontic power, of a

power that each partner has to cancel the deontic arrangement that constitutes their relationship.

To be sure, romantic terminations often also have other consequences besides their direct deontic-cancelation effects. If Paula breaks up with John, her termination will likely cause a change in the dispositions, feelings, and expectations that John and herself have toward each other. Yet these nondeontic aspects of their relationship (i.e., the network of dispositions, beliefs, feelings, habits, and expectations attached to their relationship) may or may not terminate *ipso facto* by Paula's termination—that is, by the very act of Paula's communicating to John her choice to terminate their relationship. The only direct, essential effect of her termination is that it cancels the obligations and claims that Paula and John owe each other as romantic partners.¹⁶

We can now see why simply postulating that voluntary fidelity arrangements are promises made under the condition that the parties have not terminated their relationship does not really constitute an alternative to the view I have proposed. For these promises will precisely entail the promisor's power to liberate herself from her promissory obligation by simply communicating to the promisee her choice to liberate herself (i.e., her choice to terminate the relationship). A promise that binds the promisor only while she remains in a relationship does not constitute a relationship of personal authority *if* the promisor has the power to terminate such relationship by sheer fiat. Thus, if promises are authority-creating/sovereignty-restricting mechanisms, voluntary sexual fidelity arrangements cannot be promises. As I have argued, they must be commitments.¹⁷

My argument so far has been that by releasing herself, the obligor can terminate her commitment obligation. The obligor's unjustified release, however, may cause a serious setback to the moral creditor's well-being and thus *wrong* him, even if her release terminates the moral creditor's claim and her obligation. For instance, under certain circumstances romantic terminations may cause significant harm to the moral creditor (e.g., psychological, financial), and the

16 More precisely, terminations directly cancel only the obligations and claims over which the involved parties have control. It seems clear to me that at least some of our relational obligations and claims may not be terminated by sheer fiat. For instance, we do not tend to think we can directly terminate all of our friendship obligations or claims by simply communicating to our friend that we thereby end our friendship.

17 The idea that some romantic loyalty obligations cannot be promissory in nature has been nicely highlighted by Elizabeth Brake, who argues that marital vows cannot be promises, for that would lead us to hold that unilateral divorce is generally wrongful, which, according to Brake, is not an acceptable conclusion ("Is Divorce Promise-Breaking?"). Understanding marital vows as constituted by commitments and not promises would dispel Brake's worry.

possibility of causing such harm may count as a consideration that strongly militates against the obligor's self-release. If the obligor, without sufficient justification, releases herself from her commitment and thus fails to act on the undefeated considerations that count against her release, we may say that she in some sense "wrongs" the moral creditor. It may be appropriate for the moral creditor, as a victim of the obligor's injurious self-release, to resent and rebuke the obligor, and for the obligor to feel remorse and to offer the moral creditor explanations for her unjustified termination. However, even if it may constitute a wrong or wronging in *this* sense, the obligor's self-release always terminates the obligatory force of her commitment. After Paula has exercised her power to release herself from her sexual fidelity obligations toward John, John is not deontically entitled to Paula's performance. Paula's practical deliberation is no longer deontically constrained on the matter, John is not entitled to demand that she refrain from having sex with other people, and it is not fitting any longer for John to blame her or for her to feel guilty if she does. Of course, Paula may have powerful reasons to refrain from having sex with other people right after her release (e.g., out of respect for John's feelings). But her commitment no longer puts her under any obligation to refrain from doing so.¹⁸

Other voluntary loyalties are similar to sexual fidelity obligations; they are commitments and not promises. Consider the case of some of the loyalty obligations entailed by our memberships in voluntary associations. Some of these loyalty obligations are the product of an exercise of normative power by those involved. By joining a political party, for example, incoming members

18 Thus, the alternative sense of "wrong" or "wronging" mentioned above does not entail the violation of a deontic entitlement. In other words, the fact that the moral creditor has "standing to be wronged" (in the above-mentioned sense) by the obligor's release does not entail that he is deontically entitled to the obligor's refraining from exercising her power to release herself. Indeed, it presupposes that he is not deontically entitled to the obligor's release. I think the distinctive marks of this special form of wronging relationship include: (1) the victim's right to demand *explanations* (not apologies) for the "wrongful" behavior; (2) the fittingness of the victim's *complaints* (where the target or direction of such complaint is different from blame's, for under the account of deontic entitlement I have adopted here, blame is only an appropriate reaction to an unjustified breach of obligation—that is, to a violation of a deontic entitlement); and (3) the fittingness of *resentment* (where, again, the reason, target, or direction of resentment is the fact that the obligor did not give the victim's interests sufficient weight in her decision to release herself, and not the breach of an obligation. Again, when the target of the victim's resentment is the breach of an obligation owed to the victim, resentment constitutes, at least in part, blame). Nico Cornell ("Wrongs, Rights, and Third Parties") insightfully defends the idea that wrongings may in certain cases not be equivalent to rights violations. The idea of wronging without breach of obligation that I am putting forward here, however, is different from Cornell's in important regards. I will leave an exploration of such differences for another occasion.

often sign up for a package of loyalty obligations. By pledging such loyalty to the party, they assume, for instance, an obligation to avoid plotting with their political adversaries against the party's goals. As long as they remain in the party, their practical deliberation is constrained accordingly, and blame and guilt are appropriate reactions in cases of breach. Yet, members of voluntary associations like political parties typically have the power to liberate themselves from their loyalty obligations by fiat—that is, by simply communicating to the party their decision to cancel their membership.

As in the case of sexual fidelity obligations, by exercising one's power to release oneself from partisan loyalty obligations, one may under certain circumstances wrong one's moral creditors. For example, quitting the party in order to join the adversaries in the middle of a political battle may constitute a case of injurious self-release. The obligor's self-release may render apt her remorse, her explanations, and the moral creditors' resentment and rebuke toward the obligor. However, after her release, her comrades are no longer *entitled* to her partisan loyalty. Her practical deliberation is no longer deontically constrained toward the performance of partisan loyalty acts, and blame and guilt are no longer appropriate reactions to her failing to undertake such acts. To be clear, our memberships in voluntary associations frequently entail relationships of authority. The members of a party may owe the annual membership fees to the party, and only the party (i.e., its administrative board) can release them from this obligation. The point is that at least *some* of our voluntary associative obligations, particularly those that constitute bonds of associative loyalty, seem to be commitments and not simply relationships of personal authority such as promissory relationships.¹⁹

Let us consider one final example to illustrate the explanatory potential of commitments.

2.2. *Contract Law*

The law of contracts is the law of enforceable private agreements. Legal systems do not enforce all kinds of private agreements, however. For instance, it is a basic feature of the practice of modern contract law that loyalty obligations, such as sexual fidelity obligations or marital loyalty vows, are generally legally unenforceable. Even when their breach results in serious harm to the moral

19 Like in the case of sexual fidelity commitments, it is plausible to think that we give our voluntary associative loyalty obligations the structure of commitments and not of promises precisely because the value of our membership in this type of association is at least partially constituted by the fact that we can choose to remain in them or not. Thus, commitments and not promises are the kind of deontic arrangement that most appropriately fits the kind of value this type of association has for us.

creditor, he is not legally entitled to demand performance or damages.²⁰ Yet why are the obligations produced by these kinds of agreements not enforceable? Instead of drawing a principled distinction between the structure of these types of obligations and that of those that the law tends to enforce (e.g., those that arise from agreements made in the context of commercial exchange), some contract theorists maintain that loyalty obligations and other forms of personal promises are not enforceable because we would in one way or another undermine the value of the promised matter if we enforced them. Hence, in their view, these personal promises are *in principle enforceable*, but we ought not to enforce them because by forcefully extracting their performance, we undermine the value of performance itself. For example, since sexual fidelity's value depends on it being voluntarily practiced and therefore not coercively extracted, enforcing sexual fidelity promises would deprive sexual fidelity of its value.²¹

There is some truth to this explanation, for it is clear that legal enforcement may often do more harm than good, and this may account for some of our contractual enforcement practices. But I believe that there is a more basic aspect of the nature of this type of obligation that renders these obligations unenforceable.

20 True, the breach of such obligations may constitute one of the grounds for a divorce, which in turn may amount to some forms of compensation. Yet in such cases, besides pending property distributions and other financial adjustments between spouses, compensation is typically awarded to the partner who, as a result of the marriage, incurs a financial "opportunity cost" by relying on the *stability of marriage* (e.g., abandoning her career in order to become the main caretaker of the family). Compensation thus is not awarded for the harm caused in virtue of the breach of the loyalty obligations as such, but as a distributive measure to alleviate the often unfairly distributed burdens of marital arrangements. Accordingly, for example, sec. 25 A 2 of the English Matrimonial Causes Act 1973 provides that such compensation is awarded to repair the "undue hardship [caused by] the termination of [one of the parties'] financial dependence on the other party." For some recent cases that vindicate this rationale in English law, see, e.g., *Miller v. Miller*; *McFarlane v. McFarlane* [2006] UKHL 24; or *RC v. JC* [2020] EWHC 466 (Fam).

21 See, e.g., Smith, who maintains that

the law's refusal to order specific performance and to punish breach can be linked to its refusal to provide redress for wrongful harms like cheating on a partner, lying to a brother or failing to deliver a gift. . . . The reason the law stays out in such cases, I believe, is that there is little that law can do to redress such wrongs and, as importantly, because legal interference may make the valuable activities in issue (loving and familial relationships, gift relationships) less valuable. ("Performance, Punishment, and The Nature of Contractual Obligation," 363)

Similarly, see Kimel, who holds that a "legal prohibition on adultery (or, say, the general availability of a legal injunction against it) would only diminish the significance [of marriage], and hence value of marital faithfulness" ("The Choice of Paradigm for Theory of Contract," 245). See also Eisenberg, "The World of Contract and the World of Gift."

It is not, of course, that loyalty obligations or marital vows are less important in the eyes of the law than the other private agreements that, unlike loyalty bonds, the law declares legally enforceable (e.g., sales, leases, etc.). Our loyalty bonds may be of crucial importance to us, and their breach may often undermine our well-being to a greater extent than the breach of many of our commercial arrangements. I believe that the reason loyalty bonds are not legally enforceable may have to do with the *kind of normative relationship* they entail. Here is a possible answer: the law is concerned only with the enforcement of entitlements that constitute domains of our authority. Only claims and obligations whose violation *also* constitutes the violation of someone's personal authority merit a coercive legal response. Since the moral creditor of a commitment does not have authority over the committed matter, her claim to performance is not sufficient to justify a legal remedy. Again, for the duration that the obligor's commitment stands, her breach may render her blameworthy and justify the moral creditor's rebuke and complaint. However, as we saw, the obligor always retains her sovereignty over the committed matter, and thus the breach of her commitment never entails a violation of the moral creditor's authority.

For some writers, the contention that promises are acts that create authority or restrict sovereignty comes in the form of the claim that promises create a *right* to performance.²² In their view, the possession of a right entails what one may call a sphere of *exclusive sovereignty*: a matter over which nobody but the right-holder may have the power to determine what is owed to her. Rights are thus domains of independence from external—both factual and deontic—control.²³ At least under this sense of what it is to have a right, we may aptly say that commitments, unlike promises, do not create a *right* to the committed matter even if they create obligations and claims to performance.

Contract law seems to be concerned with the enforcement of agreements that create rights (i.e., relationships of authority), and not with the enforcement of those that merely create obligations. There is surely more to be said on why only relationships of right and not merely relationships of obligation in principle justify legal enforcement, and there are also many other justificatory problems that a complete theory of contractual enforcement must address. But distinguishing promises from commitments gives us a clue as to what the basic role of legal remedies is. The law appears to be concerned not with correcting wrongs or breaches of obligations as such, or with the reparation of the losses that arise from these acts. For breaches of commitments may often be wrongful and harmful and still fall outside the contractual realm. In modern legal systems,

22 See those mentioned *supra* note 3.

23 See Nagel, who nicely defends this claim ("Personal Rights and Public Space," 94–95).

at least to a significant extent, contractual remedies, and perhaps the private law of remedies more generally, seem to be a reaction to harmful violations of *our rights*. And—unlike the breach of a promise—the breach of commitments (such as sexual loyalty obligations and other forms of personal loyalties), even if often wrongful and harmful, does not as such constitute a violation of the moral creditor's rights.

3. COMMITMENTS AND THE NATURE OF OBLIGATION

The reader may agree with the claim defended in the previous section that promissory relationships do not accurately capture the normative structure of our voluntary sexual fidelity arrangements or of other loyalty obligations. However, she may still have trouble endorsing the existence of commitments as a possible explanation. One aspect of the notion of commitment in particular might puzzle her. Commitments, as I have contended, constitute obligations over which the obligor retains control: in a commitment, the obligor can always *opt out* of her obligation by fiat—that is, by communicating to the moral creditor her choice to release herself. But are commitments really obligatory if that is so? For, one may hold, it seems to be an essential truth about the idea of an obligation that being under an obligation to ϕ necessarily entails that it cannot ultimately be up to the obligor's own choice to determine whether she is under an obligation to ϕ . Many philosophers endorse this understanding of the nature of obligation. Hobbes, for instance, maintains that “*promises* therefore . . . are Tokens of the Will; that is, . . . of the *last act* of deliberating, whereby the liberty of non-performance is *abolisht*, and by consequence are *obligatory*; for where Liberty ceaseth, there beginneth Obligation.”²⁴ And, Williams, in his famous critique of moral obligation, contends that

Moral obligation is inescapable. I may acquire an obligation voluntarily, as when I make a promise: in that case, indeed, it is usually said that it has to be voluntarily made to be a promise at all, though there is a gray area here, as with promises made under constraint. In other cases, I may

24 Hobbes, *Philosophical Rudiments Concerning Government and Society*, II. 10. See also Hart, who similarly maintains that

the force of the preposition “to” in the expression “having a duty to Y” or “being under an obligation to Y” (where “Y” is the name of a person) . . . indicates the person to whom the person morally bound is bound. This is an intelligible development of the figure of a bond (*vinculum juris: obligare*); the precise figure is not that of two persons bound by a chain, but of one person bound, the other end of the chain lying in the hands of another to use if he chooses. (“Are There Any Natural Rights?” 181)

be under an obligation through no choice of mine. But, either way, once I am under the obligation, there is no escaping it, and the fact that a given agent would prefer not to be in this system or bound by its rules will not excuse him; nor will blaming him be based on a misunderstanding.²⁵

Though their reasons for making these claims are of course different, both Hobbes and Williams seem to endorse the view that obligations always have an *authority-creating/sovereignty-restricting force*. Being under an obligation necessarily entails lacking sovereignty over the obligatory matter—that is, lacking the moral power to control our own deontic world on such matter. Hence, for these authors, commitments are a normative impossibility: our obligations always obtain at the expense of our sovereignty. If commitments do not, they cannot truly render acts obligatory.

But should we agree with this picture of the nature of obligation? An answer to this question should not be given by exclusive reference to an *a priori* concept of obligation where obligations are by definition authority-creating/sovereignty-restricting mechanisms as doing so would simply be begging the question. Instead, a sound response must derive from our exploration of the roles that obligations play in our normative world—that is, the ways in which obligations are connected to human interests or values. In other words, the form or structure that our obligations take must track their function or role.

It is certainly true that some obligations perform their role only if the obligor lacks control over them. Your interest in receiving truthful information and in being able to rely on it is served by my having an obligation not to lie to you. And your interest in preserving your bodily integrity is served by my having an obligation to avoid injuring you. The whole point of these obligations (i.e., to protect your interests in truthful information and bodily integrity by deontically constraining my agency toward the avoidance of acts that undermine these interests, such as lying to you or injuring you) would obviously be defeated if *I* could release myself from these obligations by fiat. Similarly, as we saw, the purpose of the obligations that arise from voluntary relationships of authority, such as promissory ones, will not be fulfilled if the obligor has a power to release herself from her duty. I have suggested that the point of these obligations is to serve our authority interest—that is, an interest in forming relationships of authority or exclusive deontic control.

But other obligations are different. As I have argued, some obligations fulfil their role in our normative world only if we, as obligors, have the power to release ourselves from them. The kind of value that is served by some of these obligations (e.g., sexual fidelity obligations and some of our voluntary

25 Williams, *Ethics and the Limits of Philosophy*, 196–97.

associative loyalties) is realized only if we can preserve our sovereignty over the matters that such obligations render obligatory. These obligations serve what I have called our allegiance interest.²⁶ Hence, while the authority-creating/sovereignty-restricting conception of the structure of an obligation nicely captures the role of *some* of our obligations, it simply falls short in matching the structure of at least one other class of obligations—namely, commitments.²⁷ There is, of course, a sense in which commitments limit the freedom of the obligor over the committed matter. As we saw, the obligor in a commitment remains obliged until she *communicates* to the moral creditor her intention to thereby liberate herself from her obligation. Therefore, a mere change of mind or uncommunicated desire not to be obliged does not succeed in liberating her from her commitment. Yet besides this minimal constraint, the obligor in a commitment remains sovereign over the committed matter: she always has the power to liberate herself by fiat.

Our puzzled reader may, at this stage, agree with the proposal that there are some claims that do not constitute entitlements of personal authority. Commitments are part of our moral world, she may grant. However, she may again insist on holding that those subjected to commitments are not under relationships of directed obligation. She may contend that directed obligations necessarily entail both relationships of what I called deontic entitlement *and* of personal authority. Thus, even if commitments in *some sense* bind us, their binding force must be different from the force of obligation. I am afraid, however, that at this point the disagreement with our reader may be purely terminological.

I think that our ordinary notion of obligation is broad enough for it to include deontic requirements that constitute relationships of personal authority (e.g., promissory obligations) and those that do not (e.g., commitments). And, if this is correct, it is our skeptical reader who would have to explain to

26 To be sure, commitments might well serve other human interests too. Yet, as I propose we understand them here, it is an essential feature of (validly made) commitments that they serve our allegiance interest. Naturally, what they cannot do is at the same time serve our allegiance *and* our authority interests. Our allegiance and authority interests are two mutually exclusive possible normative grounds of our obligations.

27 Although, as I said, commitments are inherently bipolar and interpersonal in nature and thus categorically different from so-called self-promises, the possibility of commitments should offer some support to those arguing for the existence of genuinely binding self-promises, over which the promisor arguably retains some control to liberate herself from her promise. If commitments are genuinely binding, the obligations that self-promises arguably create will not be the only kind of obligation over which the obligor retains control. For defenses of the possibility and coherence of self-release in the philosophical literature on self-promises, see Rosati, "The Importance of Self-Promises," 127–36; and, more recently, Oakley, "How to Release Oneself from an Obligation"; and Schaab, "On the Supposed Incoherence of Obligations to Oneself."

us why we should endorse exclusively an authority-creating/sovereignty-restricting conception of obligation and not a more ecumenical one. My central point here, however, is not terminological but theoretical. I have argued that there is a class of deontic requirements that, while entailing relationships of deontic entitlement, do not amount to relationships of personal authority or exclusive deontic control. Whether this type of deontic requirement deserves to be called an obligation or something different may be an important question, yet it does not undermine the substantive claim that such deontic requirements are part of our moral world.²⁸

4. CONCLUSION

Obligations are often constitutive aspects of our well-being. And, as autonomous beings, we have powers to shape some of our obligations at will. It is not the case, however, that our powers to shape our deontic world are reducible to our capacity to form relationships of personal authority. As I hope to have shown, we have the power to assume voluntary obligations over which we as obligors preserve control. These voluntary obligations are commitments.

Further exploration of the forms in which we may shape our deontic world could shed light on voluntary deontic arrangements that differ in structure from both promises and commitments. For example, it may be the case that some of our obligations are such that they obtain only if we voluntarily assume them, but that once we have assumed them, neither the obligor *nor* the moral creditor has control over them. Perhaps normative powers such as citizenship by naturalization may produce deontic relationships of such structure. Once naturalized, a person arguably assumes certain duties and claims toward the state and toward one's fellow citizens that neither oneself *nor* the state or one's fellow citizens have the power to terminate by fiat. Perhaps we have an interest in being able to

28 There are at least two kinds of questions about terminology one may ask regarding the relationship between commitments and obligations. First, one can ask whether the idea of commitment captures a sufficiently relevant part of the phenomenology of the notion of obligation in a way that is sufficient to justify (from the point of view of ordinary thought and talk) identifying commitments as obligations. Second, one may ask whether it would be *good* for us to adopt a conception or definition of obligation that incorporates commitments, regardless of what ordinary language and thought dictate. The first is a merely semantic question that, though perhaps interesting in itself, I do not consider to be of much philosophical interest. The second is an interesting (and difficult) question about terminological or definitional normativity, yet I do not wish to offer an answer to it here. Again, my primary task has been to identify a novel theoretical kind (i.e., commitments) and show its capacity to account for important aspects of our moral phenomenology, rather than to defend a claim about the correct use of the term "obligation."

voluntarily assume such *enduring bonds*, as we might call them. I cannot explore this type of deontic arrangement here. For our purposes, it is sufficient to note that the identification of commitments shows us that our basic powers to craft our deontic world go beyond our power to restrict our sovereignty by granting personal authority or exclusive deontic control to others. By identifying the existence of commitments, we thus open up inquiries into all the other forms our normative powers and our obligations and claims may take.²⁹

Harvard Law School
cmolinarodriguez@law.harvard.edu

REFERENCES

- Brake, Elizabeth. "Is Divorce Promise-Breaking?" *Ethical Theory and Moral Practice* 14, no. 1 (February 2011): 23–39.
- Chang, Ruth. "Commitment, Reasons, and the Will." In *Oxford Studies in Metaethics*, vol. 8, edited by Russ Shafer-Landau, 74–113. Oxford: Oxford University Press, 2013.
- Cornell, Nicolas. "Wrongs, Rights, and Third Parties." *Philosophy and Public Affairs* 43, no. 2 (Spring 2015): 109–43.
- Dannenberg, Jorah. "Promising Ourselves, Promising Others." *Journal of Ethics* 19, no. 2 (June 2015): 159–83.
- Eisenberg, Melvin. "The World of Contract and the World of Gift." *California Law Review* 85, no. 4 (July 1997): 821–66.
- Fruh, Kyle. "The Power to Promise Oneself." *Southern Journal of Philosophy* 52, no. 1 (March 2014): 61–85.
- Gilbert, Margaret. *Rights and Demands: A Foundational Enquiry*. Oxford: Oxford University Press, 2018.
- Grotius, Hugo. *The Rights of War and Peace*. Edited by Richard Tuck. 3 vols. Indianapolis, IN: Liberty Fund, 2005.
- Hart, H. L. A. "Are There Any Natural Rights?" *Philosophical Review* 64, no. 2 (April 1955): 175–91.
- Hills, Alison. "Duties and Duties to the Self." *American Philosophical Quarterly* 40, no. 2 (April 2003): 131–42.

29 I am grateful for helpful comments from David Owens, Sandy Steel, David Enoch, Ruth Chang, Maximilian Kiener, Felix Koch, Jed Lewinsohn, Cathy Mason, Luca Passi, Massimo Renzo, Felipe Jimenez, Elad Uzan, Laura Valentini, and the participants and organizers of the Oxford Jurisprudence Seminar. I am particularly grateful to the late Joseph Raz; this essay is dedicated to him.

- Hobbes, Thomas. *Leviathan*. Edited by Richard Tuck. Cambridge: Cambridge University Press, 1996.
- . *Philosophical Rudiments Concerning Government and Society*. London: R. Royston, 1651.
- Kant, Immanuel. *The Metaphysics of Morals*. Edited by Mary J. Gregor. Cambridge: Cambridge University Press, 2015.
- Kimel, Dori. “The Choice of Paradigm for Theory of Contract: Reflections on the Relational Model.” *Oxford Journal of Legal Studies* 27, no. 2 (Summer 2007): 233–55.
- . “Personal Autonomy and Change of Mind in Promise and in Contract.” In *Philosophical Foundations of Contract Law*, edited by Gregory Klass, George Letsas, and Prince Saprai, 109–12. Oxford: Oxford University Press, 2014.
- Liberto, Hallie. “The Problem with Sexual Promises.” *Ethics* 127, no. 2 (January 2017): 383–414.
- Nagel, Thomas. “Personal Rights and Public Space.” *Philosophy and Public Affairs* 24, no. 2 (Spring 1995): 83–107.
- Oakley, Tim. “How to Release Oneself from an Obligation: Good News for Duties to Oneself.” *Australasian Journal of Philosophy* 95, no. 1 (2017): 70–80.
- Owens, David. *Shaping the Normative Landscape*. Oxford: Oxford University Press, 2012.
- . “A Simple Theory of Promising.” *Philosophical Review* 115, no. 1 (January 2006): 51–77.
- Owens, David. “The Value of Duty.” *Aristotelian Society Supplementary Volume* 86, no. 1 (June 2012): 199–215.
- Pufendorf, Samuel. *On the Duty of Man and Citizen according to Natural Law*. Edited by James Tully. Translated by Michael Silverthorne. Cambridge: Cambridge University Press, 1991.
- Raz, Joseph. *The Authority of Law: Essays on Law and Morality*. 2nd ed. Oxford: Oxford University Press, 2009.
- . “Liberating Duties.” *Law and Philosophy* 8, no. 1 (April 1989): 3–21.
- . *Practical Reason and Norms*. Oxford: Oxford University Press, 2002.
- . *The Roots of Normativity*. Edited by Ulrike Heuer. Oxford: Oxford University Press, 2022.
- Rosati, Connie S. “The Importance of Self-Promises.” In *Promises and Agreements: Philosophical Essays*, edited by Hanoeh Scheinman, 124–55. Oxford: Oxford University Press, 2011.
- Schaab, Janis David. “On the Supposed Incoherence of Obligations to Oneself.” *Australasian Journal of Philosophy* 99, no. 1 (2021) 175–89.
- Scheffler, Samuel. “Relationships and Responsibilities.” *Philosophy and Public*

Affairs 26, no. 3 (Summer 1997): 189–209.

Shiffrin, Seana Valentine. “Promising, Intimate Relationships, and Convention-
alism.” *Philosophical Review* 117, no. 4 (October 2008): 481–524.

Shpall, Sam. “Moral and Rational Commitment.” *Philosophy and Phenomeno-
logical Research* 88, no. 1 (January 2014): 146–72.

Smith, Stephen A. “Performance, Punishment and The Nature of Contractual
Obligation.” *Modern Law Review* 60, no. 3 (May 1997): 360–77.

Williams, Bernard. *Ethics and the Limits of Philosophy*. London: Routledge
Classics, 2011.

WHAT IS THE POINT OF NONDOMINATION?

Devon Cass

THE REPUBLICAN TRADITION of political thought makes the following distinctive claims: (1) goodwill and virtuous self-restraint are insufficient to realize freedom; and (2) suitable law is constitutive of freedom.¹ I will refer to these as the *insufficiency* and *constitution* claims. The first of these has received the most attention and is often used to motivate the republican view. A slave with a kindly master is unfree, no matter how unlikely their master's interference. Thus, republicans argue, the absence or low probability of others' interference is not enough for freedom. Instead, noninterference must be enjoyed *robustly*: across a range of possible worlds, with no agent having an uncontrolled capacity to interfere with another. To be free in this sense is to enjoy *freedom as nondomination*.

Against the republican conception of freedom as nondomination, two objections have been persistently raised. The first objection is that it is problematically *moralized*.² The second objection is that it is *impossible* to realize.³ These

- 1 In the contemporary literature, these two claims are developed and defended by “neo-republican” theorists such as Philip Pettit, Frank Lovett, and Quentin Skinner. In addition, these claims are defended (on different grounds) by Kantians such as Anna Stilz, Arthur Ripstein, Japa Pallikkathayil, and Thomas Sinclair. See, for example, Pettit, *On the People's Terms*; Lovett, *A General Theory of Domination and Justice*; Skinner, “Freedom as the Absence of Arbitrary Power”; Stilz, *Liberal Loyalty*; Ripstein, *Force and Freedom*; Pallikkathayil, “Deriving Morality from Politics”; and Sinclair, “The Power of Public Positions.” In the canon, these claims have been defended by both “Italian Atlantic” and “Franco German” republican traditions. See Pettit, “Two Republican Traditions.” In this paper, I focus on the contemporary “neo-republican account,” and from here on I will use the term “republican” to refer to this camp.
- 2 See Carter, “A Critique of Freedom as Non-Domination”; Dowding, “Republican Freedom, Rights, and the Coalition Problem”; Kramer, “Liberty and Domination”; List and Valentini, “Freedom as Independence”; McMahon, “The Indeterminacy of Republican Policy”; and Waldron, “Pettit's Molecule.”
- 3 See Simpson, “The Impossibility of Republican Freedom”; Kolodny, “Being under the Power of Others”; Sharon, “Domination and the Rule of Law”; and Dowding, “Republican Freedom, Rights, and the Coalition Problem.” Apart from the *moralization* and *impossibility* objections, several theorists argue that appeal to freedom as nondomination

concerns have led to skepticism about whether republicanism's distinctive claims about the insufficiency of goodwill and constitutive importance of law can be adequately defended.

I believe these concerns with moralization and impossibility are indeed forceful. But at the same time, I think critics have risked obscuring insights of the republican approach, insofar as they take the insufficiency and constitution claims to be misguided entirely. My aim in this paper, then, is to articulate and defend a pair of different claims in the same spirit. I do so by examining the point or purpose of protecting persons from domination. On the standard view, nondomination is regarded as a conception of negative liberty, in competition with others in the literature such as the liberal “noninterference” view. On the account I develop, however, a different concern is at stake: that persons have a civic status as equals, which realizes an important form of social or “relational” equality.⁴ Protection from nondomination—more precisely, the provision of robust noninterference through law—is a central part of what this status involves. Understanding the point of nondomination in this way, I will argue, offers a compelling defense of alternative insufficiency and impossibility claims: goodwill is insufficient for, and suitable law is constitutive of, equality of civic status.

Niko Kolodny has also suggested that the republican concern with domination is better understood as a concern with social equality rather than freedom.⁵ But along with this suggestion, he seems to hold that the distinctive republican commitments are altogether indefensible. This I believe to be mistaken, and, as we will see later, leaves him with a problematic conception of social equality. Thus, I want to show that a satisfactory understanding of social equality—spelled out in terms of equality of civic status—draws on insights of the republican approach and regards goodwill as insufficient for, and suitable law constitutive of, the realization of that ideal. And while the notion of equality of civic status is not foreign to republicanism, I will make clear over the course of

is not necessary to explain our judgments of paradigm cases like the kindly slave master. They argue that freedom as noninterference, correctly understood, can explain these cases. See Goldwater, “Freedom and Actual Interference”; and Carter, “How Are Power and Unfreedom Related?”

- 4 On the social or “relational” conception of equality, see Anderson, “What Is the Point of Equality?”; Scheffler, “What Is Egalitarianism?”; Lippert-Rasmussen, *Relational Egalitarianism*; and Schemmel, *Justice and Egalitarian Relations*. Some theorists, such as Anderson and Schemmel, also identify a connection between nondomination and relational equality. On these accounts, however, nondomination appears to be taken as a notion of freedom that is a constituent part of social equality. See also Schuppert, “Non-Domination, Non-Alienation, and Social Equality” and Garrau and Laborde, “Relational Equality, Non-Domination, and Vulnerability.”
- 5 See Kolodny, “Being under the Power of Others.”

the paper that it is useful to more carefully distinguish that concern from the standard republican approach than is typically done.

The paper proceeds as follows. In section 1, I explain the notion of freedom as nondomination and how the insufficiency and constitution claims are typically defended. In section 2, I discuss the impossibility and moralization objections, evaluating and critiquing the responses recently made by republicans. In section 3, I characterize what I call the “status-conferring” function of protecting persons from domination and contrast it with the standard “freedom-constituting” account. I argue that the proposed approach provides a compelling defense of alternative insufficiency and constitution claims and avoids the moralization and impossibility objections. In section 4, I conclude by suggesting some other advantages of the approach developed.

1. NONDOMINATION AND SOCIAL JUSTICE

A person is free from domination with regard to a particular choice if no agent has an “uncontrolled” capacity to interfere with that choice. While there is room for republicans to differ over what exactly is required for the relevant kind of control, Lovett and Pettit write that “there is widespread agreement . . . that control must robustly protect *A* against interference: that is, protect them across possible variations in *A*’s own preference in the choice and, more importantly, across possible variations in *B*’s preference as to how *A* should choose.”⁶ A slave may enjoy noninterference in the actual world under a benevolent master. But the slave is not free from domination, since they do not enjoy noninterference in the relevant range of possible worlds, e.g., those in which the master is no longer benevolent. The slave remains subject to the master’s uncontrolled capacity to interfere. Thus, the insufficiency claim: no amount of goodwill or virtuous self-restraint is enough to realize freedom.

The constitution claim follows almost immediately, to the extent that law is the best means we have available to make noninterference robust.⁷ On the republican account, the law is constitutive of freedom because what it is—or at least part of what it is—to be free involves being protected from interference in possible worlds where the goodwill of would-be interferers no longer remains. This contrasts with the “liberal” view of freedom, on which, as Bentham says,

6 Lovett and Pettit, “Preserving Republican Freedom.”

7 To be sure, other ways of making noninterference robust are conceivable. Imagine a new technology is invented—a brain chip that gives one a shock if they form the intention to interfere with another.

“all coercive laws are, as far as they go, abrogative of liberty.”⁸ Spelling out this contrast, Pettit writes that

republicans do not say, in the modernist manner, that while the law coerces people and thereby reduces their liberty, it compensates for the damage done by preventing more interference than it represents. They hold that the properly constituted law is constitutive of liberty in a way that undermines any such talk of compensation.⁹

We can say, then, that on the republican view law is “freedom constituting”: when suitably designed and implemented, law brings freedom into existence. By contrast, on the liberal view the law is “freedom promoting”: law limits freedom in the first instance but brings about more of it overall.¹⁰ On the freedom-promoting view, law is a causal means to noninterference in the actual world; on the freedom-constituting view, law is a constitutive means to *robust* noninterference. While the best system of law will interfere in the actual world (and so be “abrogative” of liberty in that sense, as Bentham says), it will not interfere in an uncontrolled or arbitrary manner—and thus will not reduce freedom as nondomination.

Republicans not only hold that nondomination is the correct understanding of freedom; they claim, further, that freedom as nondomination is the fundamental concern of justice. Thus, there is a second respect in which the republican approach differs from the so-called liberal view, since liberals tend to regard justice as involving a plurality of concerns—a balance of freedom and equality, for example. But in order to develop the general account of nondomination into a theory of social justice, we require an account of three issues: (1) the *domain* of choices that require protection from domination; (2) the *manner* in which those choices are to be protected; and (3) the nature of the *distribution* of undominated choice across persons.¹¹ On Pettit’s account, a theoretical device known as the “eyeball test” is used to fill in these details, on the assumption that the provision of nondomination is required to express, borrowing Dworkin’s phrase, “concern and respect for persons as equals.”¹² The eyeball test invokes the republican notion of the free person, or *liber*, who can

8 Bentham, “Anarchical Fallacies.”

9 Pettit, *Republicanism*, 35.

10 For a recent discussion of this issue, see Goldwater, “Freedom and Actual Interference,” 150–53.

11 Cf. Southwood, “Republican Justice.”

12 Pettit, *On the People’s Terms*, 78.

“walk tall without reason for fear or deference.”¹³ When the test is passed, each person can “look others in the eye” without having to “bow or scrape, toady or kowtow, fawn or flatter.”¹⁴

First, consider the *domain* issue. Pettit does not hold that every choice must be protected for justice to be realized. Instead, justice concerns nondomination in the domain of the *basic liberties*.¹⁵ The basic liberties are a package of liberties, familiar to both liberal and republican traditions, that include “those of speech, association, residence, employment, ownership, and exchange.”¹⁶ Intuitively, even if you dominate a trivial or nonbasic choice of mine, such as my using your pool, that is not enough for me to fail the eyeball test: I can still look you in the eye without reason for fear or deference.¹⁷ But if you dominate my *basic liberties*—my ability to speak in public, say, or my ability to move freely—then I would seem to have good reason to avoid any eye contact that might invite you to exercise your dominating power; and I may have reason to ingratiate myself to you, “keeping you sweet,” such that I might continue to enjoy noninterference across these options.

Second, consider the issue of the *manner* in which nondomination is realized. In principle, people’s basic liberties could be protected from domination in a range of different ways. Suppose, for example, that a technologically advanced state has a secret police force that prevents violations of people’s basic liberties before they happen.¹⁸ This may be enough to make this state’s citizens free from domination; but it might not be enough for them to pass the eyeball test. For instance, if this police force is indeed secret, it would not be a matter of common awareness that each citizen is protected, in the sense that each citizen knows they are protected, knows that others know they are protected, and knows that others know that they know, and so on.¹⁹ By contrast, however, Pettit maintains that nondomination of persons, in order to realize justice, must be realized through a system of *public* laws and norms. Public laws and norms provide protection, objectively, which, intersubjectively, is a matter of common awareness such that persons enjoy “a publicly established and acknowledged status in relation to others.”²⁰

13 Pettit, “Free Persons and Free Choices” and “The Basic Liberties.”

14 Pettit, *On the People’s Terms*, 88.2012

15 In more detail, Pettit identifies the basic liberties with those that are both “co-exercisable” and “co-satisfiable.” See Pettit, *On the People’s Terms*, 94, 98.

16 Lovett and Pettit, “Preserving Republican Freedom,” 366.

17 I borrow this example from McCammon, “Domination.”

18 As in Dick, *The Minority Report*.

19 Pettit borrows this understanding of common awareness from Lewis, *Convention*.

20 Pettit, *On the People’s Terms*, 83.

Finally, consider the *principle* question. We have identified the basic liberties as the relevant domain, but this leaves open the question of how nondominated choice within the basic liberties ought to be distributed across persons.²¹ One could hold the view that nondominated choice should be strictly equally distributed, such that each person has the same choices with regard to “speech, association, residence, employment, ownership, and exchange.” Unsurprisingly, Pettit adopts a less demanding principle than this. His preferred principle is *sufficientarian* “in the currency of free or undominated choice,” requiring that all have resources and protection to meet “a certain threshold . . . of free choice” required for them to pass the eyeball test.²² While sufficientarian in this sense, this principle is egalitarian in another: when all possess the resources and protection needed to pass the eyeball test, Pettit claims, persons will have an *equal* status as nondominated. Intuitively, *some* degree of inequality with regard to choices within the basic liberties—e.g., not all have the exact same choices of employment or ownership or residence—need not make some fail the eyeball test or fall short of having an equal status as nondominated.

In sum, then, this is the contemporary republican account: justice requires that each person has a public status as nondominated, in virtue of no agent having the uncontrolled capacity to interfere with their basic liberties, achieved through a system of public laws and norms.

2. THE MORALIZATION AND IMPOSSIBILITY OBJECTIONS

I now turn to examine the *moralization* and *impossibility* objections. I will explain both in turn and evaluate the responses that republicans have given in the recent literature, drawing the conclusion that the objections raise substantial problems for the republican account.

2.1. Moralization

A conception of freedom is moralized if its definition involves other moral notions. On a moralized conception of freedom, certain forms of interference—taxation, imprisonment—may not count as infringements of freedom if they are *justified*. If there is disagreement about what counts as appropriate justification, then, there will be disagreement about whether certain forms of interference count as restricting freedom. Suppose we disagree about whether it is just to incarcerate someone for recreational drug use. If we adopt a moralized

21 More precisely, there remains a question about how to distribute nondominated choice within the basic liberties, consistent with equal concern and respect. Pettit, *On the People's Terms*, 78.

22 Pettit, *On the People's Terms*, 88.

conception of freedom, then we will disagree about whether incarceration on these grounds makes a person unfree. In a similar way, people might disagree about whether a system of private property restricts or protects freedom, since, as G. A. Cohen points out, protecting one person's freedom to own several homes removes others' freedom to live in those homes.²³

Moralized conceptions of freedom are typically regarded as problematic for several reasons.²⁴ For one, they seem to violate a kind of "ordinary language" desideratum that "the conception displays adequate level of fidelity to ordinary-language use."²⁵ This desideratum would be violated, for instance, by moralized conceptions of freedom that count the guilty prisoner as perfectly free, despite being behind bars.

But there is another reason why the issue of moralization is troubling for republicanism in particular. As we have seen, the project of contemporary republicanism is to take freedom as the foundational political concern. As such, it would be especially problematic for the approach if freedom as nondomination involved other moral notions. It would then be these other (unexplained) values, and not just freedom as nondomination, that are at the foundation of the theory.

It is no surprise, then, that Pettit resists the moralization charge. He gives the following reply.²⁶ There is simply a matter of fact whether people have a capacity for uncontrolled interference. For example, it is either true or false whether a person enjoys noninterference robustly, in possible worlds where goodwill breaks down, and so on. If true, then as a matter of fact freedom of nondomination will be realized, and if false, then not. Whether nondomination is realized in a given case depends only on descriptive facts about the presence of control.

Does this reply succeed? Christian List and Laura Valentini argue that it does not, since it has unacceptable implications. They ask us to consider a democratic decision that involves a conflict of interest between different members of society (as is inevitable). Consider the decision whether to introduce a new inheritance tax. While the rich do not support the introduction of this new tax, everyone else does, and so the tax is introduced. Are the rich dominated?

23 Cohen, "Capitalism, Freedom, and the Proletariat."

24 For a recent argument in opposition to the commonly held view, see Bader, "Moralizing Liberty."

25 List and Valentini, "Freedom as Independence," 1051.

26 Pettit says that "unlike 'legitimate,' 'nonarbitrary' is not an evaluative term but is defined by reference to whether as a matter of fact the interference is subject to adequate checking. Interference will be nonarbitrary to the extent that, being checked, it is forced to track the avowed or avowal-ready interests of the interferee; and this, regardless of whether or not those interests are true or real or valid, by some independent moral criterion" ("Republican Liberty," 117). I have adapted the point in terms of "control" rather than "interest tracking," since the former notion has replaced the latter in the more recent work.

While it is true that each person exercises an equal share of control over the decision, this does not seem to give them “control” in any sense that would preserve their freedom.²⁷

A natural response would be to say that since the procedure was fair, or something else, it should not be regarded as involving domination. Or one could say this with respect to the outcome. But in either case, we would be invoking fairness, or some other value, and so the judgment would be moralized.

A more plausible and perhaps more likely republican response (one that List and Valentini do not consider) might draw on the distinction between free persons and free choices. The inheritance tax does not bring the rich below the threshold of sufficiency needed for them to enjoy a status as free persons. Although the rich may be dominated with respect to their choice to keep their inheritance free from taxation, this is not enough for them to be dominated as persons, because their basic liberties are still protected. Being less able to inherit wealth does not threaten anyone’s ability to look others in the eye “without reason for fear and deference.”

This response, however, only pushes back the moralization concern without resolving it. Invoking the notion of the “free person” along with the eyeball test, I contend, involves a kind of moralization. The notion of the “status of a free person” is not wholly descriptive, but instead (tacitly) involves moral commitments. As we saw previously, having the status of a “free person” involves meeting a particular threshold of undominated choice within the basic liberties. But this invites the question of why the relevant “status as free person” does not consist in a different set of liberties, or a different principle of distribution. Some might argue, for example, that the economic liberties should be counted as “basic.”²⁸ Others might argue that the sufficientarian principle is far too permissive for the relevant kind of equal status required by justice. Thus, both elements of what it is to be a free person—the delineation of which liberties are “basic,” as well as how undominated choice is to be distributed across persons—involve a choice among a range of options. Moreover, these would appear to be exactly the kinds of choices that call for resolution by appeal to moral concerns. On this basis, then, I believe the *moralization* objection holds.

2.2. Impossibility

The impossibility objection claims that, as a matter of living in society, people are inevitably subject to others’ uncontrolled ability to interfere, if not on the

27 Unlike, perhaps, when you ask your partner to lock up the keys to the liquor cabinet, which later frustrates your akratic desires. See Pettit, *On the People’s Terms*, 171.

28 For such an account, see Tomasi, *Free Market Fairness*.

part of individuals, then by *groups*, whether these groups are actual or merely possible agents.

Consider the following cases put forward by Thomas Simpson:

Nearly Coordinated Masters: There are three masters and a slave. No one master is strong enough to interfere in the slave's choices, but any two can do so by coordinating their efforts. Master 1 is ready to coordinate with either Master 2 or Master 3, but the latter are both benevolent and do not.

Uncoordinated Masters: There are three masters and a slave. No one master is strong enough to interfere in the slave's choices, but any two can do so by coordinating their efforts. All three masters are benevolent.²⁹

In both cases, the slave appears to be dominated, as Simpson suggests. The slave does not enjoy noninterference robustly: if the masters no longer remain goodwilled, then they will coordinate and interfere with the slave. As such, the slave will have reason to "keep them sweet" and would fail the eyeball test.

The next, and crucial, step in Simpson's argument is to claim that, in *any* society, there are possible groups that are relevantly similar to the uncoordinated or nearly coordinated masters. In any society, for any person, he claims, there will inevitably be

a collection of agents who have an uncontrolled power to invade her, by coordinating. It does not matter how improbable it is that they will do so; the possibility that they may makes her unfree. No one can protect her from this interference except other people. But they too have the same power to coordinate, so the same danger arises again.³⁰

Domination between *individuals*, Simpson agrees, or between individuals and small groups, might be removed through suitable laws and norms. However, this will not be enough to remove the dominating ability of every group. Republican justice, he reminds us, requires popular control of government. And if "The People" can control the government, they surely can interfere with any individual. Thus, Simpson concludes, "just by living among other people . . . one is dominated."³¹

Lovett and Pettit are not convinced by this argument, however. In response, they identify what they regard as a salient difference between the ability of

29 Simpson, "The Impossibility of Republican Freedom." Cf. Kolodny, "Being under the Power of Others."

30 Simpson, "The Impossibility of Republican Freedom," 37.

31 Simpson, "The Impossibility of Republican Freedom," 37.

“The People” to control government as opposed to the ability, or lack thereof, of the “The People” or other potential groups to interfere with an individual citizen’s basic liberties. The difference is what it takes for a group to be a *capable* dominating agent as opposed to a merely *potential* one. In order for a group to be capable of domination, three conditions must be met by its members: “first, they must each desire that they together interfere; second, they must be aware in common of their mutual desire; and third, they must be aware in common of a strategy whereby they can act on that desire.”³² Call the latter two conditions the *awareness* and *strategy* conditions.

In a just republic, these conditions will not be met by potential groups that might otherwise interfere with an individual’s basic liberties. When suitable republican laws and norms obtain, it will be incredibly costly for a person to signal a willingness to form such a group, as doing so would make one liable to punishment by law, ostracism, and other forms of sanction. Thus, even if the members of a potential group wish to interfere with one another, they will be prevented from making those intentions known and forming a joint strategy. As such, the awareness and strategy conditions will not be met. By contrast, Lovett and Pettit hold that these conditions may be met by “The People” acting to control the government, if suitable laws and norms are in place, if citizens are not too apathetic, and so on.³³ In a just republic, laws and norms allow citizens to make their intentions to team up against abuses of government power publicly known and to coordinate a joint strategy. “The People,” then, are capable of controlling government, but they do not dominate.

I contend, however, that this clarification of what republicans are committed undermines their defense of the insufficiency and constitution claims and reveals a significant limitation of the view so defended.³⁴ To see this, we need to take a small step back. Along with drawing the distinction between capable and merely potential dominating teams, Lovett and Pettit, *pace* Simpson, argue that the slaves in the Nearly Coordinated Masters and Uncoordinated Masters cases are *not dominated*. They claim that

for all that Simpson’s descriptions of the scenarios imply, the awareness and strategy conditions might not be satisfied. If they are not, the masters will not have the capacity to interfere, and thus not constitute a capable team holding sway over the slave. Thus, it is false to claim, as he does, that necessarily the nearly coordinated and uncoordinated masters dominate the slave.

32 Lovett and Pettit, “Preserving Republican Freedom,” 377.

33 Lovett and Pettit, “Preserving Republican Freedom,” 381.

34 For another reply, see Simpson, “Freedom and Trust.”

Lovett and Pettit, to remain consistent, must deny that the slave is dominated in these cases; otherwise, they would be unable to maintain that “The People” do not dominate as well.

This move, however, puts pressure on whether the distinctive republican claims can be adequately defended. Consider, first, the insufficiency claim. In both the Uncoordinated Masters and Nearly Coordinated Masters cases, it appears that goodwill *is* in fact sufficient for freedom. This is because the goodwill of the masters, as a matter of common awareness between them, prevents the “strategy” and “awareness” conditions from being satisfied. As such, the goodwill of the masters in these cases makes it such that there is only a potential dominating team, not a capable one.

Recall, also, that the insufficiency claim involves the thought, as Pettit puts it, that you “cannot make yourself free . . . by cozying up to the powerful and keeping them sweet.”³⁵ But this thought would seem to now be contradicted by Lovett and Pettit. Consider the following variation of Simpson’s case:

Unrobustly Coordinated Masters: There are two masters and a slave. No one master is strong enough to interfere in the slave’s choices, but the two can do so by coordinating their efforts. Master 1 is ready to coordinate with Master 2. Master 2 is also ready to coordinate with Master 1. Master 2, however, has a weakness for flattery and obsequiousness, such that whenever the slave fawns, toadies, or kowtows he becomes unwilling to interfere and signals this to Master 1.

In this case, the slave could ensure their freedom by “cozying” up to Master 2 and “keeping him sweet.” Doing so would ensure that the “strategy” and “awareness” conditions are not satisfied, and thus that the two masters do not form a capable dominating team. As such, Lovett and Pettit’s clarification of freedom as nondomination appears to result in the same “absurdity” that Pettit levels against the “noninterference” view: it entails that one can, in the kind of case imagined, make themselves free through ingratiation.³⁶ If there can be no freedom through ingratiation, however, it is not at all clear why should it make a difference whether ingratiation keeps a single master from exercising their capacity to interfere as opposed to preventing a potential dominating team from becoming capable.

Similarly, it is no longer quite so clear why the law has the kind of *constitutive* importance that republicans want to ascribe to it. If the slave is free from

35 See Pettit, “The Instability of Freedom as Noninterference,” 705, and “Freedom and Probability,” 216.

36 Pettit, “The Instability of Freedom as Noninterference,” 704–8.

domination in both *nearly* and *uncoordinated* cases, what further role is there for legal protection *qua* freedom constituter? Lovett and Pettit would, rightly, point out that there is a high probability that in either case the slave would come to be dominated, and that legal protection would remove this threat.³⁷ But this response only shows that suitable laws may make the slave more likely to continue to enjoy freedom as nondomination, not that the law has a freedom-constituting role.

One might reply that my point here does not undermine the fact that, *in normal circumstances*, the law will nonetheless have constitutive importance.³⁸ Freedom as nondomination, one might point out, may be constituted by more than one means. As Pettit says

to enjoy nondomination . . . is to have inhibitors present in your society—maybe these, maybe those—which prevent arbitrary interference in your life and affairs. And the presence of suitable inhibitors—suitable institutions and arrangements—represents a way of realizing your nondomination.³⁹

In the Nearly Coordinated Masters and Uncoordinated Masters cases, the greater strength of the slaves (that makes the masters' coordination necessary), along with the failure of the awareness and strategy conditions, act as "suitable inhibitors," or a constitutive means of freedom. To be sure, however, conditions such as these are quite unlikely to obtain; and so, given the normal conditions of human life, the law—or legal, political, and social institutions more broadly—are likely to be the best constitutive means available to realize nondomination.

This, I think, is a fair response. Nevertheless, it is not entirely satisfactory. The issue is that, in the cases considered above, there remains something very normatively problematic with the relationship between the slave and masters. Whether or not it renders them unfree, the slave still has an objectionable legal status, and so there remains an injustice that republicans (or anyone) ought to condemn. Moreover, I contend, the correction of this injustice is exactly of the sort we might expect to be (partly) constituted by suitable law. Thus, despite the republican insistence on the constitutive importance of law, they appear unable to vindicate that claim where we ought to expect them to—in these (admittedly strange and unlikely) cases.

37 Lovett and Pettit, "Preserving Republican Freedom," 379.

38 I would like to thank a referee for the *Journal of Ethics and Social Philosophy* for suggesting this response.

39 Pettit, *Republicanism*, 108.

It might be replied that we need to again appreciate the distinction between free persons and free choices. Even if the slave is not dominated in the Nearly Coordinated Masters and Uncoordinated Masters cases, the slave does not have the status of a free person. Introducing legal protection, then, would constitute that status. But while this response is promising, it is not available on the approach that Lovett and Pettit have taken. On their approach, what it is to have a status as a free person is nothing more than being free from domination as a matter of common awareness. If this is what it is to have the status of a free person, then that status might well be possessed by the slaves in the Nearly Coordinated Masters or Uncoordinated Masters cases, or a suitably specified variant. Imagine, for example, the following case:

The Society of Kindly Masters: There are two classes: one comprised of benevolent masters; the other, physically powerful slaves. The slaves are much more physically strong than the weak masters, such that it would take a group of masters to interfere with any one slave. No master wishes to interfere, however. Further, there are entrenched and public social norms that forbid ganging up on any slave, as doing so violates the demands of a kind of “noblesse oblige”: to do so would tarnish a master’s virtue.

This case is stated such that it is incredibly improbable that any slave would suffer interference. Further, the strategy and awareness conditions are not met, and so there is no capable team of masters. No slave suffers domination, and this is matter of common awareness, given the nature of the norms in the society. These conditions, then, may be enough for the slaves—and, for that matter, *everyone* in the society—to have a kind of public status as nondominated.

We have every reason, however, to think that this society would nonetheless fall short of realizing justice. But the republican account seems ill equipped to tell us why. Even if all might be free from domination, there nonetheless remains an objectionable kind of social hierarchy. The fact that the masters enjoy legal permission to interfere serves to mark the “slaves” as inferior.⁴⁰ Suppose, further, that there exist widespread inegalitarian patterns of deference and modes of address that perpetuate a common awareness that the “slaves” are regarded as “less than.” We may even suppose that these practices are regarded as fitting given the status of the slaves under law. After all, it is only out of a sense of “noblesse oblige” that the masters remain an incapable team and do not interfere.

40 The label of “slave” may not seem entirely apt here, given the absence of interference, but they are nonetheless social inferiors in a significant sense.

In this case, the masters' goodwill and lack of coordination is not enough for the slaves to be their social equals. And it is this problem that a suitable system of law might remedy in a constitutive manner. What this suggests, then, is that the republican ideas about the constitutive importance of law and the insufficiency of goodwill may not be best understood in connection with *freedom*, as they typically are. Instead, these ideas might be better explained and defended by examining the relationship between legal protection and social equality or equality of civic status. I pursue this idea next.

3. THE STATUS-CONFERRING FUNCTION OF LEGAL PROTECTION

On the standard republican view, we have seen, a person is free if no other agent has an uncontrolled ability to interfere with their basic liberties. By inhibiting interference, legal protection makes the enjoyment of noninterference *robust* and so *constitutes* freedom. By contrast, we saw that on the freedom as noninterference view, the inhibiting function of law *promotes* freedom—in the way that restricting some choices (to kill, steal, and so on) gives people more freedom overall. In addition to inhibiting behavior, however, law has an *expressive* or *communicative* function.⁴¹ It is this function, or rather an aspect of it, that I now wish to explore. An important aspect of this communicative function concerns the kind of status people are given under political institutions, or how they are *ranked*, e.g., as inferior to others or as equals.

My claim is that the equal provision of robust noninterference through a system of law functions to give persons a civic status as equals, thereby realizing an important form of social or “relational” equality. While it may be inevitable that some will enjoy greater esteem or recognition than others in domains such as athletics or the arts, suitable law can rank persons as *equal citizens*—hence the term “civic” status. Realizing this equality is important because it is necessary for government to express respect for its members as equals, which, on many accounts, is regarded as the first commitment of political theory.⁴² No government shows respect for its members as equals, I contend, if some are marked out as having an inferior or “less than” status.

It may be that it is impossible for the law—or any other social or political institution—to entirely remove the uncontrolled ability to interfere from all actual or possible agents (such as “The People” or other groups), as critics

41 See Anderson and Pildes, “Expressive Theories of Law”; Sunstein, “Law’s Expressive Function”; and Shiffrin, “Speaking amongst Ourselves.” Most recently, see Schemmel, *Justice and Egalitarian Relations*, ch. 2.

42 See Dworkin, *Sovereign Virtue*; Kymlicka, *Contemporary Political Philosophy*; Pettit, *On the People’s Terms*.

maintain. But this does not undermine the significance of suitable laws' status-conferring function. Recognition of this function, I will argue, provides a compelling defense of alternative *insufficiency* and *constitution* claims, understood in terms of social equality instead of freedom. Moreover, this account avoids the impossibility and moralization objections.

To begin, recall three of the cases considered previously: Uncoordinated Masters, Nearly Coordinated Masters, and Society of Kindly Masters. We saw that in each of these cases republicans hold that the slaves are not dominated, although the cases differ in terms of the probability of the slaves becoming dominated. And the slaves in these cases may even enjoy a kind of public "status as nondominated." Certainly, however, in each case the ideal of social equality is not realized. The slaves do not enjoy a status as *civic equals*. The goodwill of the masters, or their lack of coordination, is not sufficient to realize social equality, even if it is thought to give the slaves freedom. The fact that there exists legal permission to interfere with the slaves is enough to mark them out as inferior, despite there being no group willing and capable of acting on that permission.⁴³ The slaves are not protected as equals by a shared law and, for this reason, there remains an important sense in which they are subordinate to their masters. In general, then, I contend that in a society where some members' basic liberties are protected by law, then that protection must be given to persons as equals in order for them to enjoy an equal civic status.⁴⁴

To realize social equality, it is important that people are protected *as equals*, and not merely equally. To see this, suppose that in the Society of Kindly Masters the masters decide to legally entrench what they perceive as the requirements of "noblesse oblige." The slaves are given equal legal protection, not *as equal members* of the community, but instead on the basis of the masters' desire not to tarnish their virtue. Although the slaves may come to enjoy equal protection in such a way, this would not, intuitively, be enough to realize social equality. The protection and powers given to individuals under law must communicate respect for persons as equals. While this example is somewhat fanciful, there are real world analogues. Think, for example, of the issue of whether same-sex couples should be given the legal power to *marry* and not just enter civil partnerships. Even if both arrangements grant the same legal powers, one may argue that the ability to enter marriage should be extended on the grounds that is necessary for same-sex persons to fully enjoy a status as civic equals.

43 A different concern, raised by Fabian Wendt, is that slaves are denied self-ownership. See Wendt, "Slaves, Prisoners, and Republican Freedom."

44 More may be required: that differences in the worth (in Rawls's sense of that term) of liberties are not too great, that resources and capabilities are not distributed too unequally, and so on.

There are, then, two ways to understand the importance of legal protection and the robust noninterference it provides. On the standard republican view, such protection constitutes freedom by removing others' uncontrolled capacity for interference. By contrast, on the account developed here, such protection, given to persons as equals, is required to realize social equality understood in terms of equality of civic status. This approach is perhaps not so far off from the standard republican view, one might think, given that the goal of republican justice is sometimes described in terms of giving persons an equal status as free persons. But there is an important difference to highlight. On the standard republican account, what it is to have a status as a free person is defined in terms of enjoying freedom as nondomination (over the basic liberties) as a matter of common awareness. By contrast, on the account I propose, being protected from domination is not sufficient for equality of civic status; and protection from domination is only necessary when it is needed to secure equality of civic status. On the one hand, nondomination is not sufficient because, as we saw in the *Nearly Coordinated Masters*, *Uncoordinated Masters*, and *Society of Kindly Masters* Cases, one can be protected from domination yet fail to have an equal civic status. On the other hand, even if Simpson and others are correct that each person is inevitably dominated by possible teams, such as "The People," this need not undermine equality of civic status. Domination by "The People" or other possible groups does not socially downgrade particular people in the way that occurs when an individual dominates others.⁴⁵ In this way, this account avoids the impossibility objection. Justice forbids agents having a capacity for uncontrolled interference when, and because, it marks out some as inferior or as second-class citizens. Equality of civic status requires that persons are protected from the dominating power of other individuals, but not from "The People" or other possible groups. Borrowing a line from Rousseau, we might say that social equality can be realized when "each citizen is in a position of perfect independence from all the others and of excessive dependence upon the city."⁴⁶

What about the moralization objection? The approach I propose actually allows moralization and, indeed, requires it. This is because the ideal of people

45 Kolodny notes an interesting point in this regard: it is not clear that it is even intelligible to think of social hierarchy between an individual and a group agent. He asks, "what would it even mean for me to be the 'equal' of Indonesia, say, or the Roman Catholic Church?" ("Being under the Power of Others," 112). Nevertheless, there may be objectionable cases of certain groups, such as powerful corporations, dominating individuals. I take it, however, that these cases can be explained by the concern that particularly powerful individuals within these groups, e.g., executives, are rendered superior to others.

46 Rousseau, *On the Social Contract*, 2.12.3.

sharing a status as equals is incomplete and in need of specification. We require an account of the *kind* of status to be shared as equals. One way this account can be filled in is in connection with a particular conception of the person. For example, Rawls's conception of the person as a bearer of the "two moral powers" might work to delineate a set of basic liberties, the protection of which is constitutive of a kind of equality of status.⁴⁷ Alternatively, we could invoke the conception of the free person or *liber* found in the republican tradition.⁴⁸ This route would move us closer to the approach taken by contemporary republicans. But we would not be regarding nondomination, a conception of the freedom, as the fundamental principle of justice; instead, we would be working with the ideal of social equality and a moral conception of the person.⁴⁹

For our purposes here, however, we can leave this issue open, and instead focus on how the proposed account vindicates these alternative insufficiency and constitution claims understood in terms of social equality instead of freedom. Based on the account thus far, one may remain doubtful: it might be thought that the legal provision of robust noninterference to persons as equals is *one* way to realize equality of status, or social equality. But perhaps it is not the *only* way. If people have the right dispositions toward one another, or if the right kinds of social norms exist, then one might think that we would have all the social equality we could want and introducing law would not provide any further enhancement. This kind of picture seems to be suggested, for instance, by Niko Kolodny's characterization of social equality as involving people having "resolute dispositions" not to interfere with one another, as "something to which they are entitled."⁵⁰

47 The two moral powers are the capacity to form, revise, and pursue a conception of the good and the capacity for a sense of justice. See Rawls, "The Basic Liberties and Their Priority." An important difference to highlight between Rawls's view and the one taken here is as follows. Rawls requires the equal basic liberties on the grounds that the equal status they confer is necessary for self-respect and the development and exercise of the two moral powers. By contrast, on my account, the equal basic liberties are necessary for social equality, which is taken to have noninstrumental importance. For further discussion, see Cass, "The Priority of Liberty."

48 Pettit, "The Basic Liberties."

49 One could also (in principle) work out a conception of equality of status in connection with other ideals, such as persons having a status as "equal self-owners" in a libertarian vein, or as "equal owners of the means of production" in a Marxist manner.

50 Kolodny, "Rule over None II." Kolodny does not mean this to be a complete account of social equality, so he could in principle agree with the account I give. But nothing he says suggests he would support the *insufficiency* and *constitutions* claims about social equality that I defend.

I contend, however, that there is reason to think that any set of interpersonal dispositions or social norms will fall short of fully realizing social equality. Interpersonal dispositions and social norms, however specified, will be deficient in several respects. These defects are that the *importance* and *basis* of people's entitlement to noninterference will be uncertain and may fail to be a matter of common awareness; further, people will not have the power to make *claims*. The introduction of suitable law, then, is required to remedy these defects and fully realize social equality.⁵¹ I will explain each defect and their solutions in turn.

First, consider the issue of importance. In any community, there are bound to be a range of esteem hierarchies that rank people unequally along various dimensions. Some will rank highly in athletics, business, the arts, politics, and so on. And different communities will place greater or lesser importance on some domains over others. Given these kinds of facts, a shared commitment to noninterference may not be enough to manifest social equality. To illustrate the point, consider a community in which all members place the utmost importance on one domain of achievement or another—skill in combat and athletics, say. If, as is inevitable, people perform unequally in this domain, then people will come to have an unequal status in a significant sense. Given that there is a particularly salient domain of esteem, those who perform especially well in that domain are likely to be regarded as “better than” or superior, and those who perform poorly, as inferior. In a community such as this, it is not clear that a shared commitment to noninterference would do much to mitigate the presence of an otherwise pervasive social ranking.

The protection of people's basic liberties as equals through law provides a remedy to this problem. The basic liberties assign people a status as equal citizens, and this status may have a kind of eclipsing effect over various esteem hierarchies.⁵² Though esteem hierarchies may provide some reason for people to regard others as superior or inferior, it might be said that that legal protection can give people an “exclusionary reason” to regard one another as equals.⁵³

Next, consider the issue of the *basis* of people's enjoyment of noninterference. As we have seen, social equality does not simply involve the concern that individuals enjoy noninterference robustly, but that they do so as equals. Without law, however, it may remain uncertain whether people are entitled to noninterference on the basis of their equality and not on some other basis. In

51 My account draws on Hart, *The Concept of Law*.

52 This point bears similarity to Rawls's discussion of the basic liberties and self-esteem/respect. See Rawls, *A Theory of Justice*. Also, for a helpful complementary discussion, see Porro, “Esteem, Social Norms and Status Inequality.”

53 On this notion, see Raz, *Practical Reason and Norms*.

this regard, Kolodny's locution that social equality involves noninterference as "something" to which others are entitled is problematic. There is a range of possibilities here. A slave master, as we have seen, could be disposed to not interfere with their slave, not because the slave has any moral standing, but because it shows a bad character to do so. Alternatively, the powerful might be disposed not to interfere with the weak based on a kind of perceived moral standing—but it may not be an *equal* moral standing: perhaps they regard them as inferior, though still worthy of some moral concern. Again, the introduction of an equal civic status conferred by suitable legal protection provides a remedy to this problem, making it explicit and a matter of common awareness that people are entitled to noninterference on the basis of their equality. Legal norms make possible a kind of "mediated accountability," such that people may come to owe each other noninterference on the basis of an institutionally defined status as an equal citizen, and not as a particular individual.⁵⁴

Finally, consider the capacity of persons to make *claims* under law. A slave master may regard their slave as being entitled to noninterference, but the slave would nonetheless not be able to *claim* noninterference. In order to make a claim, there must be procedure and set of rules in place by which a person can have any violations adjudicated and remedied. Even if people are disposed not to interfere with one another, and noninterference is widely supported by social norms (think again of the Society of Kindly Masters), this would not be enough to give persons the power to have their entitlement to noninterference upheld. Even if claims are exceedingly unlikely to be exercised, their possession nonetheless makes a difference for the kind of status people have: it will be common knowledge not only that people are owed certain forms of treatment, but also that they have the power to guarantee that treatment or demand compensation for any violation.⁵⁵ Suitable law gives persons a forum to enforce and challenge norm violations on their own behalf; by contrast, social norms tend to be enforced by the judgments and opinions of the community at large.⁵⁶ Thus, legal protection, by giving people the power to make claims, provides a status benefit that interpersonal dispositions and social norms cannot.

54 I borrow the term "mediated accountability" from Brennan et al., *Explaining Norms*, ch. 3. Some Kantians argue for what seems to me to be a complementary idea. They claim that without a shared system of law, any enforcement of one's rights amounts to arrogating authority in a problematic way, imposing one's will *unilaterally*. A suitable system of law, by contrast, subjects all equally to an *omnilateral* will. See Sinclair, "The Power of Public Positions." Again, though, this is seen as an issue of *freedom*, not of social equality.

55 Feinberg "The Nature and Value of Rights."

56 Brennan et al., *Explaining Norms*.

One might argue that my points do not show that law is required to solve the problems raised; instead, it might be claimed that we need to more carefully specify the relevant kinds of dispositions or social norms needed.⁵⁷ For instance, we can imagine a community in which everyone is resolutely disposed not to interfere with one another on the basis of each person's equal standing; and, further, we can imagine that this standing is taken to have a kind of priority over unequal rankings in various domains of esteem. And we might imagine that there are social norms in place that allow persons to make claims.

This community, however, would in effect have a system of law, on a widely accepted account of what law is. On a Hartian conception, to have a legal system just is to have a union of primary and secondary rules, where primary rules are social norms and secondary rules are rules about those rules that "specify the ways in which the primary rules may be conclusively ascertained, introduced, eliminated, varied, and the fact of their violation conclusively determined."⁵⁸ The primary rules at issue are those that specify that people should not be interfered with. But notice that the line of argument just considered stipulates that there are rules about those rules (secondary rules) in place. There are rules specifying the basis and importance of the primary rule prohibiting interference; and there are rules of adjudication specifying how to resolve possible disputes, thereby giving people claims. This is enough, then, for that community to count as having a legal system.

This account, then, provides a defense of alternative insufficiency and constitution claims that differs from the standard republican line. The account I propose points to three defects that law remedies. Legal protection defines the importance and basis of noninterference and gives all claims as civic equals. Goodwill and virtuous self-restraint, and whatever associated dispositions or social norms, are insufficient to solve these problems. But law is uniquely suited to do so in a constitutive manner.

4. CONCLUSION

I have argued that the importance of protection from domination—more precisely, the provision of robust noninterference to persons as equals through law—is better understood in terms of its status-conferring function as opposed to its freedom-constituting function. This account vindicates the spirit of republicanism's distinctive commitments while at the same time avoiding

57 This kind of concern has been raised with regard to the freedom-constituting function of law by Frye, "Freedom without Law"; and Guillery, "Domination and Enforcement."

58 Hart, *The Concept of Law*, 92.

the moralization and impossibility objections that have long bothered critics. In conclusion, I want to consider two further possible advantages of the proposed approach.

The first is that, unlike the standard republican account, I do not claim that negative liberty must be regarded in terms of nondomination, and so my account is compatible with other conceptions of freedom. This might be an advantage insofar as we want different notions to play different theoretical roles; for instance, in cases where concerns with status and freedom seem to come apart. Consider, for example, the Patriot Act, which (oversimplifying) gave government a wider range of permissions to monitor, search, and detain persons suspected of terrorist activity.⁵⁹ Although the act granted the government the capacity to interfere with *anyone* in the society, one might think that those for whom freedom was most reduced were those that were most likely to be searched or detained.⁶⁰ In addition, however, we might think that these kinds of policies can affect the status of certain groups *as a whole*—in this case, Muslims—even if particular members of those groups suffer more than others with regard to a loss of freedom. Similarly, policies such as “stop and search” may have harmful status effects for entire groups, even if there are unequal effects on freedom with regard to individual members of those groups (e.g., poor versus wealthy Black Americans). I do not mean to say there are crystal-clear judgments here, only that there appear to be different dimensions of assessment, and it is a benefit of my account that it can allow their assessment in a more fine-grained way.

The second advantage involves the implications of my account for distributive justice. The contemporary republican account is often critiqued for being too permissive. For instance, we might think it is unjust for some people to barely reach Pettit’s sufficiency threshold, while others greatly surpass it.⁶¹ On the status-based approach I have proposed, there is room to develop a more demanding account. Consider, for instance, what it would take for people in a society to pass the eyeball test versus what would be required for persons to regard each other *as equals*. Among other things, this would require that the difference between those at the top versus those who barely pass the threshold

59 For an argument that freedom as nondomination cannot adequately capture what is significant about conflicts between liberty and security, see Goodin and Jackson, “Freedom from Fear.”

60 Consider, for instance, the hypothetical example of “Musa—a young, single, devout Muslim student from Pakistan who studies nuclear physics in the United States.” See Carter and Shnayderman, “The Impossibility of ‘Freedom as Independence,’” 144.

61 Southwood, “Republican Justice.”

is not too great, suggesting the addition of range constraints.⁶² Again, clear answers are not immediately obvious, but working out the details may be fruitful.⁶³

Institute of Philosophy, NOVA University of Lisbon
devoncass@fcsh.unl.pt

REFERENCES

- Anderson, Elizabeth. "What Is the Point of Equality?" *Ethics* 109, no. 2 (January 1999): 287–337.
- Anderson, Elizabeth, and Richard Pildes. "Expressive Theories of Law: A General Restatement." *University of Pennsylvania Law Review* 148, no. 5 (2000): 1503–75.
- Bader, Ralf M. "Moralizing Liberty." In *Oxford Studies in Political Philosophy*, vol. 4, edited by David Sobel, Peter Vallentyne, and Steven Wall, 141–66. Oxford: Oxford University Press, 2018.
- Bentham, Jeremy. "Anarchical Fallacies." 1843. In *The Works of Jeremy Bentham*, vol. 2, edited by John Bowring. Charleston, NC: Nabu Press, 2010.
- Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood. *Explaining Norms*. Oxford: Oxford University Press, 2013.
- Carter, Ian. "How Are Power and Unfreedom Related?" In Laborde and Maynor, *Republicanism and Political Theory*, 58–82.
- Carter, Ian. "A Critique of Freedom as Non-Domination." *The Good Society* 9, no. 3 (2000): 43–46.
- Carter, Ian, and Ronen Shnayderman. "The Impossibility of 'Freedom as Independence.'" *Political Studies Review* 17, no. 2 (May 2019): 136–46.
- Cass, Devon. "The Priority of Liberty: An Argument from Social Equality." *Law and Philosophy* 40, no. 2 (April 2021): 129–61.
- Cohen, G.A. "Capitalism, Freedom, and the Proletariat." In *Liberty*, edited by David Miller, 163–82. Oxford: Oxford University Press, 1991.

62 See discussion in Schemmel, *Justice and Egalitarian Relations*, esp. ch. 8.

63 I would like to thank two referees for the *Journal of Ethics and Social Philosophy* for helpful comments on this paper. I would also like to thank the late Geoffrey Brennan, Jesse Hamby, Josef Holden, Seth Lazar, Lars Moen, and Philip Pettit, as well as audiences at the Australian National University and an online workshop on republicanism at the London School of Economics organized by Nicolas Cote. The research for this article was supported by funding from the Social Sciences and Humanities Council of Canada as well as the Portuguese Fundação para a Ciência e a Tecnologia, reference PTDC/FER-FIL/6088/2020.

- Dick, Philip K. *The Minority Report*. New York: Knopf Doubleday Publishing Group, 2009.
- Dowding, Keith. "Republican Freedom, Rights, and the Coalition Problem." *Politics, Philosophy and Economics* 10, no. 3 (August 2011): 301–22.
- Dworkin, Ronald. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press, 2002.
- Feinberg, Joel. "The Nature and Value of Rights." *Journal of Value Inquiry* 4, no. 4 (December 1970): 243–60.
- Frye, Harrison P. "Freedom without Law." *Politics, Philosophy and Economics* 17, no. 3 (August 2017): 298–316.
- Goldwater, Jonah. "Freedom and Actual Interference." *Journal of Ethics and Social Philosophy* 17, no. 2 (April 2020): 133–58.
- Goodin, Robert, and Frank Jackson. "Freedom from Fear." *Philosophy and Public Affairs* 35, no. 3 (Summer 2007): 249–65.
- Garrau, Marie, and Cécile Laborde. "Relational Equality, Non-Domination, and Vulnerability." In *Social Equality: On What It Means to Be Equals*, edited by Carina Fourie, Fabian Schuppert, and Ivo Walliman-Helmer, 45–64. New York: Oxford University Press 2015.
- Guillery, Daniel. "Domination and Enforcement: The Contingent and Non-Ideal Relation between State and Freedom." *Politics, Philosophy and Economics*, 19, no. 4 (November 2020): 403–23.
- Hart, H. L. A. *The Concept of Law*. Oxford: Oxford University Press, 1961.
- Kramer, Matthew. "Liberty and Domination." In Laborde and Maynor, *Republicanism and Political Theory*, 31–57.
- Kolodny, Niko. "Being under the Power of Others." In *Republicanism and the Future of Democracy*, edited by Geneviève Rousselière and Yiftah Elazar, 94–114. Cambridge: Cambridge University Press, 2019.
- . "Rule over None II: Social Equality and the Justification of Democracy." *Philosophy and Public Affairs* 42, no. 4 (Fall 2014): 287–336.
- Kymlicka, Will. *Contemporary Political Philosophy: An Introduction*. Oxford: Oxford University Press, 2002.
- Laborde, Cécile, and John Maynor, eds. *Republicanism and Political Theory*. Malden: MA, Blackwell Publishing, 2009.
- Lewis, David K. *Convention: A Philosophical Study*. Hoboken, NJ: Wiley-Blackwell, 1969.
- Lippert-Rasmussen, Kasper. *Relational Egalitarianism: Living as Equals*. Cambridge: Cambridge University Press, 2019.
- List, Christian, and Laura Valentini. "Freedom as Independence." *Ethics* 126, no. 4 (July 2016): 1043–74.
- Lovett, Frank. *A General Theory of Domination and Justice*. Oxford: Oxford

- University Press, 2010.
- Lovett, Frank, and Philip Pettit. "Preserving Republican Freedom: A Reply to Simpson." *Philosophy and Public Affairs* 46, no. 4 (Fall 2018): 363–83.
- McCammon, Christopher. "Domination: A Rethinking." *Ethics* 125, no. 4 (July 2015): 1028–52.
- McMahon, Christopher. "The Indeterminacy of Republican Policy." *Philosophy and Public Affairs* 33, no. 1 (January 2005): 67–93
- Pallikkathayil, Japa. "Deriving Morality from Politics: Rethinking the Formula of Humanity." *Ethics* 121, no. 1 (October 2010): 116–47.
- Pettit, Philip. "The Basic Liberties." In *The Legacy of H. L. A. Hart: Legal, Political, and Moral Philosophy*, edited by Matthew H. Kramer. Oxford: Oxford University Press, 2008.
- . "Freedom and Probability: A Comment on Goodin and Jackson." *Philosophy and Public Affairs* 36, no. 2 (Spring 2008): 206–20.
- . "Free Persons and Free Choices." *History of Political Thought* 28, no. 4 (Winter 2007): 709–18.
- . "The Instability of Freedom as Noninterference: The Case of Isaiah Berlin." *Ethics* 121, no. 4 (July 2011): 693–716.
- . *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press, 2012.
- . *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press, 1997.
- . "Republican Liberty: Three Axioms, Four Theorems." In Laborde and Maynor, *Republicanism and Political Theory*, 83–101.
- . "Two Republican Traditions." In *Republican Democracy: Liberty, Law and Politics*, edited by Andreas Niederberger and Philipp Schink, 169–204. Edinburgh: Edinburgh University Press, 2013.
- Porro, Costanza. "Esteem, Social Norms and Status Inequality." *Ethic Theory and Moral Practice* 24, no. 4 (September 2021): 901–15.
- Rawls, John. "The Basic Liberties and Their Priority." In *Liberty, Equality, and Law: Selected Tanner Lectures on Moral Philosophy*, edited by Sterling M. McMurrin. Salt Lake City: University of Utah Press, 1987.
- . *A Theory of Justice*. Rev. ed. Cambridge, MA: Harvard University Press, 1999.
- Raz, Joseph. *Practical Reason and Norms*. London: Hutchinson, 1975.
- Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press, 2009.
- Rousseau, Jean-Jacques. *On the Social Contract*. Indianapolis: Hackett Publishing, 1988.
- Scheffler, Samuel. "What is Egalitarianism?" *Philosophy and Public Affairs* 31,

- no. 1 (Winter 2003):5–39.
- Schemmel, Christian. *Justice and Egalitarian Relations*. Oxford: Oxford University Press, 2021.
- Schuppert, Fabian. “Non-Domination, Non-Alienation and Social Equality: Towards a Republican Understanding of Equality.” *Critical Review of International Social and Political Philosophy* 18, no. 4 (2015): 440–55.
- Sharon, Assaf. “Domination and the Rule of Law.” In *Oxford Studies in Political Philosophy*, vol. 2, edited by David Sobel, 128–55. Oxford: Oxford University Press, 2016.
- Shiffrin, Seana. “Speaking amongst Ourselves: Democracy and Law.” *The Tanner Lectures on Human Values*, 2017. https://tannerlectures.utah.edu/_resources/documents/a-to-z/s/Shiffrin%20Manuscript.pdf.
- Simpson, Thomas W. “Freedom and Trust: A Rejoinder to Lovett and Pettit.” *Philosophy and Public Affairs* 47, no. 4 (Fall 2019): 412–24.
- . “The Impossibility of Republican Freedom.” *Philosophy and Public Affairs* 45, no. 1 (Winter 2017): 27–53.
- Sinclair, Thomas. “The Power of Public Positions: Official Roles in Kantian Legitimacy.” In *Oxford Studies in Political Philosophy*, vol. 4, edited by David Sobel, Steven Wall, and Peter Vallentyne, 28–52. Oxford: Oxford University Press, 2018.
- Skinner, Quentin. “Freedom as the Absence of Arbitrary Power.” In Laborde and Maynor, *Republicanism and Political Theory*, 83–101.
- Southwood, Nicholas. “Republican Justice.” *Critical Review of International Social and Political Philosophy* 18, no. 6 (2015): 669–78.
- Stilz, Anna. *Liberal Loyalty: Freedom, Obligation, and the State*. Princeton, NJ: University Press, 2009.
- Sunstein, Cass R. “Law’s Expressive Function.” *The Good Society* 9, no. 2 (1999): 55–61.
- Tomasi, John. *Free Market Fairness*. Princeton, NJ: Princeton University Press, 2013.
- Waldron, Jeremy. “Pettit’s Molecule.” In *Common Minds: Themes from the Philosophy of Philip Pettit*, edited by Robert Goodin, Geoffrey Brennan, Frank Jackson, and Michael Smith, 151–54. Oxford: Clarendon Press, 2007.
- Wendt, Fabian. “Slaves, Prisoners, and Republican Freedom.” *Res Publica* 17, no. 2 (May 2011): 175–92.

ON AN ANALYTIC DEFINITION OF LOVE

Tyler J. VanderWeele

LOVE is a notoriously difficult concept to characterize. The numerous uses and nuances of the word are varied and complex. The forms of love that can be found within human life, and concerning the things we love, are very diverse. It has thus been suggested that there is no analytic definition of love, and that none is possible. Hacker writes, “There is no analytic definition of love that captures in its net the use of the word.”¹ I hope to show in this paper that more can perhaps be accomplished toward providing an analytic definition of “love” than is suggested by this claim.

The present paper will put forward a proposal for the conceptual grammar—that is, a characterization of our use of language—concerning the expression “He/she loves . . .” where this expression is followed by some object that may, or may not, be a person. I will argue, as have others, that uses of such expressions involve either desire for some good or desire that good come to the object specified, but I will argue, in contrast to others, that it is the nonexclusive disjunction of these two desires that can be used to provide a precise analytic characterization of our use of “love” in ordinary language in English.²

The task in this paper is an analytic one, concerning our ordinary use of the word “love.” I will thus not directly be considering questions of what love *ought* to look like or how some might argue the word “love” *should* be used. I will only very briefly touch on the question of the various reasons for love or causes of love.³ And I will do so principally insofar as this is required to defend

1 Hacker, *The Passions*, 269.

2 Aquinas, Taylor, Sidgwick, and Stump, for example, all make reference both to the desire to be with, or be united with, the beloved and to the desire to contribute or have good come to the beloved object, sometimes requiring both desires to constitute love (e.g., Stump), whereas here it is being proposed that the nonexclusive disjunction of the two can be used to characterize our use of “love” in ordinary language. See Aquinas, *Summa Theologica*, I.II.26.4; Taylor, “Love,” 157; Sidgwick, *The Methods of Ethics*, 244; and Stump, “Love, By All Accounts,” 27.

3 Cf. Kraut, *Love De Re*; Velleman, “Love as a Moral Emotion”; Kolodny, “Love as Valuing a Relationship”; Stump, “Love, By All Accounts”; Frankfurt, *The Reasons of Love*; Jollimore, *Love’s Vision*; Setiya, “Love and the Value of a Life.”

the proposed analytic definition, though I will also return to the question at the paper's conclusion, when addressing some practical considerations and implications arising from the proposed definition. These other topics are undoubtedly important. However, there is arguably also value in focusing on ordinary language, both to better understand how that language is used in practice and to help resolve a number of puzzles and paradoxes concerning love.

While the principal aim of the paper is to propose and defend an analytic definition of love, doing so will also require considering the characteristics of love as they pertain to the proposed definition; questions concerning the strength and intensity of love, and of loving well; and issues related to the reciprocity of interpersonal love. Consideration of these issues will be necessary to provide an adequate account of how we in fact use our language concerning love and to explain a number of paradoxes that seemingly arise with our use of that language. In defending the proposed analytic definition, I will also argue that it is applicable to a diversity of objects, including not only love of inanimate objects, places, activities, ideals, and people but also love of God, love of neighbor, love of stranger, love of enemies, self-love, and love of pets. I will conclude with some of the practical considerations that arise from the proposed definition as it pertains to cultivating love within society, within marriages, and within political life as well as to the role of the media in shaping our loves.

1. LOVE AS THE DISPOSITION TO DESIRE A PARTICULAR GOOD OR GOOD FOR A PARTICULAR OBJECT

I propose that when the expression "He/she loves ..." is employed and followed by some object, then "love" as used in such expressions can be defined as "a disposition toward either (i) desiring a perceived good or desiring union with it, either as an end itself or with it being a source of delight in itself or (ii) desiring good for a particular object for its own sake."⁴ I refer to the first of these dispositions as *unitive love*, and I refer to the second of these dispositions as *contributory love*.

The distinction between love considered under the aspect of desiring union and under the aspect of contributing good is not new. It appears clearly in

4 I believe that this characterization pertains to many uses of the expressions "I love ...," but that such expressions, in the first person, also regularly constitute a specific type of perlocutionary speech act (Austin, *How to Do Things with Words*) wherein the speaker principally intends to evoke some sort of emotional, physical, or reciprocal response from the person to whom the words are addressed. I will not consider in detail such uses of this expression here and will restrict attention to the arguably somewhat simpler third-person expression "He/she loves"

Aquinas, an illuminating exposition is given by Stump, and it likewise appears elsewhere.⁵ In contrast to Stump, however, I propose that love is not desire but rather the disposition toward desiring. This arguably comes closer to the account given by Aquinas, who describes love as the principle of movement toward the object that is loved.⁶ The proposed definition here, however, adds further qualifications. With regard to unitive love, a disposition toward desiring a particular good is not on its own sufficient for love; rather, that good must be desired *either as an end itself or with it being a source of delight in itself*. This qualification will be important when considering means to various ends and whether it can be properly said that the means themselves are loved. With regard to contributory love, a disposition toward desiring the good for someone or something is not sufficient for love; rather, the desiring of good for the beloved object must be *for its own sake* rather than solely for the sake of some subsequent benefit that may accrue to oneself. There may indeed be subsequent benefit, but if that subsequent benefit to oneself is in fact the only end desired and if the desiring the good of the other is entirely a means, then, as discussed below, we would not generally refer to this as love.

Before proceeding with the analysis, I would like to be more precise about the claims I will be defending and about the nature of the analytic definition of love being proposed. As noted above, I define unitive love as “a disposition toward desiring a perceived good or desiring union with it, either as an end itself or with it being a source of delight in itself” and contributory love as “a disposition toward desiring good for a particular object for its own sake.” I will make the case that what is being asserted in any use of the expression “He/she loves . . .” followed by the specification of some object is an instance of unitive love, contributory love, or both. The first claim I thus seek to defend in this paper is that whenever the expression “X loves Y” is used, then it is always the case that at least one of the following two statements is in view:

1. “X has a disposition toward desiring Y, or union with Y, either as an end itself or with it being a source of delight in itself,” or
2. “X has a disposition toward desiring good for Y for its own sake.”

The second claim I will seek to defend is that, conversely, whenever at least one of 1 or 2 is in view, then *at least some people*, in their ordinary speech, would assert “X loves Y.” The second claim is in some sense weaker than the first. The qualification of “at least some people” arises because of cases in which one of

5 Aquinas, *Summa Theologica* (hereafter *ST*) I.II.26.4; Stump, “Love, By All Accounts”; Sidgwick, *The Methods of Ethics*.

6 See Aquinas, *ST* I.II.26.1; alternatively, in the same question, Aquinas describes love as “the first change wrought in the appetite by the appetible object” (*ST* I.II.26.2).

unitive or contributory love is present but the other is absent. I will argue that in all such cases, many people would, in their ordinary speech, still say “X loves Y,” but others may not. Some may refrain because, especially in interpersonal relationships, there is a sense that unitive and contributory love ought both be present. Thus, when one of unitive or contributory love is present and the other absent, although the use of “X loves Y” is common in ordinary language, others may say “that is not *really* love.” This equivocation, or disagreement, over the use of “love” is a feature of our ordinary language about love and is a consequence of the disjunctive nature of how the word “love” is ordinarily used. Expressions such as “some may call that love, but it is not *really* love” arise in part because we often desire some distinction between those cases in which unitive and contributory love are both present and cases in which only one is present.

What is intended with the proposed analytic definition of love is a descriptive characterization of our ordinary language use of expressions of the form “X loves Y,” but one that also recognizes and respects what is sometimes equivocation over these statements. In summary, the proposed descriptive characterization is thus that the nonexclusive disjunction of 1 and 2 is a necessary condition for “X loves Y” in ordinary language and a sufficient condition for at least some people being willing to assert “X loves Y.”⁷

The disjunctive nature of the proposed analytic definition will be important throughout the discussion. The characteristics of unitive and contributory love are in many ways distinct and thus arguably best analyzed separately. In what follows, I will argue that many of the paradoxes about love, and disputes over objects of love, can potentially be resolved when we realize that certain aspects or properties of love pertain either only to contributory love or only to unitive love.

2. CHARACTERISTICS OF UNITIVE LOVE

We tend to use love for a diverse range of objects. We love our children. We love desserts. We love our country. We love justice. We can love people, pets, things, places, communities, activities, ideals, the divine, and so on.⁸ For many of these loves, we desire the objects themselves or some form of union with them, and this was defined above as unitive love, a disposition toward desiring a perceived good or desiring union with it, either as an end itself or with it being a source

7 Velleman (“Love as a Moral Emotion,” 353) and Setiya (“Love and the Value of a Life”) effectively argue against the position that desiring to be with the other and desiring the other’s good are *necessary* conditions for love. However, under the proposed disjunctive definition, neither is necessary; rather, it is their disjunction that is proposed as necessary.

8 Hacker, *The Passions*.

of delight in itself. It is in this unitive sense that we might love ice cream. We are not trying to contribute toward the ice cream's good. We simply desire it.

However, simply desiring something is not sufficient for using the word "love" in ordinary language; that desire must have some consistency to it. We must have a disposition toward desiring it.⁹ If someone who normally feels ambivalence toward ice cream suddenly desires ice cream on a particularly hot day, we would not ordinarily say that that person loves ice cream. It is only concerning someone who consistently desires ice cream—who has a disposition toward desiring it—for whom we would say that the person loves ice cream. That consistency for desiring the object does not need to be exceptionless to qualify as love; if someone does not feel like having ice cream on a particular day, that does not disqualify him from loving ice cream. Rather, once again, a disposition toward desiring the object must be present. Similar issues pertain to activities. We might say someone loves skiing if that person has a disposition toward desiring to ski. For someone who has a desire to ski in order to "just try it once," we would not ordinarily say "she loves skiing." Trying it once, and enjoying it, may create a disposition toward desiring it and the love might ensue shortly thereafter, but the disposition toward desiring is necessary. There must be some familiarity with the object to be able to say that such a disposition is present.

Love, as defined above in its unitive sense, is arguably also applicable not just to things and activities but also to people, pets, places, communities, or even ideals. We may say we love a person (or pet or community) if we have a disposition toward desiring to be united with (i.e., to be with) that person (or pet or community). We may say we love a place if we have a disposition toward desiring to be united with (i.e., to be in) that place. We might say "he loves justice" if the person has a disposition toward desiring justice.¹⁰ In all of these cases, there is a desire that one's life be somehow united with some perceived good. In ordinary

9 "Disposition" here is being used in the sense of a tendency or proneness (cf. Hacker, *Human Nature*, 118–21; Hacker, *The Passions*, 302–3).

10 One might attempt to unify all the phenomena that relate to the word "love" with the notion of "a disposition toward desiring the good for something (either for oneself or for something else)." However, the phenomenon of love for an ideal renders this difficult. When expressions such as "love of justice" or "love of beauty" or "love of truth" are used, that for which one is desiring the good is less clear. This could perhaps coherently be rendered as good for one's community, for oneself, or for all, but it is not clear that it is the good of the community or oneself that is principally in view with expressions like "love of justice" or "love of truth." However, love of an ideal seems still to descriptively fit the proposed definition of unitive love as desiring the object itself; and it is the proposed definition of unitive, rather than contributory, love that seems to correspond to love of an ideal. With the love of justice (or truth, beauty, etc.), it is not the case that we desire good *for* justice; we desire justice (and may contribute to bringing it about since, indeed, our desires tend to prompt action), but it would be odd to say that one desires "good for justice."

language, we sometimes speak of desiring the perceived good itself, and other times only of desiring some form of union with it. We would usually speak of desiring to *be with* a pet, a person, or a community or to *be in* a place—that is, to be united with it in some way.¹¹ In other cases, when speaking of love of an activity or of an ideal, we speak of desiring the good itself, not of being united with it or of being with it. We say “I desire skiing” or “I desire to ski” but not “I desire to be with (or to be united with) skiing.” However, again, in all of these cases, there is a desire that one’s life be somehow united with the perceived good.

In instances of unitive love for people or pets or communities or places, we might also have a disposition toward desiring their good (i.e., contributory love), but we might not. Unitive love may be present without contributory love. Someone may love a particular beach, and want to be there, without necessarily having any desire to make it better or contribute toward its good.

In the proposed definition for unitive love, a disposition toward desiring a perceived good or desiring union with it is not, on its own, considered sufficient to be called love. Rather, the perceived good, or union with it, is to be desired *either as an end itself or with it being a source of delight in itself*. In many cases, the perceived good that is desired will be both an end in itself and a source of delight in itself. Such may be the case in one’s desire to be with one’s children, or to ski, to be in one’s hometown, or to have justice. In other instances considered below, the good that is desired may be desired either only as an end or only as a source of delight, but not both. However, in certain cases, one may have a disposition toward desiring some perceived good but desire it only as a means, not as an end and not as a source of delight, and it is not clear we would then always refer to love for that object. Someone, for example, may have a disposition toward desiring exercise not because it brings any delight—she may in fact find it unpleasant and painful—but because, without exercise, she quickly experiences depression. Exercise is sought purely as a means; it is not an end, and it does not in itself bring delight. In such cases, we would not usually say “she loves exercise,” even though she has the disposition toward desiring it.

However, in other cases, that which is desired purely as a means may nevertheless be a source of delight. I may take a medicine that tastes bad and has

11 We thus say “I desire to be with my baseball team” but not typically “I desire my baseball team.” Sometimes we can speak of desiring the thing or of desiring to be united with the thing, with the same meaning in view. We might say either “I desire ice cream” or “I desire to be united with (i.e., to eat) ice cream.” However, in other cases, we might appropriately speak of either desire for the object or desire to be united with the object, but the meaning may differ in these two uses. “He desires to be with her” indicates as its object a union of presence; “he desires her” has sexual connotations. Likewise, “he desires to be in Chicago” indicates a union of presence; “he desires Chicago” suggests some form of conquering or authority over it.

some unpleasant side effects but very quickly heals me. I may marvel at its efficacy. I may do so to the extent that I take a genuine delight in it. I may exclaim "I love this medicine." While the medicine is a means, it is also a source of delight in itself. In the case of exercise considered above, if a different person finds exercise similarly unpleasant, and exercises only to alleviate depression, yet also marvels at how well it works in alleviating depression, she might then say "I do love exercise, not so much for its own sake, but because of what it does for me." The difference between these cases is essentially in the subjective attitude the person takes to the means of exercise, viewing it only as a necessary evil for the purpose of some good (in the first case) or, alternatively, taking delight in the efficacy of it as a means (in the second). We usually would not speak of "love of exercise" in the first case, but we might in the second. Nevertheless, both with the medicine and with the second case of exercise, the good for which the person has a disposition to desire is only a means; it is not an end.

Likewise, there are arguably cases in which one might have a disposition toward desiring some end in itself, even when it does not bring notable delight in itself, and then we might arguably still properly speak of love. A parent with a difficult child may desire to spend time with the child, potentially independent of any good that she hopes to contribute, simply because it is *her* child. The child may behave dreadfully and may bring a great deal of sorrow, and the parent may even come to expect this. However, the parent may still have the disposition toward desiring to be with the child, as an end in itself. We would generally still say in such circumstances that "the parent loves the child." This arguably extends beyond contributory love: the parent wants to be with the child even beyond anything that she might contribute to the child's well-being. As another example, a man may live in a country that is at war, on the brink of devastation. There is nothing further the man can do to help his country. The man may nevertheless have a desire to remain in his country, even as it falls, simply because it is *his* country. There is little or no delight that the man is taking in his country, but yet he has a disposition toward desiring to be united with his country, as an end in itself. It is arguably proper then to speak of "love of country" because of the man's disposition to remain in it, even though this is not a source of delight.¹²

12 Even in these extreme examples, the attainment of any good, perceived of as an end, might be argued to bring at least a modicum of delight simply by the attaining of that desired end, even if this delight is minimal in comparison to the sorrow accompanying the attaining of that end. Strictly speaking, it may be possible to slightly simplify the proposed definition of unitive love to "a disposition toward desiring a perceived good, or desiring union with it, as a source of delight in itself," but again it would be somewhat odd in these extreme cases just considered to speak of the person's love for their country, or their love for their child, principally as a source of delight. Language concerning the beloved objects being sought as ends in themselves is arguably more appropriate.

The requirement that the beloved object is either an end itself or a source of delight in itself excludes certain pathological desires from being love. A kleptomaniac who derives no joy from the actual taking of trivial possessions from others would not generally be said to love these trivial objects, even though he has a disposition toward acquiring them. However, the requirement that the beloved object is either an end itself or a source of delight in itself does not necessarily exclude certain addictions from being loves. A person may be addicted to drinking liquor, see the damage the addiction is causing and desire to stop, and yet be unable, but nevertheless find some delight in the drinking. It would then arguably still be appropriate to speak of the person's love of liquor, because there is a disposition toward desiring it, and the object is a source of, at least some, delight. In such cases, we might be inclined to say "he wants to quit, but he does love drinking." This love is not incompatible with also wanting to refrain from alcohol, or perhaps even also simultaneously hating the beloved object.¹³ However, so long as the disposition toward desiring the object with its being a source of delight in itself persists, it is arguably appropriate to speak of love. Once again, when someone has a disposition toward desiring a perceived good or desiring union with it, either as an end itself or with it being a source of delight in itself, then we would typically say that that person loves that object or perceived good. Let us turn now to contributory love.

3. CHARACTERISTICS OF CONTRIBUTORY LOVE

Contributory love was defined above as "a disposition toward desiring good for a particular object for its own sake." We most characteristically think of loves of this type for persons. The mother loves her child. The husband loves his wife. The friends love one another. However, the objects of such contributory love extend beyond persons. One might have a disposition toward desiring good for

13 The possibility of simultaneously loving and hating some object may arise because one has both the disposition toward desiring it and the disposition toward avoiding it (under some other aspect or for some other reason), as in the case of addiction. Similarly, one might find the process of learning Spanish particularly difficult, but in spite of the hard work, even frustration, over the memorization and learning, one comes back again and again to desiring to learn it—if one has a disposition toward desiring to learn it and if one cannot escape the desire—then one might say, albeit perhaps in a somewhat resigned manner, "Well, I guess I do love learning Spanish." Again, love and hate of learning Spanish may be simultaneously present. However, it is also possible that love and hate may be said to be simultaneously present for some object if only one of unitive or contributory love is present and the other absent; one may have a disposition toward desiring good for someone, whilst being physically repulsed by their presence; or alternatively, one might be obsessed with another person physically and sexually and nevertheless be disposed to wish that person evil.

a place—for example, one’s hometown, community group, or school. We might then speak of “love” also in this contributory sense. Such contributory love may be present with or without unitive love. Someone may have a disposition toward desiring good for one’s hometown, for its own sake, without any sort of disposition toward desiring to return to it. Contributory and unitive love may both be present together, but one may be present without the other. As with unitive love, desire alone is not sufficient for contributory love. For desire to constitute love, that desire must be sufficiently consistent that a disposition toward desiring the good for the object can be said to be present. A sudden feeling of goodwill, or desire for the good of another, does not qualify as love. There must be some familiarity with the object for it to be the case that the person has a disposition toward desiring its good.

With contributory love, aspects of one’s own well-being, including for instance the satisfaction of one’s desires, depend on the well-being of the object under consideration.¹⁴ That one’s own well-being depends on the well-being of the other is effectively a necessary condition for contributory love; it follows directly from the definition of contributory love as a disposition toward desiring the good of the other. One might be tempted to take this as a sufficient condition for, or as a characterization of, contributory love, and indeed some have.¹⁵ We might think that one’s well-being depending on the well-being of another must be love. However, there are arguably exceptions to such dependence being a sufficient condition for contributory love—for example, when one’s own well-being depends on another’s well-being only instrumentally. Someone in a prison camp may be told that his life depends on keeping the others in the camp healthy. The person may not genuinely care about the others, but because his own life depends on the health of the others, he very quickly forms a strong disposition toward desiring their good. However, if the person only cares about the health of the others because his own life is at stake, we generally would not say that he loves the others in the prison camp. To handle such cases, it was proposed above that contributory love be defined as “a disposition toward desiring good for a particular object *for its own sake*.” We would only

14 So-called desire satisfaction theories of well-being effectively equate well-being with the satisfaction of one’s desires (or the desires one would have if fully informed; cf. Heathwood, “Desire-Fulfillment Theory”); however, one need not embrace the *equating* of well-being with desire satisfaction to take the position that well-being depends *in part* on the satisfaction of one’s desires.

15 Sullivan, in considering definitions of love, writes: “When the satisfaction or the security of another person becomes as significant to one as one’s own satisfaction or security, then the state of love exists. So far as I know, under no other circumstances is a state of love present, regardless of the popular usage of the word” (“Conceptions of Modern Psychiatry,” 20).

speak of love for another (in the contributory sense) if there is a disposition toward desiring the other's good for their own sake. This does not mean that no further benefit accrues to, or is desired by, the lover, only that at least one of the reasons for desiring the good of the other is for the other's sake. In the case of the prison camp, the person whose life depends on the health of the others in the camp might eventually come to desire their well-being for their own sake. That might coexist with his own well-being being a reason for desiring their health, and if the desire for their health for their own sake is sufficiently consistent that we could reasonably say that he has a disposition toward desiring their good for their own sake, then we might again properly speak of love.

Contributory love was defined above as a disposition toward *desiring* the good of the beloved object, rather than a disposition toward *acting* to bring about the good of the beloved or toward actually acting to bring about the good of the beloved. In using the term "contributory love," it is not necessarily the lover that is contributing the good; it is only that the lover desires that there be some good contributed to the beloved, but the source of that good may be the lover, another person, or even the beloved's own actions. Of course, in most cases a disposition toward *desiring* the good of the beloved object will bring about a disposition toward *acting* to bring about the good for the object and also actual actions to do so; these will all typically go hand in hand. It can thus be difficult to discern the characterizing features of love. However, there are arguably cases in which the disposition toward acting is present, but the disposition toward desiring is not, or vice versa.

If the disposition toward acting to bring about good is present, but without a disposition toward desiring that good, it is not clear that we would refer to this as love. Consider a physician who consistently acts to contribute to the health of his patients and is disposed to do so because this is his job. With regard to such circumstances, Hacker comments, "good doctors do not love their patients and conscientious shepherds do not love their sheep, although they care for them and protect them."¹⁶ If the doctor considers providing treatment for patients merely as his duty, as fulfilling the role for which he is paid, then indeed the doctor would not generally be said to "love his patients." Likewise, a shepherd who protects the sheep and keeps them alive simply because this is his job would generally not be said to love the sheep. In both cases, on account of their respective roles, there is a disposition toward *acting* to bring about the good of the patients and the sheep respectively, and yet we would generally not speak of loving patients or sheep. On the other hand, a doctor may, over time, come to know a number of his patients more deeply; his interest in their

¹⁶ Hacker, *The Passions*, 302.

well-being may increase. He may develop a disposition toward desiring the good of those patients for their own sake, and if that were the case, then we might properly speak of his loving the patients.¹⁷ A similar phenomenon might even be thought possible for the sheep. But it is not just a disposition toward acting to bring about the good of the object that qualifies as love but rather the disposition toward desiring the good of the beloved object for its own sake.¹⁸ Neither the disposition toward acting nor even actual action to contribute to the good of the other is a sufficient condition for love.

Moreover, actual actions to bring about the good of the other are not only not a sufficient condition for love but also arguably not even a necessary condition. A man may love his wife in the contributory sense of having a disposition toward desiring her good for her own sake but may not be able to act. One might imagine the husband being imprisoned in jail in a foreign country, with a life sentence and with no mode of communication with his wife being possible; he may also know that his wife has only three months to live because of a terminal cancer. In such a state, he is unable to directly act to contribute to her good and will not be able to do so during the remainder of her life.¹⁹ Yet we would likely still say that he loves her, on account of his continued disposition toward desiring her good for her own sake, even though he may be entirely unable to act on those desires.

A disposition toward desiring the good of the beloved object will in general result in a disposition toward acting to bring about the good of the object but may not always. Examples are conceivable in which someone genuinely has a disposition toward desiring the good of another but does not have the disposition toward acting to achieve that good because of, for example, lethargy, sloth,

17 Moreover, if this were the doctor's general pattern of interacting with patients, we might even be willing to say that the doctor loves not just those particular patients but even patients generally, as a class. The particular object of contributory love may be the specific individual patient, or in some cases, for some doctors, it might be the class of patients. Contributory love as a disposition toward desiring the good of the object for its own sake does arguably require familiarity, but if the object is taken as the class of all patients, the doctor might be properly said to have familiarity with this class and to have a disposition toward desiring patients' good for their own sake.

18 Hacker in fact uses the examples of the doctor and shepherd to argue against the idea of love being a disposition, but his examples are sufficient only to show that a disposition toward *acting* does not constitute love (*The Passions*); the examples are not in tension with the characterization of love as a particular disposition toward *desiring* the good of the other.

19 It is not entirely clear whether it would be appropriate to say that the husband has a disposition toward acting to bring about the good of his wife. Saying that he has a disposition toward acting might still be considered to be appropriate insofar as in the counterfactual scenario in which he were not imprisoned, he would act according to his desire for his wife's good. However, clearly, in actuality, he cannot act.

lack of knowledge, inattentiveness, or some deficiency of character, or simply because the disposition toward desiring the good of the other is weak. These considerations then bring us to questions of the strength of love and of loving well to which we will now turn, along with other potentially problematic or paradoxical examples concerning love.

4. THE STRENGTH AND INTENSITY OF LOVE

A disposition toward desiring the good of another for his or her own sake conceivably might not result in a disposition toward acting, possibly because the disposition of desiring is simply too weak. It may be dominated by other dispositions and motivations; action may not result, nor even any notable disposition toward action. The issue here then pertains to the strength or intensity of love.

If the account of love as a type of disposition toward desiring is correct, then the potency of love might be considered both with respect to the strength of the disposition in terms of the frequency of its operation and with respect to the intensity of desire that arises from it. A love might be considered weak either with respect to the disposition not being strong or with respect to the desire that results from the disposition not being intense. In some cases, the desire may be intense but the disposition not regularly operative: a woman may have an intense desire to ski each year upon the first snowfall, but once that desire is satisfied, it may not arise again until the following year. In other cases, the disposition may be strong insofar as it is frequently operative, even though the desire it generates is not especially intense: a man might consistently desire to have a sparkling water every day following his coffee; the desire for the water may not be especially intense, but it is always present; the disposition is strong even though the desire is not intense. The strength of the disposition in terms of frequency of operation might be considered both with regard to its numerical recurrence and with respect to the circumstances under which the disposition is operative. If a slight alteration of circumstances renders the otherwise regular disposition inoperative, then we would not generally speak of a strong love. In contrast, if the disposition toward desiring is operative even in the face of various obstacles and costs, we might speak of a strong, or, in the contributory case, perhaps even sacrificial, love.

We are arguably most likely to speak of loves as “loves,” rather than mere “likings,” when at least either the disposition is strong or the corresponding desire is intense, or both. A love that is both intense in the desire it generates and strong in terms of the frequency of its being felt may often occupy an important place in one’s life. Hacker thus also speaks of perhaps yet another dimension of love’s potency—the “depth of love,” which he describes as “the

manner in which it penetrates manifold levels of the personality of the lover, by its grip on the imagination, by the weight of the reasons for thought and action with which it furnishes the lover, by the extent to which it motivates behavior, and by its giving meaning to one's life."²⁰ Certainly, the intensity and strength will have an effect on love's depth, thus described. But the depth of love is also related to the fact that love, and the relationship of the lover to the beloved object, in some sense has a history—a beginning, development, present state, and future expectations.²¹ It is possible for the present state of a love to be not especially intense in feeling, or not frequently felt as a disposition, and yet still have depth because of the history of love within that relationship, as might occur in long-standing friendships, in marriage, or with family relationships.

However, returning to the considerations above, the issues of the strength and intensity of love can help us make better sense of cases in which the disposition toward desiring may be present but the disposition toward acting absent. In most circumstances, a strong, intense love will result in a disposition toward acting and in actual action. Thus, many of our judgments concerning the love of one person for another are based on actions. In much of life, it is easier to assess the actions of another than it is to assess their dispositions toward desiring. In most contexts, this is reasonable because a strong, intense love will typically result in action. However, again, this may not necessarily be the case, as a result of some physical impossibility, the love not being very strong, or the love being dominated by other loves and motivations. It might also not result in action, or at least not in actions that achieve the good of the other, because of lack of understanding as to how to act or as to what constitutes the good of the other, or out of mere inattentiveness or lethargy. This brings us to considerations of loving well.

5. LOVING WELL

Conceiving of love as a disposition toward desiring union with the beloved or toward desiring good for the beloved can give rise to what, on the face of it, may seem like paradoxical situations. Someone may desire the good of another

20 Hacker, *The Passions*, 304. He describes the "intensity" of love as that which is "exhibited in essentially associated emotions; the intensity of joy at reciprocated love, the intensity of the anxiety felt when the person one loves is endangered, the intensity of the delight in joint activities and successful projects, the intensity of the longing for the person one loves when absent and of the joy when reunited, in the grief of bereavement, and so forth" (304). But presumably the *intensity* of all of these essentially associated emotions arises principally from the intensity of the desire itself to be united with the beloved, or to contribute to the beloved's good, or both.

21 Kolodny, "Love as Valuing a Relationship"; Hacker, *The Passions*.

but, because of a lack of understanding, act in such a way so as not to attain that good, or even so as to cause harm. Alternatively, someone might desire the good of another and act so as to attain it, but the beloved may understand the good very differently and may not see the actions as loving. Are such examples then truly instances of love?

Concerning the first situation, an overly indulgent mother might genuinely desire what is good for her son and believe she is providing what is good for him by her actions. However, by giving in to the son's every whim, she may in fact be causing him harm. Her always giving in may prevent him from developing patience, fortitude, and self-control. Her continually providing him with sweets may cause harm to his health. Does she then love her son? In the definition of love proposed above, one would say yes, she does love him (even in the contributory sense) if the mother desires the son's good for his own sake. However, while the mother desires the son's good and genuinely believes that her actions are contributing to it, she is mistaken as to what contributes to his good. We might then say that she loves her son, but she is not loving him *well*. Her beliefs and actions are such that although she desires his good, she does not in fact contribute toward it. In this example, the mother's beliefs about what actually constitutes the son's good are in some sense deficient. As a variation of this example, one might alternatively imagine that the mother in fact understands what is good for her son, realizes it is best not to give in to his desires, and understands that excessive sweets may harm his health, but when he begs for them, she nevertheless, in spite of knowing better, gives in. She may thus genuinely desire the good of her son, and even know what constitutes that good, but because of weakness of character or lack of fortitude, she does not act in such a way to attain the desired end. Once again, we might say that she loves her son, but she is not loving well. In each of these cases, while the disposition constituting love is present, some deficiency in knowledge or character prevents the mother from appropriately acting to achieve the goal of her desires—her son's good.

We may thus say that love is indeed constituted by a particular disposition toward desiring the beloved or union with the beloved, or desiring good for the beloved, but that loving *well* is constituted by having knowledge, strength of character, and, whenever possible, actions so as to in fact render the attaining of good for the beloved, or union with the beloved, more likely. It would follow from this that stronger or more intense interpersonal loves will characteristically also entail the desire to love well, since this may be what is needed to actually attain the desired good.

Consider now one further related example in which the lover genuinely desires the good of the beloved and acts rightly so as to attain it, but the beloved does not understand his own good in a similar way. It is possible that the lover's

idea of what is good is, in some sense, correct, but if the beloved does not view the lover's action as loving, as seeking his own good, is this indeed still love? We might now imagine a mother who does genuinely want to develop her son's character and to provide him with a healthy diet, and so acts accordingly and resists her son's demands and desires for excessive sweets. We might imagine her son being very upset at this, persisting in his desires and perhaps feeling unloved because his mother does not give in to them. Here, we would again say that the mother loves the son because she consistently desires his good. We might even say that she is, for the most part, loving him well because she is resisting his demands for what is not good for him. She might even feel torn about giving him sweets, desiring to please him and wanting his love, but desiring also his ultimate good.²² However, if as a result of her actions or the way in which she goes about resisting, the son feels unloved, then we might be hesitant to say that she is loving well.

Human experience clearly indicates that one of the strongest and most universal human desires is the desire to feel loved. Feeling loved—both feeling desired by another and feeling that the other is seeking one's own good—is a part of one's own well-being; it affirms one's value and dignity.²³ Thus, if someone is genuinely seeking the good of another, this will include seeking to make the person *feel* loved—seeking to make the other feel that one desires the good of, and desires to be with, that person. Because feeling loved is a part of the good of the beloved, contributory love for the beloved will in general entail trying to communicate that love to the beloved. Loving well will, to the extent possible, involve trying to effectively communicate to the beloved that desire to contribute to the good of the beloved and to be with the beloved. In the case of the mother, it will involve not only resisting her son's inappropriate demands but also explaining why she is doing so, and that she loves him and wants what is good for him, and that this is the reason for her acting as she does. Such communication may not be effective, but contributory love will, if the mother is loving well, attempt to have the son feel that he is loved, even while resisting his desires.

If someone has a genuine, consistent desire for the good of the beloved, such love will also not be entirely inattentive to the beloved's desires being unmet. The lover who desires the good of the beloved will realize that unfulfilled desire is partially constitutive of some lack of good on the part of the

22 In more extreme cases, contributory and unitive love may be opposed. Someone may desire to be with the beloved as a source of delight but also know that in the circumstances, this is bad for the beloved, so that to contribute to their good one ought to depart.

23 See Velleman, "Love as a Moral Emotion," and Pieper, *About Love*, for an account of how someone's being and feeling loved can give rise to this sense of their own value and worth or an affirmation of the goodness of their being.

beloved, even if circumstances are such that it is indeed genuinely best for those desires to remain unfulfilled. The loving mother will not give in to her son's demands but will nevertheless be attentive to the fact that his desires remain unsatisfied. Loving well—the desiring and rightly understanding and acting so as to achieve the good of the other—will not involve giving in to those demands and desires, but it may involve trying to change them. The seeking the good of the beloved may require trying to change the son's perceptions and desires so that they also are genuinely for what is good.

6. THE UNION OF UNITIVE AND CONTRIBUTORY LOVE AND THE RECIPROCITY OF LOVE

Especially with human relations, unitive and contributory love often occur together. Indeed, this is generally what we desire in our relationships. It is arguably what underlies friendship. It is what most hope for in marriage. However, the coming together of unitive and contributory forms of love can occur with objects other than persons. It may occur for a country in which one desires both to remain within one's country and to contribute to its good. It may occur with a work project concerning which one desires both to spend time on the project and to advance it toward its goal.

The disjunctive nature of the proposed analytic definition of "love" may also explain occasional disputes over what constitutes love in interpersonal relationships or what the essence of love truly is. As noted above, we sometimes hear statements along the lines of "we may call that 'love,' but it is not really love" or that "it is not true love." This can arise both with unitive and contributory forms of love. It might be said of a young man with strong attraction to a woman, "he desires to be with her, but he does nothing to help her; that is not true love."²⁴ Unitive love may be present, but contributory love mostly lacking. However, the reverse may be the case. In a context in which relations between husband and wife have grown cold, it might also be said, "she wants to help him, but she doesn't really even like being with him any longer; that is not real love."²⁵ Here, contributory love might be present, with unitive love weak. What may

24 A more extreme example might be constituted by attempts at purely sexual relationships. Each person may have a disposition to desire sexual union with the other as a source of delight. Someone might even say "I *love* her," perhaps intending nothing more than "loving to have sexual intercourse with her," but this may again be met with the response "that is not really love."

25 The doctor considered in section 3, who had developed a disposition to desire the good of a patient for her own sake, might be analogous. We might hesitate to say that he loved the patient if the desire to be with the patient was entirely absent.

be in view in these cases is the fact that either unitive or contributory love is absent. In ordinary language, we tend to use the word “love” when at least one of unitive love or contributory love is present, but we may not make reference to “true love.”²⁶ With respect to interpersonal relationships, there is often a sense that the two *ought* to go together or that the deepest forms of love, or real love, involve both. There can thus be some ambivalence, especially when the question is explicitly raised, as to whether it is appropriate to call a particular disposition “love” when only one of unitive or contributory love is present. In such cases, some may refer to this as love in their ordinary speech, and others may not.

The possibility that only one, versus both, of unitive and contributory love is present effectively creates space for dispute as to whether it is appropriate to say “X loves Y.” Within the context of interpersonal relationships, some may be willing to make this assertion, and others may not because either unitive or contributory love is lacking.²⁷ There can thus be disagreement over the use of the word even when there is agreement on the facts, and expressions such as “that is not really love” perhaps especially arise in such contexts. The claim made in this paper is thus that the disjunction of unitive and contributory love is a necessary condition for our ordinary language use of “X loves Y,” but this disjunction is only a sufficient condition for it being the case that *some* people will, in their ordinary speech, assert that “X loves Y.” Given that we use “X loves Y” disjunctively for many objects of love but often expect unitive and contributory love to accompany one another in interpersonal relationships, we perhaps then ought to expect some ambiguity over the appropriateness of “love” when the object of love is a person. This equivocation is a feature of our language concerning love, which makes it yet more difficult to analyze. However, when we appreciate the disjunctive nature of many of our uses of “love,” we can see also that this very equivocation helps us understand how the word “love” is used: ordinary language expressions such as

26 See Earp, Do, and Knobe, “The Ordinary Concept of True Love,” for experimental evidence that when “true love” is used in English it pertains either to love being especially good or valuable or to the reality of certain psychological states, rather than to love that is highly prototypical. The analysis of love as valuing a relationship given by Kolodny might be understood as an account of the conditions under which we would typically describe love truly as *real* interpersonal love—when we might say, “X really does love Y” (“Love as Valuing a Relationship,” 151). He does, however, also clearly recognize ordinary language uses of “love” outside those characterized by his account (137).

27 While the expectation that unitive and contributory love ought both be present is perhaps especially strong when the object of love is a person, that expectation can arise with other objects of love as well. Even if someone consistently desires good for his town of origin, we might hesitate to say that the person loves his hometown if he never desires to return to it. Some may be willing to speak of the person’s “love for his hometown”; others may say “that is not really love.”

“that is not *really* love” helpfully allow us to distinguish between cases in which just one, versus both, of unitive and contributory love is present.

The expectation that unitive and contributory love will often accompany one another is, however, not merely a linguistic phenomenon; it is arguably grounded in the very nature of these various aspects of love. Some of the reason both unitive and contributory love often seem to accompany one another may be because of the effects that each can have on the other. In contributory love, by desiring and seeking the good of the other, we may more closely engage with the other, and it is then also possible that our affections become more attached to the other and that we then more consistently desire to also be with the beloved. In unitive love, in seeking to be with the other and in actually being with the other, we see opportunities to contribute to their good, our affections may become more attached, and we may more strongly desire their good, giving rise to a stronger contributory love as well.

Both unitive and contributory love also have the capacity to evoke reciprocity from the beloved. Contributory love can give rise to consistent desire of the beloved to be with the lover, since the lover has become a source of good for the beloved. Unitive love can likewise sometimes give rise to consistent desire of the beloved to be with the lover, insofar as the lover’s unitive love may lead to the beloved feeling valued and thus wanting to further experience this sense of worth. This may also subsequently give rise to the beloved’s having contributory love toward the lover, possibly in part to demonstrate a sense of worthiness to the lover, so as to be further desired by the lover. The lover’s contributory love will often likewise give rise to contributory love on the part of the beloved, a phenomenon closely related to what is sometimes referred to as “reciprocal altruism.”²⁸ The relative weights of the various explanations for such reciprocal altruism are disputed, but the phenomenon itself is well documented.

Certainly, there are exceptions to the phenomenon of contributory love giving rise to unitive love, to unitive giving rise to contributory, or to unitive or contributory love evoking reciprocity. Contributing toward another’s good may evoke reciprocity, but it can sometimes evoke resentment; such resentment can, moreover, result in the lover’s desiring to withdraw. Unreciprocated unitive love can sometimes result in a sense of revulsion or disgust of the beloved toward the lover. This can occasionally result in a redoubling of the lover’s efforts but can also, especially if the repulsion persists over time, result in the lover ceasing to seek the beloved’s good or presence. While contributory and unitive love often evoke reciprocity, and often one leads to the other, these phenomena clearly do not always play out in this manner.

28 Trivers, “The Evolution of Reciprocal Altruism.”

However, when unitive and contributory forms of love are both present, the effects both on the lover and on the beloved can be powerful, and yet more so when they are reciprocated. Unitive and contributory love together will then involve seeking both to be with and to benefit the other. The ensuing actions to contribute to the good of the other may further reinforce the disposition to desire the other's well-being. The seeking the good of the other will strengthen the affection underlying the disposition to desire the other. That one's own well-being depends on the well-being of the other strengthens the effect yet further. The seeking to be with the other, and the being in the presence of the other, may further motivate seeking the good of the other and may provide additional insight as to how to do so. The desire for union with the other may prompt a lover's desire for reciprocity, to have the beloved also desire union with the lover.²⁹ This may further prompt a desire to contribute to the beloved's well-being in hope of bringing about that reciprocity. The beloved's knowing this desire and feeling loved, the lover's seeing the beloved's delight at receiving some good, or the confirmation that the beloved recognizes that such contributions were intended as an expression of love will strengthen both forms of love yet further. These experiences can powerfully affect the lover.

When the beloved similarly has both contributory and unitive love toward the lover, the same dynamic again comes into play, potentially creating a powerful interplay of reciprocity, a virtuous cycle of love. Moreover, when both unitive and contributory love are mutual, the very presence of the lover contributes to the good of the beloved insofar as this is what the beloved desires: because of the beloved's desire to be united, the lover's unitive love itself can also become a contributory form of love. The result of the reciprocity of both unitive and contributory love—the merging of the two and the feedback and interplay of reciprocity—can of course be emotionally powerful, potentially culminating in a sense of ecstasy. When the intensity of desire is heightened yet further in romantic love—with unitive love taking on a sexual dimension, along with the powerful physiological and emotional forces that accompany this—being entirely overwhelmed by, and with, the beloved is not uncommon. The power of the union of unitive and contributory love, especially when it is reciprocal, can feel euphoric and enrapturing, as documented throughout art, literature, poetry, drama, film, and personal narratives. Perhaps because of this potential for reciprocal contributory and unitive love together to have such strength, intensity, and depth, there is sometimes hesitation to use “love” for

29 Interpersonal unitive love will in general entail a disposition toward desiring a reciprocation of love from the beloved in part because the union that is principally desired is more likely to be attained if the beloved also desires it and in part because the desired union may itself be constituted by a mutual desiring to be with the other—a union of affections.

any of the various less powerful loves and sometimes hesitation to utter the words “I love you,” both as concerns whether the use of the word is appropriate and as concerns what may ensue if these words are spoken to another.

7. SPECIFIC OBJECTS OF LOVE

In this section, I would like to briefly address various types of love or objects of love that may initially seem to fall outside the proposed analytic definition of love as a particular disposition toward desiring (union with) the other or desiring the good of the other. Each of these objects of love unquestionably merits greater attention than will be given here, and my primary goal in this section will be simply to argue that contrary to what may seem at first glance, these other types of love or objects of love do in fact conform to the proposed definition. In what follows, I will briefly consider love of God; love of neighbor, stranger, and enemy; self-love; and love of pets.

In monotheistic religions, considerable emphasis is placed on “love of God.” Love of God, in its unitive sense, would, if rendered according to the proposed definition, be “a disposition toward desiring union with God, either as an end itself or with it being a source of delight in itself.” From the standpoint of these religious traditions (both in general and especially within religious mystical traditions), such union is often construed as the primary end of human life. Love of God, in its contributory sense, would be rendered as “a disposition toward desiring good for God for God’s own sake.” Theologically, this might be seen as somewhat more problematic, but perhaps it is still intelligible with regard to the possibility of bringing about the good on earth that God desires in order to please God. Perhaps the greater challenge that the notion of love of God poses to the proposed analytic definition of love in this paper is the fact that, in these traditions, the love of God is *commanded*. If love is a disposition toward desiring, then how can desire be commanded? Even here, however, I believe a resolution is possible in terms of the theological understandings within the religious traditions themselves. One possible cause of unitive love is simply the apprehension that something constitutes a good for oneself, and thus is in some sense desirable. Such an apprehension is not sufficient for love, because it may be dominated by other desires, or one may become distracted, but such an apprehension can, sometimes at least, be a cause of love. In the case of love of God, in its unitive sense, the command to love might be understood to be fulfillable insofar as it is a command to understand God as the greatest good and therefore something to be desired.³⁰ Moreover, if it is thought that from

30 Cf. Aquinas, *ST II.II.44.1.4*.

God being the highest good it follows that God is worthy that each should contribute to God's good by seeking to carry out on earth what God allegedly desires, then contributory love for God could likewise be commanded.

Another potential cause of love may be a vow, agreement, or covenant. While it may be inappropriate to vow to desire a particular object, it is arguably not inappropriate to vow to be with or to seek the good of an object. One arguably cannot appropriately make vows concerning desires, but one can do so for actions. Moreover, a possible course of action is to seek to alter the dispositions of one's desires. Furthermore, once such a vow is in place, the vow itself can give rise to desire insofar as one desires to fulfill the vow. This does not on its own constitute love if these desires arise only *for the sake of the vow*. However, the vow may create and sustain a desire to be with the other and a desire for the other's good. If eventually that sustained desire gives rise to a disposition toward desiring to be with the other *as its own end* or to contribute to the good of the other *for the other's sake*, then the vow has itself effectively created love. A dynamic of this sort is arguably sometimes operative in marital vows, and later in the next section, I will consider this in greater detail. However, in certain instances it seems that a vow itself might, at times, give rise to love. This notion of vow or covenant was especially important in the Jewish understanding of the law and the love of God and was appropriated also by the Christian tradition.³¹ Such covenants or vows likewise provide another way to understand the command to love God.

The Jewish law, the New Testament, and the Qur'an also speak of, and even command, "love of neighbor," with the Jewish law extending this to "stranger" (Deuteronomy 10:19) and the New Testament even to "enemy" (Matthew 10:44). The context in all of these cases suggests love in a contributory sense of seeking good for one's neighbor or stranger or enemy, which would thus, under the proposed definition, be rendered as a "disposition toward desiring good for one's neighbor [stranger/enemy] for his or her own sake."³² Such contributory love, as a disposition, seems conceptually unproblematic except for the fact that, once again, it is commanded. However, under the exposition Aquinas gives, such contributory love for neighbor follows from and is derivative from one's love

31 Levenson, *The Love of God*.

32 Such love of neighbor may require a different analysis from the preferential love that arises within the context of a relationship (Kolodny, "Love as Valuing a Relationship"). In instances of love of neighbor in which there is no prior relationship, we might be more inclined to say "X is showing love" than "X loves Y" or "X really does love Y" (see footnote 17), where Y denotes the name of a person, though we might well still say "X loves his neighbor" or "X loves the stranger." The object of love with love of neighbor in the analytic characterization proposed here is arguably the person as a member of the class of neighbors.

for God.³³ As framed above, love of God, in its contributory sense, would be the disposition toward desiring to contribute to God's good, where that good is itself understood as bringing about the good on earth that God desires. However, the good that God desires includes good for one's neighbor, for the neighbor's sake. Thus, a disposition toward desiring to contribute to good for God entails a disposition toward desiring good for one's neighbor, for the neighbor's sake. In Aquinas's understanding, love of God and love of neighbor constitute a single precept with the latter effectively encompassed by the former, but with the latter explicitly articulated so as to make that entailment clear.³⁴ Within this understanding, it becomes comprehensible how love of neighbor, conceived of principally as a disposition toward desiring the good of one's neighbor, could be commanded, because it follows from love of God. While love of neighbor or stranger or enemy is arguably to be understood principally as love in its contributory sense, Stump argues that within the Christian tradition, there is also a unitive sense of love for neighbor or enemy, insofar as there ought to be a desire for final union with neighbor and even enemy, in union with God, in the life to come.³⁵ In any case, it seems love of neighbor, stranger, and enemy, and even the command to love, can be given a coherent rendering under the proposed analytic definition.

Let us now briefly consider self-love. The expression "self-love" is sometimes used simply to indicate selfishness or self-centeredness. There is, however, also a more morally neutral form of self-love, which might be understood as love in its contributory sense with oneself as the object of that love. Self-love thus understood would then be rendered as "a disposition toward desiring good for oneself for one's own sake." This disposition of self-love is natural and essentially inescapable. It is arguably this disposition of self-love, thus understood, that provides the sense and force to the command "Love thy neighbor *as thyself*." However, with respect to self-love, when this natural disposition of self-love severely neglects the good of others, we often speak of "selfishness," and when this disposition of self-love mentally dominates someone's life, we sometimes speak of "self-centeredness." However, self-love more generally can be understood simply as contributory love, with oneself as the object. Whether there is also a unitive sense of self-love is somewhat more obscure. It is not, on the face of it, clear what it would mean to desire oneself or union with oneself.³⁶ Accordingly, Hacker notes, "One cannot fall in love with oneself. . . . Nor can one yearn for reciprocity from oneself. . . . One cannot wish to share one's

33 Cf. Aquinas, *ST* II.II.25.1, II.II.44.2.

34 Aquinas, *ST* II.II.44.2.

35 Stump, "Love, By All Accounts."

36 Cf. Stump, "Love, By All Accounts."

experiences with oneself. . . . One cannot miss oneself, long for oneself, or long to be reunited with oneself.”³⁷ However, if the “union” were understood as that of perception, then a coherent understanding might be retained. Unitive love, with oneself as the object, might then be rendered as a “disposition toward desiring the perception of oneself, either as an end itself or with it being a source of delight in itself.” That perception of oneself might be principally physical or more generally how one views oneself as a person. There can certainly be a healthy natural desire to find some satisfaction or pleasure in the view one has of oneself, and we sometimes might refer to that as “self-esteem.” However, when that disposition toward desiring the perception of oneself physically, as a source of delight, becomes extreme, we would often refer to this as “vanity.” When that disposition toward desiring and delighting in the perception of oneself, as a person, becomes excessive, we might sometimes refer to this as “pride.” Thus, while the primary sense of self-love is arguably love in its contributory sense, it seems also that there is a coherent secondary sense of self-love in its unitive form, with this union being understood as self-perception.

Finally, as to objects of love, let us consider love of pets. An owner’s love of a pet, in consistently desiring to be with it or contribute to its good, seems unproblematic. The question of the pet’s love of its owner is perhaps more complex. Hacker claims that while an owner may love a pet, it is not truly proper to speak of a pet loving its owner because “the complexity of the emotion of love involves cognitive, cogitative, mnemonic, and imaginative abilities that are beyond the powers of non-human animals that lack a developed language.”³⁸ Nevertheless, we often do, in ordinary language, also hear speech about a pet’s love for its master. Here the distinction between unitive and contributory love and the disjunctive nature of the proposed definition may help. If unitive love is conceived of as a pet’s disposition toward desiring union with its master (i.e., being with its master), with such union being a source of delight in itself, then, arguably, manifesting such a disposition toward desiring is within the behavioral repertoire of a pet, even without language.³⁹ Hacker grants that “pets may show great affection for their master or mistress, and even pine away.”⁴⁰ However, contributory love, which in this case would be construed as a pet’s disposition toward desiring good for its master *for the master’s own sake*, is arguably outside of a pet’s behavioral repertoire. While a pet may have a disposition toward desiring good for its master, there is nothing within the pet’s behavioral repertoire that

37 Hacker, *The Passions*, 325.

38 Hacker, *The Passions*, 273.

39 This point arguably also pertains to an infant’s love for his or her mother.

40 Hacker, *The Passions*, 273.

allows us to distinguish whether the pet's behavior is done for the master's own sake or to somehow benefit the pet itself. Here, with regard to contributory love, one might agree with Hacker that such contributory love "involves cognitive, cogitative, mnemonic, and imaginative abilities that are beyond the powers of non-human animals that lack a developed language." However, that pets arguably can manifest unitive love may explain why we hear speech about pets loving their masters and not only speech about masters loving their pets.

Love may clearly be manifest in a variety of relationships. However, the nature of these types of relationships and also the cultural background and customs within which they operate can be highly variable. The forms of union that are appropriate between a parent and child, or between friends, or between spouses are all very different. Likewise, the good that one can contribute to a child, parent, friend, spouse, enemy, or stranger may be very different. In light of the diversity of "historically conditioned forms of reaction, expression, and response, as well as historically different motives for different kinds of actions," Hacker argues that, in contrast to other emotions:

The historicity of love is not merely a matter of different objects of love, but different emotions of love. The love of God, for example, is not "just like" the love of a man or woman, but for the fact that what one loves is God, not a human being. The *criteria* for loving God are very different from the criteria for loving a man or woman. . . . The *responses* that characterize these forms of love are altogether different, and so too are the ways in which such love is *expressed* in word and deed. . . . The differences between these social and historical forms of love are altogether unlike the differences in possible objects of fear at different times . . . where one might indeed say that the emotion is the same, the criteria for its ascription are the same, the web of subjective feeling and expression is the same, and only its objects change.⁴¹

While it is certainly the case that how love is *expressed* differs considerably by object, I hope to have shown in the preceding discussion that there is in fact more unity to the criteria for its ascription than Hacker suggests. There is diversity in the objects of love and the ways that love manifests itself, but there is a conceptual unity. With regard to diversity, Stump refers to the various types of relationships as "offices" that can powerfully shape the appropriate form of union and the appropriate good that can be contributed.⁴² Stump notes that

41 Hacker, *The Passions*, 280–81.

42 Stump, "Love, By All Accounts." Cf. Kraut, "Love De Re"; Kolodny, "Love as Valuing a Relationship," 139.

unitive forms of love will sometimes be more responsive to the intrinsic characteristics, and changes therein, of the beloved than contributory love.⁴³ However, the office, or history of the relationship or love, can also provide reasons for contributory love and unitive love to persist.⁴⁴ How each office is understood may differ by culture and by historical period, which further shapes how love in its unitive and contributory forms is experienced and expressed. There is thus diversity in the objects or offices and variability in the cultural and historical diversity as to how those offices are understood, but there is arguably yet still a unity in the criteria for ascribing love, namely, as “a disposition toward either (1) desiring a perceived good or desiring union with it, either as an end itself or with it being a source of delight in itself or (2) desiring good for a particular object for its own sake.”

8. SOME PRACTICAL CONSIDERATIONS

I will conclude this essay with some practical considerations pertaining to and possible implications of the proposed definition of love as it relates to love (i) being a disposition, (ii) toward desiring, and (iii) in one or both of its unitive or contributory forms.

Society, it seems clear, would benefit from promoting greater contributory love. If each person is seeking the good of others, for their own sake, that good will likely more effectively be attained. Of course, most, or all, also desire fulfillment in some form of interpersonal unitive love, and this too can often result in the attainment of some good, though it can likewise result in actions that hinder the good for others and possibly even for oneself and for the beloved. However, if our loves can sometimes help bring about various goods, and if both loves are understood as dispositions toward desiring good of various kinds, this raises the questions of how such dispositions can be fostered and when it is right to do so.

How might love then be facilitated? This question seems more straightforward for unitive love than for contributory love. For unitive love, the mere recognition that the object is good for oneself has some potential to bring

43 In “Love, By All Accounts,” Stump argues that an understanding of love that accommodates both unitive and contributory desires can make sense (a) of the lack of substitutability of individuals with similar intrinsic characteristics, which “responsiveness” or “quality” theories (with love arising from attributes of the beloved; e.g., Jollimore, *Love’s Vision*) have trouble explaining; (b) of there being *reasons* for love that more “volitional” accounts (with love simply being a decision to seek the other’s good; e.g., Frankfurt, *The Reasons of Love*) have trouble with; and (c) of the phenomenon of love at first sight that “relational” accounts (with love arising from a relationship over time; e.g., Kolodny, “Love as Valuing a Relationship”) have trouble with.

44 Cf. Kolodny, “Love as Valuing a Relationship”; Hacker, *The Passions*.

about love, though there may be competition from love for other goods and the resulting love consequently may not be strong.⁴⁵ With contributory love, such love might arise if someone is seen as worthy of contributing good toward, which may come from the nature of the relationship or office, as with children; or possibly from a sense of moral obligation or even a vow; or from the recognition of the inherent human dignity of the person.⁴⁶ These things are not sufficient for love but may give rise to it. These things are not sufficient for love in part because love is not simply a disposition toward action, for which a vow might be sufficient, but a disposition toward desiring. The recognition of human dignity, or of an important relationship, moral obligation, or vow, may give rise to a disposition toward desiring the other's good, but it may not.

We might wonder, however, whether desire matters if moral obligations, a relational office, or human dignity sometimes suffice for action. A disposition toward acting for the other's good could of course itself accomplish a great deal. To see the importance of desiring the other's good, in addition to acting, it may be helpful to consider marriage once again. It is not an uncommon occurrence to hear spouses complain that they have "fallen out of love" or no longer feel the love that they used to. If the experience of love were simply a disposition toward acting for the good of the other or toward being with the other, the aforementioned complaints within marriage might be less frequent. In married life, often the presence of the other becomes routine. Likewise, habits and patterns of behavior—cooking, cleaning, repairing the car, providing an income—are established that make contributing toward the good of the other become routine. Because the spouse's presence and actions occur routinely, desire is no longer strongly felt and the experience of "feeling love" fades. This can be problematic if that feeling of love *for* the other and *from* the other is itself desired.

45 Thus, in Aquinas, the cause of love is the good (*ST* I.II.27.1).

46 Velleman's exposition of love might be understood as an account of the proper grounds for love, which is an appreciation of human dignity and personhood in its universal essential aspects. However, what that love constitutes is arguably, as proposed here, a particular disposition toward desiring to be united with the other or desiring the other's good. Proper appreciation of the other's human dignity might be viewed as sufficient grounds for both dispositions (Stump, "Love, By All Accounts"; Velleman, "Love as Moral Emotion"). Velleman, however, acknowledges that both because of our limited capacity to adequately appreciate the personhood of others and because of the practical constraints of time, we end up loving certain individuals and not others, and that this love might thus arise from a variety of other particular—perhaps sometimes even incidental—reasons (Jollimore, *Love's Vision*) or because of the office or relationship (Kolodny, "Love as Valuing a Relationship"). However, such things can also in turn give rise to someone's capacity to properly appreciate someone else's value as a person (Velleman, "Love as a Moral Emotion").

The problem, then, of love in marriage, when one's presence and contributions become routine, would thus seem often to amount to that of rekindling desire.

Given the natural reciprocity of both unitive and contributory love, a rekindling of desire may be accomplished by simple expressions of desire, either to be with the other or for the other's good, and might help reassure the spouse of that desire and help elicit similar reciprocal expressions of desire. Expressions of affection, thoughtful remarks, and unexpected gifts, especially when they depart from routine, might likewise often be taken as a sign of genuine desire for the spouse's good. Even periods of absence might help the rekindling of desire for the other or help to contribute to his or her good. Conversely, expressions of disdain, contempt, or revulsion, even if only uttered in a heated moment and even if there is a commitment to remain with the other, will keep a spouse from feeling loved because such things contradict the *desire* to be with the other. The presence of the other, the other's commitment, the other's contributing toward one's good are not enough to feel loved. It is, additionally, a sense of desire from the other that will often be needed to feel loved. That desire alone may not be sufficient to feel loved—most would desire not merely to be loved but to be loved well, and such love, as discussed above, will generally result in thoughtful action. But if love is indeed the disposition toward desiring the other or good for the other, then we can make better sense of the phenomenon of sometimes not feeling loved within marriage, even when both spouses are present to one another and contributing to the other's good.

The relationship between desire and action in love is also of interest in marriage insofar as a marriage itself is traditionally initiated and formalized by vows. While it may seem that a vow could at most create a disposition toward acting to be with the other and acting to promote the good of the other, the vow itself might in fact also create the disposition toward desiring these things. At the very least, the vow may create a disposition toward desiring these things *for the sake of fulfilling the vow*. This would not on its own constitute love. The disposition toward desiring union with the other must be *an end itself or a source of delight in itself*, and the disposition toward desiring good for the other must be *for the other's sake* and not simply for the sake of fulfilling the vow. Nevertheless, the vow may help sustain the disposition toward desiring, and the accompanying actions, until a time at which the disposition toward desiring union with the other is once again for its own sake and the disposition toward desiring good for the other is once again for the other's sake, rather than simply for the sake of the vow.

Recognition of a moral obligation, of human dignity, or of one's relational office may thus each give rise to love. However, they will characteristically only do so if there is some sense of solidarity with the other, giving rise to a

consistent desire for their good. This, again, may be brought about by time spent with the other, by actions contributing toward their good, by recognition of their own intrinsic goodness, or by recognition of their being worthy to contribute good toward. Such a disposition toward desiring, and not only toward acting, is in turn important in facilitating the other's feeling valued. A disposition toward desiring is furthermore important because, not infrequently, our actions follow our passions and desires, especially when these are strong or when there is limited time to make use of reason; love might thereby facilitate right action and the seeking of the other's good.

The understanding of love as a disposition toward desiring also arguably has important implications with respect to love within the context of political life and with respect to how the media shapes these and other loves. Within a national context, it seems that love of one's community or country—possibly both unitive and contributory—will often prompt the attempt to bring about good for one's community or country, though this might sometimes be at the expense of other communities. Love for the beloved community can sometimes lead to hate for some other community. It seems plausible that love of *country* will often be more conducive to the good of many than will the love of a single *political party*. When love of both party and country are present, the relative strength and intensity of these loves may strongly shape political and societal life and well-being. It is conceivable that even if unitive love is stronger for one's party, contributory love might nevertheless be stronger for one's country, which would include those of other parties. As noted above, loving well, in the contributory sense, will also require awareness of others' understanding of the good, and, when necessary, trying to explain how their understanding might be incorrect, but above all, genuinely desiring their good.

Today, our loves, or sometimes lack thereof, are arguably also powerfully shaped by the media. The media routinely presents us with various goods that can awaken unitive love. The media likewise reports on individuals, actions, and groups, sometimes in a positive, though more often in a negative, light. This too shapes our loves and our sense of solidarity with others, with our country, and with groups with whom we may disagree. The media shapes our sense of whether we want to contribute some good to others (or of whether we think they are worthy of our attempting to do so) and of whether we want to be with them. Our witnessing of positive or negative events, conveyed by the media, shapes whether we go on to act altruistically toward others, which in turn affects the actions of those with whom we interact.⁴⁷ The media thus shapes

47 Fowler and Christakis present experimental evidence that the recipient of an action of goodwill is more likely to go on to act similarly and that the contagion effects of altruistic action may extend so far that a positive interaction between two persons can propagate

both our actions and our loves. That the content of news media has become dramatically more negative over the past several decades should thus be cause for concern.⁴⁸ The effect of each individual viewing is likely to be minimal. However, since there is both repeated exposure and vast reach, with subsequent contagion of either benevolent or hateful actions and words, the cumulative effects of media and news exposure on our love of neighbor is likely profound. If love is a disposition toward desiring and dispositions are often shaped slowly, each instance inclining one toward or away from a particular desire is important. The role of the media in facilitating or hampering love and solidarity, and the ensuing consequences for human action and human well-being, may be one of the most important, and neglected, moral questions of our age. Further reflection on this topic seems warranted.

In this concluding section, I have touched on some of the practical considerations that may arise from the proposed analytic definition of love. There are undoubtedly many others. Frankfurt, in fact, puts forward the controversial thesis that effectively all of our reasons for action are grounded in love.⁴⁹ The claim is found in Aquinas as well, who furthermore sees love as the cause of all the various passions.⁵⁰ I will not further discuss these questions here, but if anything close to these claims is true, it would indicate that love ought to receive considerably more philosophical attention than it has been given. I hope in this paper that by trying to give an analytic characterization as to how we use the word “love,” future reflection on the topic will be facilitated.

Harvard University
tvanderw@hsph.harvard.edu

through a social network and ultimately affect the interaction of two other persons, neither of whom know either person in the original pair (“Cooperative Behavior Cascades in Human Social Network”). In their experiments, they estimate that each additional contribution a subject initially makes to the public good is eventually tripled by other subjects who are directly or indirectly influenced to contribute more as a consequence.

48 Leetaru reports an approximately three-standard-deviation increase in the negative tone of news reporting from world broadcasts between 1979 and 2010 (“Forecasting Large-Scale Human Behavior Using Global News Media in Time and Space,” fig. 11).

49 Frankfurt argues that love is the “originating source of terminal value,” thereby binding us to our final ends and conferring on other things their instrumental value as means, and he thus concludes that “love is... the ultimate ground of practical rationality” (*The Reasons of Love*, 55–56).

50 On every action being done out of love, cf. Aquinas, *ST* I.II.28.6; on love as the cause of other passions, cf. *ST* I.II.25.4, I.II.27.4.

REFERENCES

- Aquinas, Thomas. *Summa Theologica*. Translated by the Fathers of the English Dominican Province. 5 vols. Notre Dame, IN: Ave Maria Press, 1948.
- Austin, John L. *How to Do Things with Words*. Cambridge, MA: Harvard University Press, 1962.
- Earp, Brian, Daniel Do, and Joshua Knobe. "The Ordinary Concept of True Love." In *Oxford Handbook of the Philosophy of Love*, ed. by Christopher Grau and Aaron Smuts. Oxford University Press, 2021.
- Fowler, James H., and Nicholas A. Christakis. "Cooperative Behavior Cascades in Human Social Network." *Proceedings of the National Academy of Sciences* 107, no. 12 (March 2010): 5334–38.
- Frankfurt, Harry G. *The Reasons of Love*. Princeton, NJ: Princeton University Press, 2004.
- Hacker, Peter M. S. *Human Nature: The Categorical Framework*. New York: John Wiley & Sons, 2011.
- . *The Passions: A Study of Human Nature*. Hoboken, NJ: Wiley Blackwell, 2018.
- Heathwood, Chris. "Desire-Fulfillment Theory." *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher, 135–47. London: Routledge, 2016.
- Jollimore, Troy. *Love's Vision*. Princeton, NJ: Princeton University Press, 2011.
- Kolodny, Niko. "Love as Valuing a Relationship." *Philosophical Review* 112, no. 2 (April 2003): 135–89.
- Kraut, Robert. "Love *De Re*." *Midwest Studies in Philosophy* 10, no. 1 (September 1987): 413–30.
- Leetaru, Kavel H. "Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone in Time and Space." *First Monday* 16, no. 9 (September 2011).
- Levenson, Jon D. *The Love of God: Divine Gift, Human Gratitude, and Mutual Faithfulness in Judaism*. Princeton, NJ: Princeton University Press, 2016.
- Pieper, Josef. *About Love*. Translated by R. Winston and C. Winston. Chicago: Franciscan Herald Press, 1974.
- Setiya, Kieran. "Love and the Value of a Life." *Philosophical Review* 123, no. 3 (July 2014): 251–80.
- Sidgwick, Henry. *The Methods of Ethics*. Indianapolis: Hackett, 1981.
- Sullivan, Harry S. "Conceptions of Modern Psychiatry: The First William Alanson White Memorial Lectures." *Psychiatry* 3, no. 1 (1940): 1–117.
- Stump, Eleonore. "Love, By All Accounts." *Proceedings and Addresses of the American Philosophical Association* 80, no. 2 (November 2006): 25–43.

- Taylor, Gabriele. "Love." *Proceedings of the Aristotelian Society* 76, no. 1 (June 1976): 147–64.
- Trivers, Robert L. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46, no. 1 (March 1971): 35–57.
- Velleman, J. David. "Love as a Moral Emotion." *Ethics* 109, no. 2 (January 1999): 338–74.

NONHUMAN ANIMALS AND EPISTEMIC INJUSTICE

Andrew Lopez

OVER THE LAST DECADE, animal ethics has undergone what is now typically referred to as a “political turn.” While animal ethicists have in the past touched upon political questions with regard to animals, much of the scholarship in the field has focused primarily on questions concerning the subjects of ethical consideration, how nonhuman animals do or do not fit within the ambit of moral consideration, and how individuals ought to conduct themselves in practical-moral matters in view of the moral standing of nonhuman animals (e.g., vegetarian or vegan diets, nonhuman animal trials in medical research, cohabitation with companion species). While there is no clean break between work concerned with these questions and work that is considered part of the political turn (note, for instance, that Peter Singer’s work has deep political implications), the political turn has rather marked an increased emphasis on matters such as political representation, political agency, state justice, sovereignty, and other issues typically considered by political theorists and political philosophers. Put differently, it is a turn toward a sustained consideration of “how political institutions, structures and processes might be transformed so as to secure justice for both human and non-human animals.”¹ More specifically, some theorists have turned to discussing not just how nonhuman animals are affected as outsiders that we need to take into consideration in our deliberations, but toward attempts to develop frameworks for their formal inclusion as members of our political communities.² Whether nonhuman animals need to be included in our epistemic communities has largely been unaddressed.

For decades, some animal activists and theorists have engaged with analyses of other forms of oppression to think through the normative status of nonhuman animals, ranging from ecofeminist approaches analyzing the relationship between women and nonhuman animals to engagements with multiculturalism,

1 Cochrane, Garner, and O’Sullivan, “Animal Ethics and the Political,” 264.

2 See Donaldson, “Animal Agora”; Donaldson and Kymlicka, *Zoopolis*; Driessen, “Animal Deliberation”; Kymlicka and Donaldson, “Inclusive Citizenship Beyond the Capacity Contract”; and Meijer, “Interspecies Democracies” and *When Animals Speak*.

intersectionality, race, and disability.³ Inspired by such efforts, this paper proposes that the application of the concept of epistemic injustice in animal ethics can be fruitful for determining how our actions negatively affect nonhuman animals as knowers. Epistemic injustice has its roots in feminist social epistemology and focuses on the role that unjust social relationships and institutions play in shaping an individual's or group's ability to know or participate in knowledge production, as well as determining who qualifies as a legitimate knower. Its emphasis, however, has been on analyses of distinctively human epistemic communities, and has had little engagement with other beings who may plausibly be knowers. To be sure, nonhuman animals and other beings in general often can and do play a role in human knowing, but they normally play the role of facilitator or tool for the human knower, such as in the use of dogs by law enforcement for the purposes of detecting contraband or as model organisms in scientific research.⁴ The general question of whether nonhuman animals themselves can be victims of epistemic injustice has not been taken up.⁵ I take up this question by connecting this work with animal cognition and animal ethology, and I argue that groups of nonhuman animals can be subject to epistemic injustice.

First, I lay out some of the key considerations of Miranda Fricker's conception of the nature of harm that I wish to circumvent: that an individual or a group must be aware of themselves as knowers and invested in being recognized as such to suffer epistemic injustice. Second, I shift the discussion of the nature of the harm of epistemic injustice away from Fricker's focus to instead consider harmful consequences that result from the disruption of the apportionment of epistemic resources for epistemic communities. Third, I complicate the picture Fricker proposes in two ways: (1) I highlight that knowledge often has an inextricable practical dimension, and so in these cases the practical cannot be cleaved from the epistemic in order to treat such cases as "incidentally" epistemic in nature, and (2) I show how skill or know-how as a kind

3 For ecofeminism, see Adams, *The Sexual Politics of Meat*; Gaard, *Ecofeminism*; Plumwood, *Feminism and the Mastery of Nature*; and Warren, "The Power and Promise of Ecological Feminism." For work exploring race, multiculturalism, intersectionality, and disability, see Boisseron, *Afro-Dog*; Deckha, "Animal Justice, Cultural Justice" and "Toward a Postcolonial, Posthumanist Feminist Theory"; Kim, *Dangerous Crossings*; Nocella, Bentley, and Duncan, *Earth, Animal, and Disability Liberation*; and Taylor, *Beasts of Burden*.

4 A model organism is a nonhuman species studied with the aim that data and findings from this study will apply across taxa and enable us to understand a broader range of biological processes. I will note, however, that an increasing number of animal researchers have come to see nonhuman animals as collaborators in knowledge production.

5 Exceptions include Tuvel, "Epistemic Injustice Expanded"; Podosky, "Hermeneutical Injustice and Animal Ethics." The question is also raised, though not discussed, by Catala, "Metaepistemic Injustice and Intellectual Disability."

of knowledge possessed by nonhuman animals circumvents the recognition requirement for epistemic injustice. Complicating this picture allows us to consider a form of epistemic injustice suffered by groups of nonhuman animals even if we grant that they lack the robust mental capacities to be aware of and invested in being recognized as a knower. This form of epistemic injustice can be understood as negative downstream effects on nonhuman animals' ability to acquire "answers" to "questions" they have an interest in answering: namely, acquiring true beliefs about conspecifics and their environment, acquisition of behaviors and skills that enable everyday successful coping, and accumulation of information for the distributed cognition involved in group decision-making.⁶ Fourth, and finally, I provide some examples from work in animal cognition and animal ethology to consider mechanisms for how epistemic injustice occurs for nonhuman animals as a result of direct and indirect consequences of human action. I consider research on elephants and ungulates to make the case.⁷ For elephants, I note the detrimental impact of poaching on elephants' ability to survive droughts and defend themselves from attack. In the case of ungulates, groups that migrate over long distances rely on older conspecifics to learn when and where to travel for grazing purposes (and so have their practical goals bound up with acquisition of (arguably cultural) beliefs and skills for successful navigation), as well as to provide information for decision-making in cases of disagreement or uncertainty. The removal of experienced conspecifics who serve as repositories of knowledge, however, can disrupt all of these efforts and constitute a form of epistemic injustice.

1. THE HARM OF EPISTEMIC INJUSTICE

Fricker's *Epistemic Injustice* details two forms of epistemic injustice: testimonial injustice and hermeneutical injustice. Briefly: testimonial injustice is an

- 6 This list represents the epistemic resources I have chosen to focus on for my argument, but it does not exhaust the epistemic resources that nonhuman animals have an interest in or ways of knowing they may be capable of (e.g., nonhuman animals may also be interested in acquiring true beliefs about interspecifics and/or predators). Additionally, though they lie outside the scope of this paper, I believe fruitful investigations could also be carried out exploring the connection between nonhuman animals and feminist work on embodied knowledge, affect, and tacit knowledge, to name a few.
- 7 In this paper, I only focus on wild animal communities, though I believe similar arguments can be made for the epistemic communities of domesticated animals or liminal animals (wild animals that nevertheless live among human beings). Though this may initially seem to extend beyond the interests I have listed for the "political turn" in animal studies, wild animals can be treated as sovereign communities. See Donaldson and Kymlicka, *Zoopolis*, ch. 6.

injustice suffered by virtue of being subject to a credibility deficit as a result of an identity prejudice manifested as a pernicious stereotype.⁸ This pernicious stereotype is systematic in that it tracks them across various dimensions of social life.⁹ Hermeneutical injustice occurs when a subject lacks the conceptual resources to properly understand their own experience or effectively communicate it to another subject.¹⁰ The former injustice is perpetrated by subjects, while the latter injustice is not, but occurs as a consequence of broader structural issues (it does, however, most often make itself apparent in interactions between individuals).

Given the differing causal patterns of each form of epistemic injustice, Fricker also characterizes the nature of the harm caused by each form of injustice differently. The primary harm caused by testimonial injustice is one of epistemic objectification: testimonial injustice fails to acknowledge the speaker's full status as a knower or informant, and treats them instead as a source of information.¹¹ Hermeneutical injustice, in turn, is characterized as a situated hermeneutical inequality: "their social situation is such that a collective hermeneutical gap prevents them in particular from making sense of an experience which it is strongly in their interests to render intelligible."¹²

These two forms of injustice also cause secondary harms, both practical and epistemic. For testimonial injustice, there can obviously be negative practical consequences as a result of one's testimony not being believed, e.g., one's claims to being a victim of theft may not be believed by one's community and result in one never recovering one's valuables. Epistemically, one may come to question their certainty in the truth of their testimony or belief, and as a result come to lose knowledge. For hermeneutical injustice, there are also negative practical consequences that result from being unable to properly conceptualize and communicate one's experience: as Fricker's example of sexual harassment shows, the failure to communicate these experiences prior to the development of the concept of sexual harassment meant that the harassee was not able to communicate the need to change the climate at her workplace, and was also

8 Fricker, *Epistemic Injustice*, 4.

9 Fricker, *Epistemic Injustice*, 27. Though not clear in *Epistemic Injustice*, in later work Fricker affirms that testimonial and hermeneutical injustice must be unintended, though this does not mean that the perpetrator is not culpable. See Fricker, "Evolving Concepts of Epistemic Injustice," 54–55.

10 Fricker, *Epistemic Injustice*, 7.

11 Fricker, *Epistemic Injustice*, 133.

12 Fricker, *Epistemic Injustice*, 7.

unable to provide good reason for seeking unemployment assistance once the harassment became unbearable.¹³

However, despite these being two different kinds of injustice with differing etiologies, Fricker's account of the primary harms caused by testimonial and hermeneutical injustice goes deeper, and reflects a concern with one's self and self-conception as a knower. Both forms of injustice, for Fricker, affect "the very construction (constitutive and/or causal) of selfhood."¹⁴ Drawing on Bernard Williams's work, Fricker argues that testimonial injustice compromises the psychological work that subjects perform to "steady one's mind" and solidify a sense of self.¹⁵ Hermeneutical injustice, in turn, can lead to someone being unable to understand who they are, or "can mean that someone is socially constituted as, and perhaps even caused to be, something they are not, and which it is against their interests to be seen to be."¹⁶

This deeper exploration of the primary harm of these two forms of injustice seems to close the door on the possibility of nonhuman animals being victims of epistemic injustice. While nonhuman animals can certainly, in a sense, be seen as merely sources of information, it is not clear that they can be subject to an injustice given that there is significant skepticism over whether nonhuman animals are concerned with the work of "steadying the mind" and solidifying a sense of self. While an animal can certainly exhibit anger or annoyance toward a human or a conspecific, it is doubtful that they are invested, whether aware or not, in being recognized as a knower for the purposes of their own self-conception and self-construction. Fricker claims that

to be wronged in one's capacity as a knower is to be wronged in a capacity essential to human value. When one is undermined or otherwise wronged in a capacity essential to human value, one suffers an intrinsic injustice. . . . The capacity to give knowledge to others is one side of that many sided capacity so significant in human beings: namely, the capacity for reason.¹⁷

Animals may be recognized as knowers, but they do not possess (and cannot come to possess) the investment in the value of being recognized as a knower.

Hence, on Fricker's account, animals cannot suffer from the primary harm of epistemic injustice. It does seem that they can suffer secondary harms from

13 Fricker, *Epistemic Injustice*, 162.

14 Fricker, *Epistemic Injustice*, 168.

15 Fricker, *Epistemic Injustice*, 52–54.

16 Fricker, *Epistemic Injustice*, 168.

17 Fricker, *Epistemic Injustice*, 44.

epistemic injustice, in that they can suffer both practical and epistemic harms. The practical harms should be clear: for instance, animals that perform attention-drawing behavior for the sake of drawing attention to a competitor (or predator) may suffer injury or loss if their attempt to inform or draw the attention of others is ignored. Secondary epistemic harms may be instantiated in cases in which human beings prevent or impair an animal's actions and behavior that contribute to its social and cognitive development and acquisition of information or knowledge. The difficulty, however, is that Fricker's account of the two forms of epistemic injustice she considers requires testimony and a robust perspective on one's self as a knower. "Killjoy" attitudes toward nonhuman animals displaying attention-drawing behavior may accept that they have some form of intentional agency, but they can still deny that the animal has a subjective experience of themselves and that their behavior provides testimony of that experience, just as a thermostat is not treated as providing testimony on the current temperature; the animal's behavior and the thermostat are both sources of information, and not informants.¹⁸ While ethologists and philosophers have argued for accounts of self-consciousness that admit of degrees and that nonhuman animals inhabit a broad range of this continuum, it is not clear whether any nonhuman animals today are capable of self-consciousness to the degree necessary for investment in oneself and interest in being *recognized as a knower* such that they can suffer epistemic injustice as outlined above.¹⁹ Additionally, there are many influential philosophers who argue against ascribing various mental capacities such as propositional content, concepts, or metacognition to nonhuman animals, and they argue that these all require language.²⁰ Rather than fight this uphill battle against the killjoys to establish nonhuman animals as having these mental capacities in the same way and to the same degree as human beings, I suggest that we can remain agnostic on the matter.²¹ In the case of belief, we can rely on a liberal form of dispositionalism that seeks merely to describe the logic of belief attributions without committing ourselves to any claims about actual mental capacities or attempting to reduce mental

18 "Killjoy" is a term used to refer to researchers who deny mental capacities to nonhuman animals that we typically consider characteristically human. For its origin, see Dennett, "Intentional Systems in Cognitive Ethology."

19 For examples in ethology and philosophy of self-consciousness on a continuum, see de Waal, "Fish, Mirrors, and a Gradualist Perspective on Self-Awareness"; and DeGrazia, "Self-Awareness in Animals."

20 For prominent examples, see Dummett, *The Nature and Future of Philosophy*; Stich, "Do Animals Have Beliefs?"; Brandom, *Making It Explicit*; and Davidson, "Thought and Talk."

21 There is, of course, a large literature on representationalist accounts of belief in nonhuman animals that do not require propositional content nor language, but since my argumentative strategy does not make use of it, I will not discuss it further.

activity to behavior.²² This allows us to consider forms of epistemic injustice that fall outside the scope of testimonial and hermeneutical injustice and do not require robust mental and linguistic capacities. By remaining open to the possibility of other forms of epistemic oppression, we can work toward avoiding what Kristie Dotson refers to as “contributory injustice.”²³

Contributory injustices are “caused by an epistemic agent’s situated ignorance, in the form of willful hermeneutical ignorance, in maintaining and utilizing structurally prejudiced hermeneutical resources that result in epistemic harm to the epistemic agency of a knower.”²⁴ Fricker’s account of hermeneutical injustice assumes that both the perpetrator and the victim lack the necessary concepts for the victim to understand or communicate their experience, but this is not always so; marginalized communities may develop their own hermeneutical resources for understanding their experiences, and these resources may not be shared with the broader society. Beyond that, these communities may have the hermeneutical resources necessary for understanding their experiences but fail to achieve uptake from others outside of these communities. In cases of contributory injustice, the perpetrator exhibits willful hermeneutical ignorance, which is “a willful refusal to acknowledge and to acquire the necessary tools for knowing whole parts of the world.”²⁵

In the case of nonhuman animals and epistemic injustice literature, contributory injustice occurs because we limit the scope of epistemic injustice to propositional knowledge and conceptual resources: we are concerned with whether a subject knows a particular proposition or whether they possess a certain concept, but we fail to account for nonpropositional forms of knowing, particularly skill. This second-order form of epistemic injustice leads to first-order epistemic injustice: the fact that we do not recognize animals as knowers (specifically, as skilled individuals) leads to epistemic injustice and harm against nonhuman animal communities, irrespective of their capability to be invested in being recognized as knowers. Fricker herself characterizes the harm of testimonial injustice as infringing upon one’s status as a knower, in that being wronged as a being capable of giving knowledge to others is to be

22 Dispositionism is the view that to talk about someone’s belief is to talk about the subject’s likelihood to act or feel in a certain way such that it corresponds to or characterizes the belief. For examples, see Marcus, “Some Revisionary Proposals about Belief and Believing” and “The Anti-Naturalism of Some Language Centered Accounts of Belief”; and Schwitzgebel, “A Phenomenal, Dispositional Account of Belief.”

23 Dotson, “A Cautionary Tale.”

24 Dotson, “A Cautionary Tale,” 31.

25 Pohlhaus, “Relational Knowing and Epistemic Injustice,” 729.

wronged as a being capable of reason.²⁶ While the capacity for reason may be indispensable for some forms of epistemic injustice, Carol Adams warns that these kinds of commitments may hide morally relevant details about beings other than humans, and that “aspects of animals’ lives and their experience of oppression may remain invisible because of a dominant metaphysics that views animals instrumentally and accepts a value hierarchy.”²⁷ In order to avoid committing this contributory injustice, we can draw from two main areas: feminist social epistemology and distributive accounts of epistemic injustice.²⁸

2. EPISTEMIC COMMUNITIES AND DISTRIBUTIVE EPISTEMIC INJUSTICE

In paying attention to the social and material aspects of knowing, feminist social epistemologists look at individual knowers as embodied, gendered, and situated spatially, socially, and historically. Beyond being simply individually situated knowers with particular standpoints who encounter each other, they have also argued that interpersonal experience and webs of relations between individuals are necessary prior to one being a knower.²⁹ As a result, philosophers like Lynn Hankinson Nelson have argued for the recognition of communities as the primary agents of epistemology.³⁰ This is not to deny that individuals do not know; rather, it is to claim that “the knowing we do as individuals is derivative, that your knowing or mine depends on *our* knowing, for some ‘we.’”³¹

Feminist philosophers have drawn out the relational and derivative status of individuals in human communities. As Annette Baier notes, a “person, perhaps, is best seen as one who was long enough dependent upon other persons to acquire the essential arts of personhood. Persons essentially are *second* persons, who grow up with other persons.”³² Lorraine Code expands on Baier’s analysis to establish our development into epistemic agents as reliant on the presence of other knowers: “in epistemic activity, ‘personal’ knowledge depends on common knowledge. Even the ability to change one’s mind is learned in a community that trains its members in conventions of criticism, affirmation,

26 Fricker, *Epistemic Injustice*, 44.

27 Adams, *Neither Man nor Beast*, 145.

28 This is not to deny that Fricker’s account captures something about at least some forms of epistemic injustice; rather than present an account that captures the nature of the wrong of all forms of epistemic injustice, I will instead assume a pluralist account of these wrongs.

29 For instance, see Code, *What Can She Know?*; Hankinson Nelson, *Who Knows*; Haraway, “Situated Knowledges”; and Longino, *Science as Social Knowledge*.

30 Hankinson Nelson, “Epistemological Communities.”

31 Hankinson Nelson, “Epistemological Communities,” 124.

32 Baier, *Postures of the Mind*, 84.

and second thinking.”³³ Given that many nonhuman animals are reliant on older and experienced conspecifics, their status as epistemic agents is reliant on and derivative of the epistemic agency of these conspecifics who serve as repositories of knowledge.

The need for rich epistemic relations with experienced conspecifics does not end upon reaching maturity or acquiring a particular skill. Nonhuman animals also rely on conspecifics for collective decision-making through distributed cognition. Here again, the feminist lens is key: feminist philosophy of science and social epistemology provides us with the tools to meaningfully consider the effects of social and material conditions on individual knowers and enables us to take a structural perspective on the dynamics of an epistemic community. Feminist (naturalized) epistemology also helps us characterize the wrong of epistemic injustice in terms of the harmful consequences that follow from the social and material conditions in which epistemic communities function. Drawing on, critiquing, and expanding on Quine’s naturalized epistemology, feminist naturalized epistemologists have emphasized social, political, and historical factors that play a role in determining what we as individuals and communities know or do not know.³⁴ For instance, in contrast to the individual epistemic subject that she finds in Quine’s work, Lorraine Code argues that her account of a feminist naturalized epistemology incorporates insights from ecology and “builds on the relations of organisms with one another and with their habitat, which comprises not just the physical habitat or the present one, but the complex network of locations and relations, whether social, historical, material, geographical, cultural, racial, sexual, institutional, or other, where organisms—human or nonhuman—try to live well, singly and collectively.”³⁵ Through feminist naturalized epistemologies, we can focus on what is at stake both epistemically and practically in cases of epistemic injustice and elaborate an account concerned with the distribution of epistemic goods, the negative effects of acquiring false beliefs, failing to acquire true beliefs, and taking up norms that do not tend toward acquiring knowledge.

Of course, this is not to say that epistemic injustice does not pay attention to social, political, and historical factors; what I have said earlier of Fricker’s work alone makes clear that epistemic injustice is tightly connected to structural forms of injustice. What I wish to highlight here is an emphasis on the conditions that affect the distribution of true beliefs, opportunities for the

33 Code, *What Can She Know?* 83–84.

34 See Quine, “Epistemology Naturalized.” For examples of feminist naturalized epistemology, see Hankinson Nelson, *Who Knows*; Antony, “Quine as Feminist”; Campbell, *Illusions of Paradox*; and Code, *Ecological Thinking*.

35 Code, *Ecological Thinking*, 90–91.

development of skills, and the flow of information between individuals, rather than the interactions between individuals that are central to Fricker's work and much of the epistemic injustice literature. David Coady persuasively argues for the existence of distributive epistemic injustice, which focuses on the just distribution of epistemic resources.³⁶ Though distributive epistemic injustice as a distinct form of epistemic injustice has been acknowledged by Fricker and others, it has been under-theorized.³⁷ However, we can draw from Coady's discussion of Alvin Goldman's veritistic social epistemology for thinking about distributive epistemic injustice and animal communities. As Coady notes, for Goldman, "intrinsically valuable true beliefs are the answers to the following kinds of questions: first, questions the agent happens to find interesting, second, questions the agent would find interesting if he or she had thought of them, and third, questions that the agent has an interest in having answered."³⁸ In the case of nonhuman animals, the "questions" they have an interest in "answering" involve acquiring true beliefs about conspecifics and their environment, behaviors and skills that enable everyday successful coping, and information for the distributed cognition involved in group decision-making. Since many intensely social nonhuman animals do not acquire these answers in isolation but in and through a community, special attention should be paid to the social and material consequences of these epistemic harms and what role aspects of the epistemic community and its institutions play in their prevention or perpetuation. For instance, Elizabeth Anderson has argued for the importance of social institutions in correcting for epistemic injustice.³⁹ Utilizing an analogy between epistemic justice and distributive justice, she argues that individual attempts by virtuous epistemic agents to address epistemic injustice are inadequate, and that structural epistemic injustice requires changing our social institutions: "the larger systems by which we organize the training of inquirers and the circulation, uptake, and incorporation of individuals' epistemic contributions to the construction of knowledge may need to be reformed to ensure that justice is done to each knower, and to groups of inquirers."⁴⁰ Anderson's point can be expanded

36 Coady, "Two Concepts of Epistemic Injustice."

37 Fricker proposes "discriminatory epistemic injustice" in order to distinguish her previous work from distributive epistemic injustice. See Fricker, "Evolving Concepts of Epistemic Injustice," 53. For recent work on distributive epistemic injustice in science, see Irzik and Kurtulmus, "Distributive Epistemic Justice in Science."

38 Coady, "Two Concepts of Epistemic Injustice," 103. The focus on intrinsically valuable true beliefs is important; otherwise, one could simply commit oneself to memorizing a phone book as a quick way of increasing the amount of one's true beliefs.

39 Anderson, "Epistemic Justice as a Virtue of Social Institutions."

40 Anderson, "Epistemic Justice as a Virtue of Social Institutions," 165.

to consider how social institutions can influence the distribution of epistemic resources. As a result, social institutions can influence the development of epistemic individuals and the epistemic community as a whole. Here, bringing distributive epistemic injustice together with feminist social epistemology allows us to cast the concern with the distribution or access to information and education (among other epistemic goods) as distinctively epistemic and ethical, and as falling under the ambit of epistemic injustice.⁴¹ For my purposes this is enough, though it is important for my account that it not be merely limited to interesting true belief, but that the account also incorporates know-how or skill. Doing so allows us to see how nonhuman animals can experience epistemic injustice with regard to their skills and their opportunities to acquire epistemic goods without having to take a perspective on oneself as a knower.

3. EPISTEMIC GOALS, PRACTICAL GOALS, AND KNOW-HOW

Christopher Hookway, in considering Fricker's account, notes that individuals "can be victims of epistemic injustice without making assertions and claims to knowledge, and without suffering from conceptual impoverishment."⁴² A single parent may, for instance, be unable to secure reliable childcare that would enable them to attend university or engage in other epistemic activities. Hookway notes that one could push back on this example and say that it is only an "epistemic injustice" insofar as the parent's goal is epistemic, but that the wrong suffered by the parent is not intrinsically an epistemic wrong because it could just as easily impede the parent in engaging in nonepistemic activities, such as going on a much-needed social night out with friends.⁴³ Access to reliable childcare for parents is not constitutive of being engaged in epistemic activities (and, of course, caring for one's child as an activity can itself lead to acquisition of knowledge of various sorts). But Hookway notes that what is useful about considering this kind of example is that "it reminds us that much of the time our engagement with the epistemic involves participation in goal-directed activities, not just in making assertions, communicating information, or using our conceptual resources to formulate problems and propositions."⁴⁴

For instance, student engagement in the classroom can involve the student asking a question, not because she is puzzled and is seeking an answer from the instructor, but for the purpose of contributing to the discussion and helping

41 Coady, "Two Concepts of Epistemic Injustice," 105.

42 Hookway, "Some Varieties of Epistemic Injustice," 152.

43 Fricker, *Epistemic Injustice*, 1; and Hookway, "Some Varieties of Epistemic Injustice," 154.

44 Hookway, "Some Varieties of Epistemic Injustice," 155.

it move forward in some direction. Instructors, in turn, can fail on the uptake of such a move: they may either dismiss the question as irrelevant, or misunderstand the student's question as simply requesting information from the instructor. A similar dynamic can be observed in interactions within research teams on what directions their research should take or what hypotheses should be explored.⁴⁵ The ignored student or researcher is recognized as an informant and an agent interested in acquiring knowledge, but they are not recognized as a collaborator who can contribute to knowledge production. The student and researcher are not making any knowledge claims, nor do their efforts serve as stage-setting for making any in the classroom or to their collaborators.⁴⁶ Nevertheless, the activity is epistemic in character while bound up with goal-directed activities. Beyond one's ability to direct discussion and investigation, however, this allows us to see how epistemic injustice can interact with know-how.

Epistemic injustice with regard to know-how functions differently than epistemic injustice as discussed by Fricker in terms of testimonial and hermeneutical injustice. Both of Fricker's types can be described as concerned with propositional knowledge, or knowing that-p. When we rely on another person's testimony for that-p, we have to rely on various social markers to reasonably infer whether they really know that-p. Testimonial injustices can occur here when we unfairly attribute a credibility deficit based on the social markers we use to infer whether they will provide reliable testimony. This is evident in examples where one treats an individual as untrustworthy because of a pervasive prejudice against his group identity or identities. Know-how can work differently, however: Katherine Hawley notes that, in cases of testimony, social identity can serve as a marker for an upstream indicator of knowledge.⁴⁷ That is, we can take social identity as an indicator of whether someone is likely or unlikely to know that-p, or whether they are reliable or trustworthy, on the basis of having a particular social identity. For know-how, we can rely on social markers to inform us on whether to trust that this person possesses the appropriate know-how, but in some cases we can also rely on downstream indicators of knowledge. These downstream indicators are often not constituents of knowledge, but consequences of it. Thus we can often treat successful action as

45 Here as well, feminist philosophers of science have contributed to the literature concerning scientific practice, knowledge, and values, though primarily with regard to securing objectivity in science, both at the level of the individual and the community level. See for example Harding, "Rethinking Standpoint Epistemology" and "Strong Objectivity"; Keller, *A Feeling for the Organism* and *Reflections on Gender and Science*; Longino, *Science as Social Knowledge*; and Okruhlik, "Gender and the Biological Sciences."

46 Hookway, "Some Varieties of Epistemic Injustice," 156.

47 Hawley, "Knowing How and Epistemic Injustice."

best explained by the possession of skill or know-how, and we can rely on the “deliverable” itself.⁴⁸ For example, let us say you and I plan to go for a hike in the Rockies, but we will each have to get there on our own, and it is a bit of a drive. Your successful drive to meet me for a hike in the Rockies is good evidence for me to believe that you know how to get to the trailhead in the Rockies by car. In this case, I do not need to rely on social markers like your race or gender as evidence that you really know how to drive to the trailhead. Your successful arrival is not constitutive of your know-how, but a consequence of your know-how.

Of course, this does not mean that know-how cannot be subject to some form of epistemic injustice. Epistemic injustice with regard to know-how occurs in various situations and can employ the same prejudices Fricker is concerned with. For example, the success of a racialized minority at a task or position in a professional field can be unfairly attributed to luck, instinct, or affirmative action, and not their training and skill. However, epistemic injustice with regard to know-how comes apart from testimonial injustice in another way. In the case of testimony, the speaker stands in relation to the listener as an informant. For Fricker, the wrong of testimonial injustice is that it fails to treat the speaker as an informant, and to some extent treats them merely as a source of information. In the case of know-how, though, this need not be the case. While I can ask you to demonstrate your know-how, I do not necessarily have to do so. I can simply observe and treat you as a source of information. I can learn and gain knowledge from you by observing and emulating you, without you intending for me to learn, or even without you knowing I am observing and emulating you.⁴⁹ Even in this latter instance, where you have no knowledge of being observed, treating you as a source of information does not necessarily result in my harming you by treating you as a source of information for the completion of a task or display of skill. But given the previously mentioned instances of epistemic injustice with regard to know-how, such as attributing success to luck, it can occur in these circumstances as well as those in which the performer is not aware of the injustice and does not even have some inchoate feeling of a wrong suffered.⁵⁰

This suggests that one does not need to take a perspective on their own knowledge and be invested in being seen as a knower in order to suffer from an epistemic injustice. If we extend this thought across species, all that would

48 Hawley, “Knowing How and Epistemic Injustice,” 293.

49 Hawley, “Knowing How and Epistemic Injustice,” 296.

50 An individual may also fail to recognize when they have learned from observing another individual’s know-how or undervalue it, and erroneously attribute all of the epistemic agency to themselves and fail to see they have learned this from someone else. I thank Lisa Guenther for this point.

be needed to consider that an animal can be subject to an epistemic injustice is an attribution of an intentional perspective (so that we believe that they act with goals or ends) without the animal having to care about being recognized as a knower *qua* rational being. For know-how, the knower does not need to (or even be able to) care about being recognized as a knower to suffer an epistemic injustice. Furthermore, though there are holdouts, many people (philosophers included) tend to believe that animals know how to do at least some things. This is especially the case for those of us who have observed animals solving problems. And, trivially, we sometimes do not want them to solve certain problems, and take steps to prevent them from learning how to do so (think of dog-proofing shelves or cabinets). All this opens the door for extending these considerations of epistemic justice to nonhuman animals who lack these capacities for reflection, given that they can acquire know-how, or be prevented from doing so. With that, I now turn to discussing how animals, as situated individuals and as members of communities, can suffer from this form of epistemic injustice that infringes upon their know-how.

4. ANIMAL CULTURES AND EPISTEMIC INJUSTICE

Despite dissenting arguments from behaviorists and other researchers in comparative psychology who hold to Morgan's Canon, animal ethologists have become more and more willing to claim that nonhuman animals may have not just sophisticated cognitive behavior, but social behavior as well.⁵¹ While dissenters warn of anthropocentrism, primatologists and philosophers have in turn warned against what they have called "anthropodenial" or "anthropotomy," understood as errors or refusals to acknowledge the qualities or abilities some nonhuman animals possess that we consider markers of the human (and markers of the animal in humans).⁵² One question considered by ethologists and philosophers is whether there is evidence of the possession of culture by intensely social nonhuman animals.⁵³ Increasingly, this question is answered

51 Morgan's Canon is a precept that cautions against positing complex humanlike psychological explanations for animal behavior that can be explained without such posits. For further discussion, see Steward, "Morgan's Canon."

52 De Waal, "Are We in Anthropodenial?"; and Andrews and Huss, "Anthropomorphism, Anthropotomy, and the Null Hypothesis."

53 The reader may expect a definition of culture. Kroeber and Kluckhohn collected 164 unique definitions of culture sixty-eight years ago, and that number has continued to grow for human cultures alone; see Kroeber and Kluckhohn, *Culture*. Grant Ramsey notes that, within the animal-cultures literature, cultures are typically defined "in terms of outcomes like traditions or group typicality." For the purposes of this paper, I will use Ramsey's definition: "Culture is information transmitted between individuals or groups, where this

in the affirmative for a variety of animals.⁵⁴ My concern, however, is how the possession of culture by nonhuman animals makes them vulnerable to epistemic injustice along the lines considered above. That is, how an account of epistemic injustice that focuses on the social and material conditions in which epistemic communities function can make sense of epistemic injustice suffered by animal communities.

There is substantial evidence from ethology showing that animals depend on input or interaction from their environment (including conspecifics) in order to acquire certain skills, and can have their long-term behavior shaped in a variety of ways by their conspecifics. While some behaviors can be reliably considered innate, others are triggered by specific environmental inputs (or lack thereof) and are important for ontogenetic and developmental purposes. Cultural behaviors differ from both, but should be especially carefully distinguished from the latter kind. Innate behaviors are reliably expressed regardless of hyper-specific environmental input (trivially, all behaviors require some environmental input in the form of nutrition, among other things). For example, rats, regardless of whether they are allowed to socialize or kept in isolation from birth, will construct similar nests, and so this behavior can be reliably considered innate.⁵⁵ However, if female rats are prevented from licking their own genitalia while pregnant, they will reliably eat their young after giving birth. This second form of behavior can arguably be said to be elicited or enabled by outside input. That is, resources from outside of the genome allow for this complex behavior encoded in the genome to be expressed under standard conditions.

Cultural behaviors cannot be explained away in such a manner, due to their uptake and performances being contingent, yet nevertheless sometimes having survival value.⁵⁶ For instance, a subset of a population of bottlenose

information flows through and brings about the reproduction of, and a lasting change in, the behavioral trait" ("What Is Animal Culture?," 346, 348).

54 See Andrews, *The Animal Mind*, ch. 8; Brown, "Animal Traditions"; Cantor, Shoemaker, Cabral, et al., "Multilevel Animal Societies Can Emerge from Cultural Transmission"; Galef and Aleen, "A New Model System for Studying Behavioural Traditions in Animals"; Goodrich, "Varieties of Culture"; Huffman and Quiatt, "Stone Handling by Japanese Macaques (*Macaca fuscata*)"; Meynell and Lopez, "Gendering Animals"; Ramsey, "Culture in Humans and Other Animals"; Sapolsky and Share, "A Pacific Culture among Wild Baboons"; van Schaik, Ancrenaz, Borgen, et al., "Orangutan Cultures and the Evolution of Material Culture"; Whitehead and Rendell, *The Cultural Lives of Whales and Dolphins*; and Brakes, Dall, Aplin, et al., "Animal Cultures Matter for Conservation."

55 Lehrman, "A Critique of Konrad Lorenz's Theory of Instinctive Behavior," 342.

56 The capacity to be sensitive to adopting behaviors and practices of conspecifics itself likely confers survival advantages, but no one behavior adopted need do so.

dolphins has been found to use marine sponges as foraging tools to find food.⁵⁷ This form of social learning has been most reliably traced as skill transmission from female parent to offspring, though males rarely engage in the practice, even if their female parent teaches them; “sponging” dolphins in almost all cases are females. This behavior is unique, however, in that it is exhibited most strongly within a particular matriline, and persists despite the fact that sponging dolphins live among other dolphins who do not sponge and regardless of whether the sponging females mate with the small number of sponging males. In other words, this practice and preference for sponging over other methods of acquiring food is reliably passed “from mother to daughter,” with inconsistent uptake “from mother to son.” Sponging stands out here, not just because of the manipulation of materials as a form of tool use, but because the behavior does not spread to the rest of the population and instead remains a form of cultural diversity within the population.⁵⁸

However, studying animal cognition and ethology among a variety of species clearly would (often, though not always) depend on their species-typical capacities, and an organism’s cognitive complexity may be expressed along various axes. With that in mind, I turn to considering the transmission of information and acquisition of behavior among intensely social nonhuman animals, specifically on the transfer of know-how and decision-making among communities of elephants and ungulates.

We can consider two examples that display the harms both to individual animals and to their communities at large: elephants and bighorn sheep. African elephants in the wild must often deal with the environmental danger of drought.⁵⁹ Elephant calves, in particular, are most vulnerable to perishing as a result of dehydration and lack of food during droughts. Research on calf survival rate has found that the mortality rate for calves was higher among younger mothers, likely due to their lack of experience of their environment and strategies for dealing with danger. However, calves and females, unlike males, are not solitary animals. Rather, calves and females tend to live in clans. Calf survival rate was positively correlated not just with the age of the calf’s mother, but with the presence of older females, or matriarchs, in the clan. Matriarchs serve as “repositories of socially transmitted knowledge,” in that they have years of expe-

57 Krützen, Mann, Heithaus, et al., “Cultural Transmission of Tool Use in Bottlenose Dolphins”; Krützen, Kreicker, MacLeod, et al., “Cultural Transmission of Tool Use by Indo-Pacific Bottlenose Dolphins (*Tursiops* Sp.) Provides Access to a Novel Foraging Niche.”

58 Krützen, Mann, Heithaus, et al., “Cultural Transmission of Tool Use in Bottlenose Dolphins,” 8943.

59 Foley, Pettolelli, and Foley, “Severe Drought and Calf Survival in Elephants.”

rience and knowledge of seasonal changes, as well as prior success in handling droughts and leading others through them.

These matriarchs have knowledge beyond a “map” of sorts, however; they also possess knowledge of strategies for dealing with danger. Other researchers found that clans with older matriarchs were on average more successful at fending off attacks by lions.⁶⁰ Importantly, these clans were also more skilled at successfully discriminating the roars of female lions from male lions. Given the significant size difference between female and male lions, defending against male lions as a group requires different strategies by elephants than defending against female lions. Once again, the presence of an older matriarch correlated positively with successful defenses against attacks by male lions. Success and efficiency also increased over time, as younger elephants benefited from the knowledge and skill of the matriarchs by following their lead and eventually picking up the knowledge and skill themselves. Given the importance of older matriarchs for this dissemination of know-how among younger elephants, one clearly infers the consequences and harms suffered by clans who do not have these knowledgeable members. Clans who, over the years, had lost older members to ivory poaching performed statistically worse with regard to both calf survival in droughts and defense against male and female lions.

Recent research that brings together movement ecology with collective behavior and collective motion highlights the importance of both social behavior within the group as well as the importance of particular individuals within the group, with much of the research focused on modeling heterogeneity, social interaction, and information flow within groups.⁶¹

Variation in cognition can influence how individuals respond to and communicate about their environment, which may scale to shape how a collective solves a cognitive task. Interactions among individuals that differ in the performance of a cognitive task can drive collective foraging behavior. The collective motion of ungulate groups can also depend on or be influenced by particular individuals in the group. For instance, among caribou, older and more experienced individuals are typically thought to hold informal leadership positions and guide migration-scale movements; however, pregnant or nursing

60 McComb, Moss, Durant, et al., “Matriarchs as Repositories of Social Knowledge in African Elephants.”

61 For examples, see Jolles, King, and Killen, “The Role of Individual Heterogeneity in Collective Animal Behaviour”; Riotte-Lambert and Matthiopoulos, “Environmental Predictability as a Cause and Consequence of Animal Movement”; Strandburg-Peshkin, Papageorgiou, Crofoot, and Farine, “Inferring Influence and Leadership in Moving Animal Groups”; Torney, Lamont, Debell, et al., “Inferring the Rules of Social Interaction in Migrating Caribou”; Westley, Berdahl, Torney, and Biro, “Collective Movement in Ecology.”

females may guide movements toward habitats with better forage opportunities.⁶² Leadership in these groups is flexible, and can involve position within the group, influence, or information flow.⁶³

Individuals within groups play an important role in fusion-fission dynamics that are part of collective motion, as groups do not always agree about when and where to go, all of which are affected by influence, leadership, genetic predispositions, life experiences, and species, among other factors.⁶⁴ Unique migratory portfolios (i.e., the variation in migratory behaviors across space and time among individuals within populations) arise out of these dynamics, with native populations having more diverse portfolios, restored populations having less diversity, and augmented populations being somewhere in between.⁶⁵ In particular, while all three types of populations exhibit various levels of movement in terms of elevation, significant differences are found for geographic migration, with native populations exhibiting greater range.

For instance, among bison, research suggests that bison tend to associate with conspecifics that possess spatial knowledge different from their own, and that what individuals know and what others know influences their patch decisions.⁶⁶ In instances of conflict or disagreement, however, the bison use group familiarity combined with their own knowledge and recent experiences to decide whether to follow or leave a group.

Bighorn sheep are also highly sensitive to the presence of older and more experienced conspecifics. Ecologists have recently confirmed a long-standing hunch that bighorn sheep and other ungulates like moose, sheep, and bison rely on the social transmission of knowledge to effectively migrate hundreds of kilometers at certain points of the year.⁶⁷ This view is supported by the fact

62 Garland, Berdahl, Sun, and Boltt, "Anatomy of Leadership in Collective Behaviour," 2.

63 Garland, Berdahl, Sun, and Boltt, "Anatomy of Leadership in Collective Behaviour"; Strandburg-Peshkin, Papageorgiou, Crofoot, and Farine, "Inferring Influence and Leadership in Moving Animal Groups."

64 Berg, Hebblewhite, St. Clair, and Merrill, "Prevalence and Mechanisms of Partial Migration in Ungulates"; Sawyer, Merkle, Middleton, et al., "Migratory Plasticity Is Not Ubiquitous among Large Herbivores."

65 Lowrey, Proffitt, McWhirter, et al., "Characterizing Population and Individual Migration Patterns among Native and Restored Bighorn Sheep (*Ovis canadensis*); Lowrey, McWhirter, Proffitt, et al., "Individual Variation Creates Diverse Migratory Portfolios in Native Populations of a Mountain Ungulate." Note: Native populations are populations endemic to a region. Restored populations are populations that have been reintroduced into a region from which they had disappeared. Augmented populations are populations that have had individuals introduced to reinforce currently present populations in the region.

66 Merkle, Sigaud, and Fortin, "To Follow or Not?"

67 Jesmer, Merkle, Goheen, et al., "Is Ungulate Migration Culturally Transmitted?"

that populations of ungulates that were reintroduced to certain habitats did not migrate, unlike the previous inhabitants of these territories. However, over several decades and generations, ecologists found that these ungulates began to migrate, and their skill in doing so improved over time. In order to successfully migrate, bighorn sheep need to know not just where to go, but when to go—experience and knowledge are crucial for successfully moving along over time at a pace and direction such that vegetation is present at that point in time and that it is relatively young for easier grazing or foraging, in a phenomenon called “green-wave surfing.”⁶⁸ Bighorn sheep that had several generations of experience with these territories and migratory patterns surfed more efficiently than transplanted individuals. Over time, however, knowledge and skill in surfing was disseminated to transplanted individuals, who steadily improved their surfing and began using the same paths as members of the historical population.

Research on mammals has found that migratory mammals are more likely to increase annually compared to their nonmigratory conspecifics, and it is possible that “the vagility of migratory mammals could aid their ability to escape anthropogenic threats in areas where non-migrants would have more difficulty moving territories.”⁶⁹ But bighorn sheep and other ungulates can still be vulnerable to the loss of this knowledge. Rapid anthropogenic climate change, as well as habitat destruction as a result of human development, can have a negative impact on their knowledge practices, and “the reliance of collectively navigating species on inter-individual cues can also result in cascading consequences when one individual makes a mistake. . . . Increasing anthropogenic change could result in greater potential for mistakes and greater cost to collective migrants.”⁷⁰ Climate change can have the deleterious effect of changing the “green-wave” schedule by which vegetation grows (if it grows at all), which could create scenarios in which bighorn sheep arrive too early or too late.⁷¹ In sum, disruption of their communities and their environments compromises their successful transmission of know-how among their conspecifics. Know-how is lost, and so they suffer epistemic harms, which compromises their ability to cope and survive, and so they also suffer practical harms. As a result, bighorn

68 Aikens, Mysterud, Merkle, et al., “Wave-like Patterns of Plant Phenology Determine Ungulate Movement Tactics.” Note, the green wave hypothesis may not apply to all ungulates, as some research suggests that bison do not follow it yet have an impact on the green wave itself; see Geremia, Merkle, Eacker, et al., “Migrating Bison Engineer the Green Wave.”

69 Hardesty-Moore, Deinet, Freeman, et al., “Migration in the Anthropocene,” 7.

70 Hardesty-Moore, Deinet, Freeman, et al., “Migration in the Anthropocene,” 7.

71 Aikens, Mysterud, Merkle, et al., “Drought Reshuffles Plant Phenology and Reduces the Foraging Benefit of Green-wave Surfing for a Migratory Ungulate”; and Cremonese, Filippa, Galvagno, et al., “Heat Wave Hinders Green Wave.”

sheep, among other species, suffer a form of epistemic injustice, given that they are prevented by direct and indirect human action from acquiring both true beliefs about conspecifics and their environment, as well as acquisition of behaviors and skills that enable everyday successful coping. These injustices are difficult to correct as well, given the time required for the vertical transmission of knowledge among generations of ungulates. Jesmer, Merkle, Goheen, et al. note that

restoring migratory populations after extirpation or the removal of barriers to movement will be hindered by poor foraging efficiency, suppressed fitness, and reduced population performance. Thus, conservation of existing migration corridors, stopover sites, and seasonal ranges not only protects the landscapes that ungulates depend on; such efforts also maintain the traditional knowledge and culture that migratory animals use to bolster fitness and sustain abundant populations.⁷²

There is a possible debate concerning the status of these behaviors as emblematic of culture: not all social learning need rise up to the level of being an example of culture. In any case, scientists invested in conserving the diverse group-specific behavior seen in animal populations have found this diversity threatened by humans.⁷³ However, we can sidestep this debate by focusing on the behaviors being transmitted from one individual to another as an instance of the acquisition of know-how, and examining how the compromise of this acquisition by human interventions (broadly) constitutes epistemic injustice. Various behaviors and practices (such as those discussed previously) that serve as candidates for the possession of culture in animals have an impact on reproduction and survival. Conservation biologists have recently made the case that cultural preservation ought to be part of the mission of conservation precisely because of this impact.⁷⁴ They note that some conservation programs have failed “by neglecting key repositories of socially transmitted knowledge.”⁷⁵ By failing to preserve these “key repositories” or by actively disrupting them, human beings negatively affect the transmission of know-how between intensely social animals, which results in significant secondary harms. The harms suffered by these animals are epistemic, given that they fail to acquire a skill through mimicry, and practical, because of the downstream

72 Jesmer, Merkle, Goheen, et al., “Is Ungulate Migration Culturally Transmitted?” 1025.

73 Kühl, Boesch, Kulik, et al., “Human Impact Erodes Chimpanzee Behavioral Diversity”; van Schaik, “Fragility of Traditions.”

74 Brakes, Dall, Aplin, et al., “Animal Cultures Matter for Conservation.”

75 Brakes, Dall, Aplin, et al., “Animal Cultures Matter for Conservation,” 1032.

consequences of lacking the aforementioned skills. For animals, the epistemic and the practical are not separable, given that these skills are deployed in goal-directed activity. Failure to achieve these goals can be disastrous.

5. CONCLUSION

In this paper, I have considered the question of epistemic injustice in the case of nonhuman animals, and have attempted to argue that they can in fact suffer from this form of injustice. I have done so by shifting the focus of discussion to the nature of the harm of epistemic injustice and considering the harm through a feminist lens. Doing so, I argue, allows us to see how the distribution of epistemic goods is negatively affected by human action in a way that compromises the acquisition of information and know-how for nonhuman animals. Whether I succeed in doing so or not, part of my aim here has also been to highlight the lack of discussion in recent epistemic injustice literature on ways of knowing other than propositional knowledge. With few exceptions, such as Katherine Hawley or Alexis Shotwell, much of the literature on other ways of knowing that has its roots in feminist philosophy and other philosophies has been sidelined.⁷⁶ A consideration of these other ways of knowing enables us to see how epistemic injustice can affect them as well, and can shine light on who experiences this epistemic injustice. In “Land as Pedagogy,” Leanne Simpson writes about gratitude for the teachings of animals like squirrels, and the knowledge they can pass on to us.⁷⁷ While we can learn much from other animals, my aim here has been to show concern for what they can learn from each other and what they can know for themselves.⁷⁸

Queen's University

Andrew.Lopez@queensu.ca

76 See Hawley, “Knowing How and Epistemic Injustice”; Shotwell, *Knowing Otherwise* and “Forms of Knowing and Epistemic Resources.”

77 Simpson, “Land as Pedagogy.”

78 I would like to thank attendees of the Dalhousie University philosophy colloquium and participants in the 2021 MANCEPT workshop on Politics, Animals, and Technology for their thoughtful feedback. Conversations with Emily Bingeman, Matthew Calarco, Julio Covarrubias, Lesley Jamieson, Letitia Meynell, Erik Nelson, Elliot Paul, and Agnes Tam on the general underpinnings of the paper and its various sections proved invaluable as well. I am especially grateful to Lisa Guenther, Arthur Hill, Will Kymlicka, Michael Luoma, and Adam Schipper for detailed feedback on the paper at various stages. I would also like to thank the anonymous referees for their comments on the draft, Noah Trammell for his careful reading, and Moisés Macías Bustos for insisting I submit the draft instead of endlessly working on it.

REFERENCES

- Adams, Carol. *Neither Man nor Beast: Feminism and the Defense of Animals*. 1994. New York: Bloomsbury, 2018.
- . *The Sexual Politics of Meat*. 1990. New York: Continuum, 2000.
- Aikens, Ellen O., Kevin L. Monteith, Jerod A. Merkle, Samantha P. H. Dwinnell, Gary L. Fralick, and Matthew J. Kauffman. "Drought Reshuffles Plant Phenology and Reduces the Foraging Benefit of Green-wave Surfing for a Migratory Ungulate." *Global Change Biology* 26, no. 8 (August 2020): 4215–25.
- Aikens, Ellen O., Atle Myrsterud, Jerod A. Merkle, Francesca Cagnacci, Inger Maren Rivrud, Mark Hebblewhite, Mark A. Hurley, et al. "Wave-like Patterns of Plant Phenology Determine Ungulate Movement Tactics." *Current Biology* 30, no. 17 (September 2020): 3444–49.
- Anderson, Elizabeth. "Epistemic Justice as a Virtue of Social Institutions." *Social Epistemology* 26, no. 2 (April 2012): 163–73.
- Andrews, Kristin. *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*. 2nd ed. New York: Routledge, 2020.
- Andrews, Kristin, and Jacob Beck, eds. *The Routledge Handbook of Philosophy of Animal Minds*. New York: Routledge, 2018.
- Andrews, Kristin, and Brian Huss. "Anthropomorphism, Anthropectomy, and the Null Hypothesis." *Biology and Philosophy* 29, no. 5 (September 2014): 711–29.
- Antony, Louise M. "Quine as Feminist: The Radical Import of Naturalized Epistemology." In *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, edited by Louise M. Antony and Charlotte E. Witt, 110–53. New York: Routledge, 2002.
- Baier, Annette. *Postures of the Mind: Essays on Mind and Morals*. Minneapolis: University of Minnesota Press, 1985.
- Berg, Jodi E., Mark Hebblewhite, Colleen C. St. Clair, and Evelyn H. Merrill. "Prevalence and Mechanisms of Partial Migration in Ungulates." *Frontiers in Ecology and Evolution* 7 (August 2019): 1–17.
- Boisseron, Bénédicte. *Afro-Dog: Blackness and the Animal Question*. New York: Columbia University Press, 2018.
- Brakes, Philippa, Sasha R. X. Dall, Lucy M. Aplin, Stuart Bearhop, Emma L. Carroll, Paolo Ciucci, Vicki Fishlock, et al. "Animal Cultures Matter for Conservation." *Science* 363, no. 6431 (March 8, 2019): 1032–34.
- Brandom, Robert. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press, 1994.
- Brown, Rachael L. "Animal Traditions: What They Are and Why They Matter."

- In Andrews and Beck, *The Routledge Handbook of Philosophy of Animal Minds*, 362–71.
- Campbell, Richmond. *Illusions of Paradox: A Feminist Epistemology Naturalized*. Lanham, MD: Rowman & Littlefield, 1998.
- Cantor, Maurício, Lauren G. Shoemaker, Reniel B. Cabral, César O. Flores, Melinda Varga, and Hal Whitehead. “Multilevel Animal Societies Can Emerge from Cultural Transmission.” *Nature Communications* 6, no. 8091 (September 8, 2015): 1–10.
- Catala, Amandine. “Metaepistemic Injustice and Intellectual Disability: A Pluralist Account of Epistemic Agency.” *Ethical Theory and Moral Practice* 23, no. 5 (November 2020): 755–76.
- Coady, David. “Two Concepts of Epistemic Injustice.” *Episteme* 7, no. 2 (June 2010): 101–13.
- Cochrane, Alasdair, Robert Garner, and Siobhan O’Sullivan. “Animal Ethics and the Political.” *Critical Review of International Social and Political Philosophy* 21, no. 2 (2018): 261–77.
- Code, Lorraine. *Ecological Thinking: The Politics of Epistemic Location*. Oxford: Oxford University Press, 2006.
- . *What Can She Know? Feminist Theory and Construction of Knowledge*. Ithaca, NY: Cornell University Press, 1991.
- Cremonese, Edoardo, Gianluca Filippa, Marta Galvagno, Consolata Siniscalco, Ludovica Oddi, Umberto Morra di Cella, and Mirco Migliavacca. “Heat Wave Hinders Green Wave: The Impact of Climate Extreme on the Phenology of a Mountain Grassland.” *Agricultural and Forest Meteorology* 247 (December 2017): 320–30.
- Davidson, Donald. “Thought and Talk.” In *Mind and Language*, edited by Samuel D. Guttenplan, 7–24. Oxford: Oxford University Press, 1975.
- Deckha, Maneesha. “Animal Justice, Cultural Justice: A Posthumanist Response to Cultural Rights in Animals.” *Journal of Animal Law and Ethics* 2, no. 189 (2007): 189–229.
- . “Toward a Postcolonial, Posthumanist Feminist Theory: Centralizing Race and Culture in Feminist Work on Nonhuman Animals.” *Hypatia* 27, no. 3 (August 2012): 527–45.
- DeGrazia, David. “Self-Awareness in Animals.” In *The Philosophy of Animal Minds*, edited by Robert W. Lurz, 201–17. Cambridge: Cambridge University Press, 2009.
- Dennett, Daniel C. “Intentional Systems in Cognitive Ethology: The ‘Panglossian Paradigm’ Defended.” *Behavioral and Brain Sciences* 6, no. 3 (September 1983): 343–55.
- De Waal, Frans. “Are We in Anthropodenial?” *Discover*, January 18, 1997. <https://>

- www.discovermagazine.com/planet-earth/are-we-in-anthropodenial.
- . “Fish, Mirrors, and a Gradualist Perspective on Self-Awareness.” *PLOS Biology* 17, no. 2 (February 7, 2019): e3000112.
- Donaldson, Sue. “Animal Agora: Animals Citizens and the Democratic Challenge.” *Social Theory and Practice* 46, no. 4 (October 2020): 709–35.
- Donaldson, Sue, and Will Kymlicka. *Zoopolis: A Political Theory of Animal Rights*. Oxford: Oxford University Press, 2011.
- Dotson, Kristie. “A Cautionary Tale: On Limiting Epistemic Oppression.” *Frontiers: A Journal of Women Studies* 33, no. 1 (2012): 24–47.
- Driessen, Clemens. “Animal Deliberation.” In *Political Animals and Animal Politics*, edited by Marcel Wissenburg and David Schlosberg, 90–104. Basingstoke, UK: Palgrave Macmillan, 2014.
- Dummett, Michael. *The Nature and Future of Philosophy*. New York: Columbia University Press, 2010.
- Foley, Charles, Nathalie Petteorelli, and Lara Foley. “Severe Drought and Calf Survival in Elephants.” *Biology Letters* 4, no. 5 (October 23, 2008): 541–44.
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Clarendon Press, 2007.
- . “Evolving Concepts of Epistemic Injustice.” In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus Jr., 53–60. New York: Routledge, 2017.
- Gaard, Greta. *Ecofeminism: Women, Animals, Nature*. Philadelphia: Temple University Press, 1993.
- Galef, Bennett G., and Craig Aleen. “A New Model System for Studying Behavioural Traditions in Animals.” *Animal Behaviour* 50, no. 3 (1995): 705–17.
- Garland, Joshua, Andrew M. Berdahl, Jie Sun, and Erik M. Bollt. “Anatomy of Leadership in Collective Behaviour.” *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, no. 7 (July 2018): 075308.
- Geremia, Chris, Jerod A. Merkle, Daniel R. Eacker, Rick L. Wallen, P. J. White, Mark Hebblewhite, and Matthew J. Kauffman. “Migrating Bison Engineer the Green Wave.” *Proceedings of the National Academy of Sciences* 116, no. 51 (December 17, 2019): 25707–13.
- Goodrich, Grant. “Varieties of Culture.” In Andrews and Beck, *The Routledge Handbook of Philosophy of Animal Minds*, 354–61.
- Hankinson Nelson, Lynn. “Epistemological Communities.” In *Feminist Epistemologies*, edited by Linda Alcoff and Elizabeth Potter, 121–59. New York: Routledge, 1993.
- . *Who Knows: From Quine to a Feminist Empiricism*. Philadelphia: Temple University Press, 1990.

- Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3 (Autumn 1988): 575–99.
- Hardesty-Moore, Molly, Stefanie Deinet, Robin Freeman, Georgia C. Titcomb, Erin M. Dillon, Keenan Stears, Maggie Klope, et al. "Migration in the Anthropocene: How Collective Navigation, Environmental System and Taxonomy Shape the Vulnerability of Migratory Species." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, no. 1746 (May 19, 2018): 20170017.
- Harding, Sarah. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity'?" In *Feminist Epistemologies*, edited by Linda Alcoff and Elizabeth Potter, 49–82. New York: Routledge, 1993.
- . "'Strong Objectivity': A Response to the New Objectivity Question." *Synthese* 104 (September 1995): 331–49.
- Hawley, Katherine. "Knowing How and Epistemic Injustice." In *Knowing How: Essays on Knowledge, Mind, and Action*, edited by John Bengson and Marc A. Moffett, 283–99. New York: Oxford University Press, 2012.
- Hookway, Christopher. "Some Varieties of Epistemic Injustice: Reflections on Fricker." *Episteme* 7, no. 2 (June 2010): 151–63.
- Huffman, Michael A., and Duane Quiatt. "Stone Handling by Japanese Macaques (*Macaca fuscata*): Implications for Tool Use of Stone." *Primates* 27, no. 4 (October 1986): 413–23.
- Irzik, Gürol, and Faik Kurtulmus. "Distributive Epistemic Justice in Science." *British Journal for the Philosophy of Science* (forthcoming).
- Jesmer, Brett R., Jerod A. Merkle, Jacob R. Goheen, Ellen O. Aikens, Jeffrey L. Beck, Alyson B. Courtemanch, Mark A. Hurley, et al. "Is Ungulate Migration Culturally Transmitted? Evidence of Social Learning from Translocated Animals." *Science* 361, no. 6406 (September 7, 2018): 1023–25.
- Jolles, Jolle W., Andrew J. King, and Shaun S. Killen. "The Role of Individual Heterogeneity in Collective Animal Behaviour." *Trends in Ecology and Evolution* 35, no. 3 (March 2020): 278–91.
- Keller, Evelyn Fox. *A Feeling for the Organism: The Life and Work of Barbara McClintock*. San Francisco: W. H. Freeman, 1983.
- . *Reflections on Gender and Science*. New Haven, CT: Yale University Press, 1985.
- Kim, Claire Jean. *Dangerous Crossings: Race, Species, and Nature in a Multicultural Age*. New York: Cambridge University Press, 2015.
- Kroeber, A. L., and Clyde Kluckhohn. *Culture: A Critical Review of Concepts and Definitions*. 1952. New York: Vintage Books, 1963.
- Krützen, Michael, Sina Kreicker, Colin D. MacLeod, Jennifer Learmonth, Anna

- M. Kopps, Pamela Walsham, and Simon J. Allen. "Cultural Transmission of Tool Use by Indo-Pacific Bottlenose Dolphins (*Tursiops* sp.) Provides Access to a Novel Foraging Niche." *Proceedings of the Royal Society B: Biological Sciences* 281, no. 1784 (June 7, 2014): 20140374.
- Krützen, Michael, Janet Mann, Michael R. Heithaus, Richard C. Connor, Lars Bejder, and William B. Sherwin. "Cultural Transmission of Tool Use in Bottlenose Dolphins." *Proceedings of the National Academy of Sciences* 102, no. 25 (June 21, 2005): 8939–43.
- Kühl, Hjalmar S., Christophe Boesch, Lars Kulik, Fabian Haas, Mimi Arandjelovic, Paula Dieguez, Gaëlle Bocksberger, et al. "Human Impact Erodes Chimpanzee Behavioral Diversity." *Science* 363, no. 6434 (March 29, 2019): 1453–55.
- Kymlicka, Will, and Sue Donaldson. "Inclusive Citizenship Beyond the Capacity Contract." In *The Oxford Handbook of Citizenship*, edited by Ayelet Shachar, Rainer Bauböck, Irene Bloemraad, and Maarten Vink, 838–60. Oxford: Oxford University Press, 2017.
- Lehrman, Daniel S. "A Critique of Konrad Lorenz's Theory of Instinctive Behavior." *Quarterly Review of Biology* 28, no. 4 (1953): 337–63.
- Longino, Helen. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press, 1990.
- Lowrey, B., D. E. McWhirter, K. M. Proffitt, K. L. Monteith, A. B. Courtemanch, P. J. White, J. T. Paterson, S. R. Dewey, and R. A. Garrott. "Individual Variation Creates Diverse Migratory Portfolios in Native Populations of a Mountain Ungulate." *Ecological Applications* 30, no. 5 (July 2020): e2106.
- Lowrey, Blake, Kelly M. Proffitt, Douglas E. McWhirter, Patrick J. White, Alyson B. Courtemanch, Sarah R. Dewey, Hollie M. Miyasaki, et al. "Characterizing Population and Individual Migration Patterns among Native and Restored Bighorn Sheep (*Ovis canadensis*)." *Ecology and Evolution* 9, no. 15 (August 2019): 8829–39.
- Marcus, Ruth Barcan. "The Anti-Naturalism of Some Language Centered Accounts of Belief." *Dialectica* 49, no. 2–4 (1995): 113–29.
- . "Some Revisionary Proposals about Belief and Believing." *Philosophy and Phenomenological Research* 50, no. S1 (Autumn 1990): 133–53.
- McComb, Karen, Cynthia Moss, Sarah M. Durant, Lucy Baker, and Soila Sayialel. "Matriarchs as Repositories of Social Knowledge in African Elephants." *Science* 292, no. 5516 (April 20, 2001): 491–94.
- Meijer, Eva. "Interspecies Democracies." In *Animal Ethics in the Age of Humans: Blurring Boundaries in Human-Animal Relationships*, edited by Bernice Bovenkerk and Jozef Keulartz, 53–72. New York: Springer, 2016.
- . *When Animals Speak: Toward an Interspecies Democracy*. New York:

- New York University Press, 2019.
- Merkle, Jerod A., Marie Sigaud, and Daniel Fortin. "To Follow or Not? How Animals in Fusion–Fission Societies Handle Conflicting Information during Group Decision-Making." *Ecology Letters* 18, no. 8 (August 1, 2015): 799–806.
- Meynell, Letitia, and Andrew Lopez. "Gendering Animals." *Synthese* 199, no. 5–6 (December 2021): 4287–4311.
- Nocella, Anthony J., II, Judy K. C. Bentley, and Janet M. Duncan, eds. *Earth, Animal, and Disability Liberation: The Rise of the Eco-Ability Movement*. New York: Peter Lang, 2012.
- Okruhlik, Kathleen. "Gender and the Biological Sciences." *Canadian Journal of Philosophy Supplementary Volume* 20 (1994): 21–42.
- Plumwood, Val. *Feminism and the Mastery of Nature*. London: Routledge, 1993.
- Podosky, Paul-Mikhail. "Hermeneutical Injustice and Animal Ethics: Can Non-human Animals Suffer from Hermeneutical Injustice?" *Journal of Animal Ethics* 8, no. 2 (Fall 2018): 216–28.
- Pohlhaus, Gaile, Jr. "Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance." *Hypatia* 27, no. 1 (Fall 2012): 715–35.
- Quine, W. V. "Epistemology Naturalized." In *Ontological Relativity and Other Essays*, 69–90. New York: Columbia University Press, 1969.
- Ramsey, Grant. "Culture in Humans and Other Animals." *Biology and Philosophy* 28, no. 3 (May 2013): 457–79.
- . "What Is Animal Culture?" In Andrews and Beck, *The Routledge Handbook of Philosophy of Animal Minds*, 345–53.
- Riotte-Lambert, Louise, and Jason Matthiopoulos. "Environmental Predictability as a Cause and Consequence of Animal Movement." *Trends in Ecology and Evolution* 35, no. 2 (February 2020): 163–74.
- Sapolsky, Robert M., and Lisa J. Share. "A Pacific Culture among Wild Baboons: Its Emergence and Transmission." *PLOS Biology* 2, no. 4 (April 13, 2004): 534–41.
- Sawyer, Hall, Jerod A. Merkle, Arthur D. Middleton, Samantha P. H. Dwinell, and Kevin L. Monteith. "Migratory Plasticity Is Not Ubiquitous among Large Herbivores." *Journal of Animal Ecology* 88, no. 3 (December 7, 2018): 450–60.
- Schwitzgebel, Eric. "A Phenomenal, Dispositional Account of Belief." *Noûs* 36, no. 2 (June 2002): 249–75.
- Shotwell, Alexis. "Forms of Knowing and Epistemic Resources." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus Jr., 79–88. New York: Routledge, 2017.

- . *Knowing Otherwise: Race, Gender, and Implicit Understanding*. University Park, PA: Pennsylvania State University Press, 2011.
- Simpson, Leanne Betasamosake. "Land as Pedagogy: Nishnaabeg Intelligence and Rebellious Transformation." *Decolonization: Indigeneity, Education and Society* 3, no. 3 (2014): 1–25.
- Steward, Helen. "Morgan's Canon: Animal Psychology in the Twentieth Century and Beyond." In *Animals: A History*, edited by Peter Adamson and G. Fay Edwards, 293–318. Oxford: Oxford University Press, 2018.
- Stich, Stephen P. "Do Animals Have Beliefs?" *Australasian Journal of Philosophy* 57, no. 1 (1979): 15–28.
- Strandburg-Peshkin, Ariana, Danai Papageorgiou, Margaret C. Crofoot, and Damien R. Farine. "Inferring Influence and Leadership in Moving Animal Groups." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, no. 1746 (May 19, 2018): 20170006.
- Taylor, Sunaura. *Beasts of Burden: Animal and Disability Liberation*. New York: The New Press, 2017.
- Torney, Colin J., Myles Lamont, Leon Debell, Ryan J. Angohiatok, Lisa-Marie Leclerc, and Andrew M. Berdahl. "Inferring the Rules of Social Interaction in Migrating Caribou." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, no. 1746 (May 19, 2018): 20170385.
- Tuvel, Rebecca Dayna. "Epistemic Injustice Expanded: A Feminist, Animal Studies Approach." PhD dissertation, Vanderbilt University, 2014.
- Van Schaik, Carel P. "Fragility of Traditions: The Disturbance Hypothesis for the Loss of Local Traditions in Orangutans." *International Journal of Primatology* 23, no. 3 (June 2002): 527–38.
- Van Schaik, Carel P., Marc Ancrenaz, Gwendolyn Borgen, Birute Galdikas, Cheryl D. Knott, Ian Singleton, Akira Suzuki, Sri Suci Utami, and Michelle Merrill. "Orangutan Cultures and the Evolution of Material Culture." *Science* 299, no. 5603 (January 3, 2003): 102–5.
- Warren, Karen J. "The Power and Promise of Ecological Feminism." *Environmental Ethics* 12, no. 2 (1990): 125–46.
- Westley, Peter A. H., Andrew M. Berdahl, Colin J. Torney, and Dora Biro. "Collective Movement in Ecology: From Emerging Technologies to Conservation and Management." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, no. 1746 (May 19, 2018): 20170004.
- Whitehead, Hal, and Luke Rendell. *The Cultural Lives of Whales and Dolphins*. Chicago: University of Chicago Press, 2015.

CRYING HAVOC AND (RE)CLAIMING RIGHTS

HOW THE LIABILITIES OF REVISIONISM AND THE JUST WAR TRADITION ALTER THE MORAL EQUALITY OF COMBATANTS

Marcus Hedahl

IMAGINE a Just War Rip Van Winkle, someone who for one reason or another found themselves temporarily cut off from the theoretical debates on the ethics of war.¹ Unlike the character in Washington Irving's famous tale, such a dreary-eyed theoretician needed to sleep only a little over a decade to miss a revolution, an insurgency in the theoretical rather than the political domain. For in the thirteen years since the publication of Jeff McMahan's *Killing in War*, there has been a seismic shift in debates about the ethics of war, a shift that challenges the foundational assumptions of the just war tradition.² This revisionist project led by McMahan, Helen Frowe, David Rodin, and numerous others begins with the contention that the moral justification for killing in war should be consistent with a broader theory of self-defense.³ In the language of

- 1 This analogy captures, albeit in a rather exaggerated manner, the way I felt after returning to discussions about the ethics of war after more than a decade away. I had served as an instructor of philosophy at the Air Force Academy before separating from the service and then focusing my graduate research not on military ethics, but rather on questions involving claim rights and directional duties. After accepting an appointment at the US Naval Academy, I was momentarily bewildered by the dramatic change in the terminology for debating the ethics of war that had taken place in my absence.
- 2 As Leonard Kahn nicely articulates in "Liability to Deadly Force in War," the claim that unjust combatants have no license to kill in war has its own rich historical roots, going back at least to Pascal's *Pensées* and Voltaire's *Philosophical Dictionary*. Nonetheless, the revisionist view, which had been relegated to the margins of discourse about the killing in war for decades, if not centuries, has ascended to become a viable rival, if not the dominant view among theorists considering the ethical issues of killing in war.
- 3 There are numerous examples within the revisionist tradition. I will consider a number of them throughout the paper. For now, it will be sufficient to note four of the more prominent examples: McMahan, *Killing in War*; Frowe, *Defensive Killing*; Rodin, *War and Self Defense*; and Fabre, *Cosmopolitan War*. Although Rodin's work predates McMahan's, and although some theorists have placed the resurgence in the revisionist tradition at the

the revisionist project, the key question—the only question, really—is whether an agent can be liable to the use of deadly, defensive force.⁴

More recently, the revisionist project has been modestly expanded, from considering how killing could make an agent liable to the use of defensive force to considering how saving or not saving lives could do so as well. With a focus firmly on this new expanded domain, this paper begins fairly modestly. I do not attempt to offer a new theory of liability or even a new principle of liability.⁵ I merely argue for two contentions: that for a principle of liability to be action-guiding, an agent must be violating the claim right of another rather than merely acting unjustly, and that the directional aspects of claim rights will therefore be more significant for settling questions of liability than they are generally taken to be. Within this expansionist project, those moral features imply that saving lives can never, in and of itself, make one agent liable to another.⁶

It is in turning these general insights about liability back on to the core revisionist considerations of killing, however, that those same modest contentions

beginning of the 2003 Iraq War, it was McMahan's work that most radically altered the landscape of the theoretical debates about the ethics of killing in war, turning revisionism from a more isolated project advanced by a few to a genuine—if not even more prominent and frequently considered—alternative to the just war tradition.

- 4 Some theorists classify the revisionist project as any attempt to reconsider any of the traditional just war theory precepts, most notably those put forward by Michael Walzer. Many revisionists, including Frowe and McMahan, are also reductive individualists. They claim that individual liability can be determined by analyzing the actions of those individuals and that the ethics of war is completely reducible to the ethics of self-defense. There are other theorists who are not reductive individualists who are nonetheless often labeled revisionists because they seek to challenge contentions of the current consensus within the just war tradition. The extent to which these theorists ought to count as members of the revisionist project depends in large part on the extent to which they are looking to radically modify the just war tradition, a phrase that, given the ways the principles and their application have changed rather dramatically across time, is better suited to the theoretical history than is the term “just war theory,” the nomenclature often favored by revisionists. In this paper, I am primarily focused on the reductive branch of the revisionist project advanced by McMahan and Frowe. Nonetheless, the arguments contained herein should be equally forceful against any theory that seeks to use the fact that soldiers can be morally responsible for participating in unjust war to conclude the need for different *ius in bello* standards for soldiers on different sides of a war.
- 5 Earlier versions of this paper were, as earlier versions of papers can be, both too bold in their aims and too dismissive of rival views. Several years ago, I presented a paper titled “Claim Rights Based Liability: The Achilles Heel of the Revisionist Just War Project.” To be clear, I am not advancing a principle of liability, nor arguing that this analysis presents an unanswerable challenge for the revisionist project here.
- 6 As I will argue in section 3, there are some rare cases in which saving the life of another could make one agent liable to another, but it is not the saving of the life, in and of itself, that does so. I will have much more to say about these exceptions in section 3.

can have a far more significant impact. Analyzing questions of liability in this new, expanded domain of saving and not saving lives can highlight important distinctions that can be missed when considering the same kinds of actions (i.e., killings) time and time again in cases of self-defense and war. Morally salient elements that would normally be concealed can be revealed, and that insight can be used to consider the paradigmatic case of killing anew. In doing so, it becomes clear that an assumption taken to be so obvious it need not even be acknowledged—that unjust killing necessarily involves the violation of rights—turns out to be false. A more nuanced understanding of rights—the moral element that is meant to be the centerpiece of the revisionist project—ends up exposing a serious limitation with several of its central aims.

To make that argument, this paper proceeds as follows: In section 1, I consider two attempts to expand the core revisionist project, advancing cases from Helen Frowe and Blake Hereth. In section 2, I pause to consider two important theoretical complications necessary to evaluate those claims. In section 3, I argue that since one agent does not have the normative authority to prevent the saving of another's life, saving lives can never, in and of itself, make one agent liable to another. In section 4, I consider how these insights are relevant for analyzing liability in war, arguing that a myopic focus on consent, rather than on the myriad moral authorities possessed by agents, undermines the revisionist claim that soldiers cannot waive, and therefore can only forfeit, their rights in war.

Despite that structure, the central argument of this paper is intended not to serve as an objection to the revisionist project, but rather to offer a critique of both revisionism in its most prevalent form and a particular interpretation of the just war tradition it seeks to replace. By realizing this limitation of the revisionist revolution, a theoretical space opens up in which one could hold on to a significant insight of revisionism, namely that soldiers have significant moral and epistemic responsibilities to avoid fighting in unjust wars, while nonetheless maintaining a significant kind of moral equality regarding the actions of combatants within war. So, in section 5, I analyze how the fact that soldiers can be morally responsible for unjust *ad bellum* wars—even if they turn out to be morally responsible for the very reasons given by revisionists—does not necessitate a change in the moral equality of soldiers embedded within traditional *jus in bello* requirements.

1. EXPANDING THE REVISIONIST PROJECT

In the revisionist project, the key moral question is whether an agent can be liable to the use of deadly force. Trying to unjustly take someone else's life, for instance, can make an agent liable to defensive, coercive, and even violent

actions to thwart that effort. In standard circumstances, for example, if I am unjustly trying to kill you, then I become liable to your use of defensive force. You would be morally justified in using force—even deadly force—against me to try to stop me from doing so, so long as that force were proportionate, necessary, and instrumental in defending yourself against my unjustified attack.⁷ In this first section, I analyze how this broad consensus has led some to consider how a similar analysis could be expanded to cases of saving and not saving lives.

In *Defensive Killing*, Helen Frowe offers an expansion of the revisionist project, arguing that saving lives can make one agent liable to another, with a case involving an attacker, a defender, and a paramedic.⁸

Paramedics Before Police: An Attacker has already killed a Victim's family, but as the Attacker is trying to strangle the Victim, the Victim hits the Attacker over the head, rendering him unconscious. The Victim then calls emergency services, who dispatch both the police (because of the attack) and an ambulance (because the assailant is injured). Unfortunately, the ambulance arrives first, and a Paramedic immediately starts reviving the Attacker. The Victim knows that if the Paramedic is successful, the Attacker will go back to what they were doing before, namely trying to kill the Victim. The Victim warns the Paramedic of her certainty regarding this outcome, but the Paramedic proceeds, noting that it is their job to save lives not to worry about the probable or even the certain outcomes of doing so.

Frowe dismisses the Paramedic's common-sense judgment, concluding that if the Paramedic proceeds in the face of the Victim's objection, the Paramedic is thereby "knowingly contributing to an unjust threat" to the Victim and "renders herself liable to defensive force."⁹ According to Frowe, so long as an agent is responsible for an unjust harm, they then can become liable to the use of deadly force. Frowe concludes that this case helps demonstrate that, in certain circumstances, unjustly saving someone's life could make one liable to the use of defensive force, even deadly force.¹⁰

7 There is some debate in the literature about whether proportionality, necessity, and instrumentality are internal or external to a principle of liability. The argument of this paper is intended to work independently of that distinction; I am focusing here only on cases in which proportionality, necessity, and instrumentality are met.

8 Frowe, *Defensive Killing*, 202. The wording and naming convention are my own, but the case is clearly articulated by Frowe in a case called *Rescue*.

9 Frowe, *Defensive Killing*, 202–3.

10 Cécile Fabre concurs with Frowe's judgments here. See, for example, "Guns, Food, and Liability to Attack in War."

In “Saving Lives, Taking Lives,” Blake Hereth considers a different expansion of the revisionist project, arguing that *not* saving someone’s life can make an agent liable to defensive harm.¹¹

Ruining the Movie Titanic Forever: Rose is in a canoe in the center of an ice-cold lake when she sees Jack, who accidentally fell into the water and is struggling to stay afloat. Jack lacks the strength to pull himself into her canoe, so he asks Rose for help. She refuses Jack’s request but only because she’s always fantasized about seeing someone drown in ice-cold water. (*Titanic* is Rose’s favorite film, but to her great disdain she realizes, “It is only fiction!” Finally, a chance to see it played out before her very eyes, and with a man named Jack nonetheless—quite the lucky day for Ms. Rose.) Yet, as he is about to come to terms with his impending demise, Jack realizes he has just enough strength to flip the canoe, crawl into it, right it, and row it back to shore. He knows he’ll lack the strength to pull Rose back into the canoe, a fact that will ultimately result in her death. Yet he knows as well that Rose could have saved them both and has decided not to do so.

According to Hereth, Rose is liable to the use of defensive force, even deadly force. So, Hereth contends, in at least some cases, unjustly failing to save someone’s life could make an agent liable to the use of deadly force.

2. THEORETICAL BACKGROUND FOR EVALUATING THE EXPANSIONIST PROJECT

Frowe seeks to expand the revisionist framework by claiming that saving another could make an agent liable to the use of deadly force. Hereth seeks to expand the revisionist framework by claiming that *not* saving another could do so as well. Before analyzing those contentions directly, however, it will be helpful first to consider some of the theoretical tools required for that task. In this section, I analyze two such complications. First, I argue that since Frowe and Hereth seek to alter the debate from the possibility of taking lives to the possibility of saving (or not saving) lives, their analysis shifts to a distinct moral domain, one with its own rich and complex normative history. So, as Hereth explicitly and quite insightfully notes, it seems at least prudent, if not necessary, to pause and consider the lessons from the ethical subfield focused on the

11 Hereth, “Saving Lives, Taking Lives.” The wording and naming convention are my own, but the case is clearly articulated by Hereth.

complications involved in saving and not saving lives.¹² Most notable, for our purposes, is the fact that in biomedical ethics the requirements of justice placed upon those giving care do not always correlate to the normative authorities of those who can demand care. Second, building off that distinction, I consider the purpose a principle of liability is meant to fulfill and whether there may be an important difference between the question of how one agent could become liable and the question of how one agent could be liable *to another*.

2.1. Lessons from Biomedical Ethics: Acting Unjustly vs. Violating Claim Rights

In this section, I consider two important lessons from biomedical ethics that will be relevant to considering contentions about liability in this new, distinct domain. First, biomedical includes questions of prioritization—who ought to get priority in the allocation of often scarce resources.¹³ Second, in part because of that fact, in biomedical ethics, the requirements of justice placed upon those giving care do not always correlate to the rights of those who can demand care.¹⁴

Some may be initially skeptical about the claim that acts of injustice need not correspond to rights of others. After all, one of the distinguishing features of duties of justice is that they often correspond to the rights of others.¹⁵ When an agent has been wronged, there may be some who would want to say that situation necessarily violates a right. That may well be true with a broad enough category of rights, yet considerations of prioritization seem to reinforce an appropriate skepticism about the possibility of an absolute correspondence between duties of justice and the existence of a certain kind of right.

When professors prioritize their students for awards or special recognition, for example, there will be all kinds of rights the students have against the professors with respect to how they choose. The students might be able to demand that the professors do all they can to mitigate the systemic and sometimes unconscious influences of race, gender, and culture on their decisions. Students might be able to demand that professors deliberate about such prioritizations

12 Hereth, “Saving Lives, Taking Lives.” Hereth uses their adroit recognition of the overlap of cases of saving lives and biomedical ethics to consider what defensive liability can tell us about biomedical ethics. Here, I take the alternative approach that builds on the same insight, considering what the findings of biomedical ethics can teach us about defensive liability.

13 See, for example, Beauchamp and Childress, *Principles of Biomedical Ethics*, 300–313; Veatch, “Physicians and Cost Containment”; Singer, “Why We Must Ration Health Care”; and Daniels and Sabin, “Limits to Health Care.”

14 See, for example, Beauchamp and Childress, *Principles of Biomedical Ethics*, 300–313; Benner, “Honoring the Good Behind Rights and Justice in Healthcare When More Than Justice Is Needed”; and Peel, “Human Rights and Medical Ethics.”

15 Broome, *Climate Matters*, 52.

in the appropriate manner (e.g., professors should not do so while drunk). It does not seem, however, that even a professor's best student has the kind of right that would allow them to demand an award, if, through some error of judgment, the professor mistakenly chose someone else. After all, the professor's decision is meant to be *constitutive* in some way. The professor's choice is meant to determine who the winner *actually is*, as opposed to the independent background facts that dictate who the winner *ought to be*. None of that ought to imply, however, that the decision is no longer one involving considerations of justice. It would be an injustice to favor a less qualified candidate over a more qualified one, even if the more qualified candidate lacked the normative authority to demand the award.

The same reasoning would apply even more forcefully for decisions regarding the allocation of scarce medical resources. Patients might be able to demand that medical professionals do all they can to mitigate the systemic and sometimes unconscious influences of race, gender, and culture on their decisions. Patients might be able to demand that medical professionals deliberate about such prioritizations in the appropriate manner. It does not seem, however, that any patient has the moral authority to demand that they be seen before someone else. After all, medical decisions during triage are meant to be *constitutive* in some way; they are meant to set a default order of care on which countless medical professionals working together must rely to save as many lives as they can. None of that ought to imply, however, that the decisions of medical triage do not involve important considerations of justice. It would be unjust to prioritize a patient over one more in need of care. The appropriate beneficiary of a distributed good does not necessarily have the kind of right that allows them to demand those goods, the way they might be able to demand other behavior regarding their rights.

By choosing a less qualified candidate or by prioritizing the wrong patient, both the professor and the medical professional are acting wrongly, and they are wronging the person who has been unjustly denied some benefit.¹⁶ Those wronged could, and often do, protest such missteps of judgment. Nonetheless, because these decisions are meant to be constitutive in some way (i.e., the decision itself is supposed to create *pro tanto* reasons for action), the patient and the student lack the kinds of normative authorities they have with respect to other rights they possess, rights that do not require those kinds of acts of constitutive authority in order to provide precise, demandable content. They do not have the same kind of authority over the teacher or medical professional they would if they were *exercising* their rights to refuse medical treatment or to

16 See Thomson, "What Is It to Wrong Someone?"

be addressed respectfully, for example. They may well be able to *complain*, but they cannot *demand as their due*.¹⁷

At issue here is a morally salient feature of a particular kind of right: the normative authority a rights bearer has over a duty bearer. Oftentimes, rights bring with them certain normative powers. A rights bearer can waive their right, they can enforce their rights, they can prioritize their rights, they can demand as their due, and, when rights are violated, they can either waive or enforce duties of compensation.¹⁸ There is a rich and long-standing debate about the source and normative significance of those authorities, with some contending that these normative authorities are the essential element to the existence of rights, and others holding that they are far less significant.¹⁹ Luckily, we need not wade into that disagreement in order to see a clear consensus: there are duties of justice (and not merely beneficence) that do not correlate to a claim that gives a rights bearer normative authority over a specifically addressed agent or agents.²⁰ When they do exist, however, such normative authorities are incredibly important, and exercising those authorities is undoubtedly a way in which agents alter the moral domain. So, without taking a stand on the foundational significance of those kinds of rights, it would be helpful to distinguish them from other, broader considerations of justice. For ease of allocation, we can refer to rights that correspond to such normative authorities as Hohfeldian rights, rights with normative authorities, or simply claim rights.²¹

17 For more on the distinction between complaints and demands, see Hedahl, "The Significance of a Duty's Direction."

18 Hohfeld, *Fundamental Legal Conceptions*; Hart, *Essays on Bentham*; Darwall, "Bi-Polar Obligation."

19 This is typically referred to in the rights literature as a debate between will theorists and interest theorists. Prominent interest theorists include Raz, *The Morality of Freedom*; MacCormick, *Legal Right and Social Democracy*; and Kramer, "Getting Rights Right." Prominent will theorists include Hart, *Essays on Bentham*; Wellman, *A Theory of Rights*; and Steiner, *An Essay on Rights*. As noted previously, thankfully, we need not wade deep into the debate about the source and ultimate significance of normative authorities to grant that when they do exist, they are, in fact, normatively significant.

20 Elizabeth Ashford nicely captures the significance of such duties when she says that one of the important aspects of our duties of justice is to seek institutional reforms that would, "make more determinate the content of [our obligations of justice] by tightening up the allocation of responsibility" ("The Inadequacy of Our Traditional Conception of the Duties Imposed by Human Rights," 220). In other words, we often have duties of justice to create more specific duties (and the corresponding more specific rights) that would endow people with the ability to engage in these kinds of exercises of moral authority, an authority they do not currently possess.

21 When I use the term "claim rights" for the rest of the paper, I intend to imply that these are claim rights with some sort of corresponding normative authority. There may be some

Once that distinction is made, it becomes clear that in cases of saving and not saving lives, violations of justice will necessarily be a broader category than violations of claim rights with corresponding normative authorities.²² In summation, a quick analysis of the domain of biomedical ethics demonstrates what ought to be obvious from the start, that violations of justice and violations of claim rights need not co-travel. In some cases, at least, a rescuer could have a duty of justice to aid another without the existence of corresponding normative authority on the part of the agent who ought to be helped.

2.2. *Why Directional Liability Matters*

Before analyzing whether saving or not saving another could make an agent liable to the use of defensive force, it will also be helpful to pause and consider the more foundational question of what makes one agent liable to another. To

who hold claim rights that can exist without any normative power, or even those who use the term to simply be synonymous with the term “right.” There may even be some subset of those theorists who are engaged in bioethical research who would be skeptical that claim rights and justice do not cut at the same joints. For any such reader, I would encourage them to read any use of the term “claim rights” as the much more cumbersome “claim rights that are accompanied by a corresponding normative authority.” I do not believe that any of the contentions in this paper would hang on that distinction.

- 22 Two things are worth noting here. First, as I will argue more elaborately in section 3.1, this distinction does not rule out by stipulation that saving or not saving another can make an agent liable. An agent may well have these kinds of normative authorities in cases in which other agents have a duty to save them. In cases in which medical resources are not scarce, for instance, patients very likely have the requisite normative authority to demand care from those who are in a position to provide it. Second, prioritization is not the only area of bioethics where considerations of justice do not cut at the same joints as considerations of claim rights and normative authorities. Climate change is, for instance, another public health domain that challenges the possibility of complete correlation between duties of justice and the existence of Hohfeldian claims. When Tuvalu’s 11,000 citizens are forced to leave their flooding country, for instance, it would be an outrage to deny them refugee status, even though there is presently no UN provision for climate refugees. (See McAdam, *Climate Change, Forced Migration, and International Law*.) In fact, we will almost certainly owe them much more than refugee status, for those harmed by anthropogenic climate change have a clear right not to be, and when they are, some form of restitution would be required. (See Buxton, “Reparative Justice for Climate Refugees.”) Unfortunately, however, for those most vulnerable to the most adverse effects of climate change, there simply are not yet sufficiently specific and specifically addressed directed obligations to prevent those harms. Yet that fact deepens rather than diminishes the sense in which justice is not being properly considered and climate victims are being wronged. If the world is aligned such that some duties of justice do not correspond to another agent’s normative authorities, in some cases, at least, we should regard that fact as a further normative failure rather than a reason to be skeptical about the existence of a right at all. I will discuss this complication more in section 4.2.

do so, we should begin by noting that there are a variety of ways in which one agent could be liable to another: liable to monetary compensation, liable to be punished, or liable to defensive harm, to name just a few. In this paper, I am focused exclusively on the liability to defensive harms, arguing that even with that more specific focus, before we can consider the *content* of a principle of liability, we must first consider its *function* and *structure*. More specifically, I argue that it is important to distinguish between principles of defensive liability that seek to determine when an agent is defensively liable full stop, and ones that seek to determine when one agent is defensively liable *to another*.

Now, it is perhaps only a slight exaggeration to say that there are as many principles of liability as there are liability theorists. In broad strokes, however, one can categorize principles of liability as either *purely objective*, *culpability*, or *agent responsibility*.²³ *Purely objective* principles of liability focus solely on states of affairs. In a purely objective principle of liability, an unjust outcome could make an agent liable to defensive violence, regardless of their level of responsibility for creating that particular state of affairs.²⁴ *Culpability* principles of liability, on the other hand, focus on an agent's moral responsibility. An agent is liable only if they have acted wrongly. An unjust attacker, for instance, could be liable to defensive harm only so long as they are morally responsible for the risk to another.²⁵ Finally, *agent responsibility* principles of liability focus on the actions that stem from exercises of moral agency. People can be agentially responsible (and therefore liable) for threats that result from actions they have performed, so long as those actions are ones that could be foreseen to possess some level of risk of unjustly threatening others.²⁶

23 This categorization is not meant to exclude more complicated principles of liability. Frowe, for example, distinguishes between direct threats and indirect threats, and therefore advances a hybrid model that ultimately classifies a broader class of people to be liable than even proponents of purely objective principles of liability do. For more on this, see, Frowe, *Defensive Killing*, 72–87; and Skerker, *The Moral Status of Combatants*, 43–47.

24 Examples here include Thomson, “Self-Defense”; Fabre, *Cosmopolitan War*; and Bomann-Larsen, “Licence to Kill?” Although McMahan has since changed his position, in 1994 he defended an objective principle in “Self-Defense and the Problem of the Innocent Attacker.” Some of these theorists explicitly build off of G. E. M. Anscombe’s “War and Murder.”

25 The best example here is Rodin, *War and Self-Defense*. Although there is some debate on how best to categorize Rodin’s principle of liability (see, for example, Skerker, *The Moral Status of Combatants*, 37–52), his focus on defensive rights and the particular actions that lead to violating them makes Rodin’s theory best classified as a culpability theory of liability, even if one needs only to commit a *pro tanto* wrong. Two less controversial examples of culpability theories of liability include Alexander, “Recipe for a Theory of Self-Defense”; and Ferzan, “Forfeiture and Self-Defense.”

26 The most prominent example is McMahan, *Killing in War*. Other examples include Strawser, “Walking the Tightrope of Just War”; Bazargan, “Killing Minimally Responsible

For the purposes of this paper, rather than simply pick a given principle of liability—or even one from each camp—it will be helpful first to pause and consider the structure and purpose of a principle of liability. In other words, it would be useful to begin not by focusing on the content of any given principle of liability, but by delineating principles by their putative function. Here, revisionists have much more widespread agreement. The purpose of a principle of liability can be summed up as follows: If an agent *A* is liable to the use of defensive force, then any proportionality calculations can have the effects on *A* (either positive or negative) diminished by some factor because *A* is liable.²⁷ Moreover, due to that diminished impact in proportionality calculations, *A* is generally not wronged when harmed.²⁸ The purpose of a principle of liability is, on this approach, to determine what state of affairs would be most morally appropriate from an agent-neutral point of view.

Of course, in order to be action-guiding, any agent neutral conception of liability will—at times, at least—have to incorporate agent relative elements as well. There are surely cases, for instance, in which it would be inappropriate for *B* to treat *A* as if they were liable even if *A* could be considered liable from a purely objective point of view. The most obvious place where those two functions come apart is when one agent does not know that another agent is, in fact, liable. A significant amount of ink has been spilled over how to analyze

Threats”; and Otsuka, “Killing the Innocent in Self-Defense.” Unlike a purely objective principle of liability, an agent responsibility principle would not determine that an agent could be liable if, through absolutely no choice of her own, their continued existence constituted a threat to others (e.g., if they were taken in their sleep and launched toward someone else). However, unlike culpability principles of liability, agent responsibility principles of liability hold that an agent can be responsible (and therefore liable) for putting others in danger, even if the agent had done nothing morally wrong—not even *pro tanto* wrong—by doing so, such as by driving a car.

27 As Rodin nicely puts it, “Proportionality and liability, far from being independent factors, are two manifestations of the same underlying normative relation” (“Justifying Harm,” 79). See also McMahan, *Killing in War*, 15–22. There is, of course, much more we could say. For instance, McMahan helpfully distinguishes between questions of narrow and wide proportionality, in which narrow proportionality only looks at the impact to the defender and the liable party, while wide proportionality looks at all the impacts. As liability theorists disagree in the principles of liability, so too will they disagree about the precise relationship between liability and proportionality, as well as the extent to which the impacts to an agent can be diminished (see *Killing in War*, 20–21). Nonetheless, there is widespread agreement at the level of generality considered above. For more on these complications at a lower level of consideration, see the excellent analysis in Quong, “Proportionality, Liability, and Defensive Harm.”

28 How often *A* could be wronged even though they are liable will depend in large part whether proportionality, necessity, and instrumentality are internal to a principle of liability or external to it.

cases in which agents have limited epistemic access to others' liability.²⁹ In this paper, I am not attempting to offer any new insights on that very long and complicated score.

Instead, I want to focus on another possibility, a possibility not yet considered within the liability literature: cases in which the relational elements between *A* and *B* are morally relevant. Recognizing these elements encourages us to envision another, distinct structure and purpose of a principle of liability: an agent *A* is liable to another agent *B* for the use of defensive force if *B* has some normative authority to amplify or diminish the factor by which the effects on *A* (either positive or negative) are diminished because *A* is liable to *B*.³⁰ Rather than solely trying to determine what state of affairs would be most morally appropriate from an agent-neutral point of view, this approach also considers when it would be morally acceptable for agents to alter their moral deliberations about *A* because of *B*'s moral authority.³¹

While the inclusion of a directional element may be initially resisted by some, doing so actually better aligns with the moral concept at the heart of the revisionist project: liability. In the legal domain in which the concept is most familiar, an agent is always liable to another. In civil law, an agent is liable to another legal person; in criminal law, an agent is liable to the state. Moreover,

29 See, for just a small subset of the examples here, Strawser, "Revisionist Just War Theory and the Real World" and "Defensive Interrogational Torture and Epistemic Limitations"; Guerrero, "Don't Know, Don't Kill"; Lazar, "In Dubious Battle"; Tomlin, "Subjective Proportionality"; and Frowe, "A Practical Account of Self-Defence."

30 There is an interesting metaethical question about what makes self-defense a permission rather than an obligation. One possibility is that agents are allowed but not required to diminish the impacts on others in any proportionality calculations. Another is that an agent ought to diminish the impact to any liable parties but they also have the normative authority to voluntarily diminish the impacts to themselves as well, such that the proportionality calculation ends up being the same as it would be if the other agent were not liable. A third is that an agent should diminish the impact to any liable parties, but because of demandingness considerations, they cannot be obligated to take violent action even if proscribed by an appropriate proportionality calculation. While that metaethical debate is interesting, I do not intend anything in this analysis to rely on its resolution, so I have framed these purposes of liability with the phrase "can be diminished" rather than "should be diminished." Thanks to an anonymous reviewer for helping clarify this point.

31 Although it would take us too far afield from the current analysis to demonstrate, I suspect that this difference (between focusing on what is right from a fully objective point of view and what is morally appropriate from the view of a deliberative and contentious moral actor) is the largest reason why theorists in the ethics of killing and practitioners concerned about the possibility of morally permissible killing have simply stopped engaging with one another. Philosophers engaged with the objective project have much less to learn from the moral complexities faced by those on the ground, and those facing those complexities have little to learn from a novel and detached ideal theory.

with liability, the kinds of normative authorities that often accompany such directional elements are extremely important as well. If the counterparty to one's legal obligation releases one from liability, then the agent is no longer liable. If, for example, a particular US citizen reaches a settlement agreement with the properly authorized IRS agent about their back taxes, they need not concern themselves with whether each and every of her fellow citizens also releases them from their duty to pay the original amount or whether the settlement meets some particular, objective principle of justice. They are simply no longer liable for the debt.

Some may remain skeptical, believing that directionality has no place in considerations of moral, defensive liability. For any such readers, it may be helpful to quickly consider a few variants of another case, Bernard Williams's famous example of Jim and Pedro. After a recent spate of protests, Pedro wants to demonstrate the dangers of such "revolutionary measures."³² He has rounded up twenty protestors to be publicly executed. At the last minute, however, Pedro notices Jim. Since Jim is considered an honored visitor, Pedro offers him the opportunity to participate. If Jim accepts Pedro's offer, Jim will have to kill only one prisoner, but if Jim refuses, Pedro will kill all twenty.³³ Consider first an iteration of this case in which all the prisoners implore Jim *not* to accept Pedro's offer. Although in his original case Williams takes the prisoners' preference for Jim's participation to be "obvious," surely with some imagination and a denial of a Hobbesian worldview, we can conceive of a number of reasons why the prisoners might demand Jim decline.³⁴ They might believe that Jim's participation would make a revolution less likely, they might believe that Jim's complicity would validate the atrocities of the government, or they might believe that Jim's actions would change the very meaning of their deaths. Consider as well a different iteration, one in which one of the prisoners steps forward and volunteers to become a sacrificial lamb. Christine Korsgaard imagines this scenario as one in which the volunteer says, "Go ahead, participate—I forgive you."³⁵ In these two iterations, regardless of how the all-things-considered judgments turn out, many might consider it be more deontically appropriate, or more morally justifiable, for Jim to participate after a prisoner volunteers and says, "Go ahead, participate—I forgive you," and less deontically appropriate, or less morally justifiable, for him to participate after all of the them demand that he refrain. The reason is that Jim does not merely have a duty not to kill;

32 Williams, "A Critique of Utilitarianism," 98.

33 Williams, "A Critique of Utilitarianism," 98.

34 Williams, "A Critique of Utilitarianism," 99.

35 Korsgaard, *Creating the Kingdom of Ends*, 296.

he has a directed duty *to the prisoners* not to do so. Not only can Jim wrong the prisoners by violating their rights, he can further wrong them by acting contrary to their expressed prioritization of the duties owed to them.³⁶ He can wrong them by disrespecting their ability to decide for themselves what matters most to them and their way of life.

Of course, neither Jim nor any of the prisoners are liable to defensive force, but in any plausible principle of liability, Pedro would be. The prisoners would be justified in using force to free themselves from Pedro; and, under certain circumstances, Jim would be justified in coming to their aid. We need not completely unpack the important caveat “under certain circumstances” to realize that just as in the iterations previously considered, the moral authority of the prisoners has some role to play. We also do not need to fully unpack that caveat to see that this directional aspect could contain several morally relevant agent-relative components, not merely because different groups of prisoners may express different prioritizations, but also because they may also have different prioritizations about different people—some may want Jim to refuse to participate but welcome the participation of a fellow villager, for example. Regardless of how the all-things-considered judgments turn out, it would be less deontically appropriate and more difficult to justify Jim’s use of force against Pedro if the prisoners all make it clear that they do not want him to do so, and more deontically appropriate and easier to justify if the prisoners all make it clear that they want him to do so. An easy way to capture that moral significance is to hold that it is not merely the case that Pedro is liable for the use of defensive force; Pedro is liable *to the prisoners* for the use of defensive force.³⁷

36 For a more detailed analysis of the directional aspects of this case, see Hedahl, “The Significance of a Duty’s Direction.”

37 Part of the lack of attention to this distinction can be seen in the revisionists near total focus on *self*-defense. As we saw in section 1, revisionists offer any number of different elaborate cases in which an agent *A* could defend themselves against another agent *B* without pausing to consider that our moral intuitions in cases of self-defense may have as much to do with the limits of morality as with the requirements of morality. If *B* commits an extremely minor moral misdeed, which through some highly unusual set of circumstances makes it the case that *A* will die unless they kill *B*, one may well share the intuition that to deny *A* the moral authority to defend themselves could well be asking too much of *A*. Yet in that particular case, the fact that *A* has the moral freedom to defend themselves need not imply that *C* would also have the authority to intervene on *A*’s behalf. It is quite possible that only a proper subset of cases of legitimate self-defense would permit other agents to intervene on their behalf. This focus on self-defense also provides an interesting contrast with the historical just war tradition. Augustine argues in *Contra Faustum Manichaeum*, for instance, that Christians should always be pacifists with respect to attacks against themselves. It is the duty to aid others who are being attacked, however, that could, in

Let us take a moment to be explicit about what the phrase “A is liable to B” in a directional, action-guiding principle of liability does and does not imply. It does *not* imply that B is the only one who can use force against A, nor that if C uses force against A without B’s consent that C necessarily *wrongs* A. A’s actions may well make it the case that they cannot be wronged. Liability, however, does not merely undercut existing directional duties, i.e., it does not merely highlight the fact that C does not wrong A by harming them. Liability undercuts those directional duties *via* altering the effects on A (either positive or negative) in proportionality calculations.³⁸ If those threatened have some authority over their rights and some authority over whether another can act in defense of them, then they must have some authority—even if it is limited—over the factor by which the effects on A are diminished in proportionality calculations. In this case, the prisoners can make it more or less morally appropriate for Jim to intervene by exercising that authority. The prepositional phrase “to B” in the claim “A is liable to B” is merely meant to capture that particular authority, one generally overlooked in considering the structure and purpose of a principle of liability. The directional elements of a principle of liability imply that in order for agent-neutral liability to become normatively action-guiding, one needs to consider not only the limitations of knowledge, but also any potential exercises of normative authority by those who are unjustly threatened.

At this point, that omission may seem fairly benign. Many may remain skeptical about how frequently these kinds of directional aspects are relevant to appropriately analyzing cases of liability. Regardless of how widespread these considerations may be, returning our focus more specifically to the possibility of defensive liability in cases of saving or not saving lives, it seems prudent to at least consider whether agent-neutral liability would need to be augmented with directional elements in order to become normatively action-guiding. After all, within the domain of saving and not saving lives, those kinds of relational considerations are widely regarded to be extremely normatively significant. Doctors used to lie to patients about their diagnoses, particularly cancer diagnoses.³⁹ Doctors also used to perform painful procedures on competent patients

rare circumstances, require intervention. Defense of others rather than the defense of self was taken to be the foundational moral case on which just war was traditionally built. If an analysis of liability is going to start instead with considerations of self-defense, then a liability theorist has to at least pause and consider whether any cases exist in which an agent could justifiably defend themselves even though others could not justifiably intervene on their behalf.

38 See note 27.

39 See, for example, Beauchamp and Veatch, “Truth Telling with Dying Patients.”

even against their violent objections.⁴⁰ Both of these practices were justified, in large part, because they were believed to bring about the best outcomes from a purely objective point of view. The revolution in biomedical ethics in the latter part of the twentieth century, a revolution that has been embraced by theorists and practitioners alike, is based on considering the duties owed by a given doctor to a given patient in addition to larger questions of justice.⁴¹ In effect, in cases of saving and not saving lives, unless allowances were made for the possibility that agent-neutral liability could be augmented with some directional elements, there would be a serious risk of running afoul of a host of other common-sense contentions of biomedical ethics.

Despite its critical tone, the analysis in this section is not intended to provide an objection against liability theory in general, nor to even begin to settle questions about content. There is no commitment about what makes one liable to defensive harms from either an agent-neutral point of view or a directional, action-guiding point of view. This section is also not meant to undermine the moral significance of agent-neutral liability. It is merely intended to highlight the morally significant directional aspects our relationships with one another will at times possess, a fact that should not be surprising given that this facet of morality is important in countless everyday encounters, and one that is even more important when assuming the moral authority to place another person's life in danger.⁴²

3. LIABILITY IN CASES OF SAVING AND NOT SAVING LIVES

The previous section established two contentions: first, that there is a difference between acting unjustly and violating someone's claim right, a distinction that is particularly salient in cases of saving and not saving lives, and second, that directional action-guiding defensive liability need not rise and fall with agent-neutral defensive liability. In this section, I build on these findings, arguing that directional action-guiding liability requires a violation of claim rights with corresponding normative authority rather than merely the creation of an unjust state of affairs. One agent cannot become defensively liable to another if

40 There are numerous examples here, but Dax Coward is a paradigm case. For more on this case, see Parsi and Winslade, "Why Dax's Case Still Matters."

41 I will return to consider these issues more robustly in section 3.

42 For the remainder of the paper, unless I am considering the views of another author, when I use the term "liability," I am referring to directional action-guiding liability. While I will not always use the phrase "action-guiding liability," I will strive to use wording to emphasize the central component of directional action-guiding liability on which I am focused here, namely that one is liable *to another*.

the latter lacks the moral authority to require the former to alter their behavior. This seemingly straightforward requirement implies that saving someone's life cannot, in and of itself, make someone liable in the way Frowe contends.

3.1. *Not Saving Lives, Normative Authority, and Directional Action-Guiding Liability*

Let us consider first the possibility that *not* saving someone's life could make an agent defensively liable to another. Here, it will be helpful to shift the focus slightly, moving away from Hereth's original question of whether not saving someone's life could make an agent liable to the use of defensive force and toward a conditional contention more significant for the claims considered in this paper: if not saving someone's life could make one agent defensively liable to another, then it is only when doing so violates claim rights. As I argue below, considering cases of not saving lives reveals that the violation of a claim right with corresponding normative authority is a necessary condition for directional action-guiding liability, a finding that has rather significant impacts for the revisionist project.

As we saw in section 2.1, in cases of not providing aid to others, not all acts of injustice involve the violation of a claim right with corresponding normative authorities. It will be helpful, therefore, to begin not with Hereth's case, but with a more straightforward case of medical ethics. Consider, for example, a nurse practitioner who acts unjustly when they misprioritize one patient over another. The nurse, the doctor who relies on their triage, and all the rest of the medical personnel involved in carrying out that prioritization may well be liable if the unjust prioritization involves a clear violation of claim rights, if it were widely known that the hospital never prioritized a non-white patient over a white patient, for example. But some lingering skepticism may well be warranted regarding the question of whether one could treat all of those complicit in an unjust prioritization as action-guiding liable if that unjust prioritization were due to some minor misjudgment that could have and should have been overcome with a little more attention to the particular symptoms of each patient.⁴³

43 Some may believe that in cases like these, agents are liable to a miniscule amount of defensive harm, and so we cannot conclude anything from our intuitions in cases like these since "S is liable to miniscule harms" and "S is not liable to any harms" are nearly indistinguishable. That possibility does raise important questions of agent-neutral liability, but it should not trouble us here. For in considering the question of how liability can become action-guiding, the question becomes, "Can the patient act *as if* the nurse is liable (in any, even minimal way)?" That difference will not be incremental, even if the nurse's responsibility and culpability could be. This distinction in the functional representations of agent-neutral liability and directional action-guiding liability serves as another reason

The primary reason to contend that directional, action-guiding liability requires a claim right violation with corresponding normative authorities is not, however, based on any intuition. The reason to do so is much more straightforward and theoretical. Once the focus has been shifted squarely to action-guiding liability, the requirement that liability to another requires a claim right violation with corresponding authorities ought to follow rather straightforwardly and uncontentiously. It stands to reason that one agent (*A*) cannot be defensively liable to another agent (*B*) for ϕ -ing if the latter (*B*) lacks the moral authority to require the former (*A*) to refrain from ϕ -ing. In other words, it cannot be morally permissible for *B* to use physical force to compel *A* to $\sim\phi$ if *B* lacks the more foundational moral authority over *A*'s actions to demand that they $\sim\phi$. The authority to physically compel someone to act a certain way requires, at the very least, the moral authority to require them to do so.⁴⁴ The cases in which *B* has the moral authority to physically compel *A* to take a given course of action must necessarily be a subset of the cases in which *B* has the moral authority to demand that *A* do so. This analytical insight ought to be at the core of any analysis of action-guiding defensive liability, a requirement that has profound implications.

Before examining those implications more thoroughly in the following subsections, however, we should first pause, noting we are now in a position to see that in Hereth's case *Ruining the Movie Titanic Forever*, Rose may well be liable to Jack. Given the fact that Rose is the only one who could rescue Jack, and given that she could do so with so little cost to herself, Jack may well have a normative authority to demand that he be saved. While I suspect Hereth's case is one in which not saving another's life could make one agent liable to another, more argumentation would have to be provided to demonstrate that Rose not saving Jack would violate a claim right with normative authorities rather than simply being unjust.⁴⁵ While settling that question would take us

to distinguish between these two different modes of defensive liability. Thanks to an anonymous reviewer for this helpful point.

- 44 Alternatively, one must be able to demand for another who has the moral authority to require them to do so. As noted earlier, when third party interventions are warranted is an interesting complication, but they too at least require the violation of *some* claim right, even if not one's own.
- 45 In order to analyze this particular case, much more would have to be said both about the correlativity between claim rights and the duties of other agents. Even granting a particular solution to the correlativity problem would not accomplish the underlying task, however, because the question is not merely when others have a directed duty that corresponds to a claim right, but when an agent has normative authorities associated with a given right. Given Jack's vulnerability and the fact that Rose is in a unique position to save him at little cost to herself, I suspect that, in this case, Jack would have the necessary normative

too far afield from the aims of this paper, we need not do so to notice an equally significant finding: even if, in some cases, not saving could violate a claim right with corresponding normative authorities, the situations in which that would be the case will be limited to cases in which specifically addressed directed obligations exist—whether because of the existence of circumstances (e.g., there are a limited number of people who could save the other), previously accepted obligations (e.g., one is a lifeguard for the area), or obligations that have already been specified and institutionally assigned (e.g., by a government, or by previous agreement). Given Hereth’s attempt to demonstrate merely that an unjust failure to save another *could* make an agent liable to defensive harm, this clarification should not be taken as a repudiation of their central aims. Instead, my hope is that it can be taken simply as a clarification: if an agent could become liable to defensive harm for not saving another, the agent would not merely have to act unjustly, but would also have to violate a claim right with corresponding normative authorities.

The contention defended in this subsection, namely that the violation of a claim right with corresponding normative authority is a necessary condition for action-guiding liability, is meant to be a rather modest one. It need not settle the question of whether not saving someone can make an agent action-guiding liable to defense force. This argument is also not meant to advance any particular principle of action-guiding liability. There may well be other necessary or sufficient conditions. It does not, for instance, settle the debate between *purely objective*, *agent responsibility*, and *culpability* principles. For although the vast majority of claim right violations involve moral responsibility, one can violate a claim right without being culpable for doing so. There’s still plenty of room for debate about whether being morally culpable or agentially responsible is also a necessary condition for action-guiding liability.⁴⁶ Even given those limited

authority over Rose. Demonstrating that fact, however, would take us far too afield from the purposes for this paper. For more on the correlativity between claim rights and duties see, Hedahl, “Directional Climate Justice.”

46 Some may be worried that merely providing a necessary condition for action-guiding liability is insufficient to the task at hand. Whether that concern is reasonable, however, will ultimately depend upon the purpose the condition is meant to serve. To wit, it will be helpful to recall that this analysis is not attempting to offer a new theory of liability, a new principle of liability, or a demonstration that the project of grounding the justification for killing based on a principle of liability is somehow fundamentally flawed. Merely providing one necessary condition might well be insufficient for any of those tasks. The goal here, however, is different: to demonstrate that a better understanding of this requirement will lead to a more nuanced understanding of the limitations of the proposed expansion of the revisionist project—and, ultimately, a more nuanced understanding of the limitations of revisionism itself.

aims, however, this finding has rather considerable consequences: any case of purported liability in which the existence of injustice need not imply the violation of a Hohfeldian claim right would require careful reconsideration to determine whether one agent could be action-guiding liable to another. This would be true regardless of whether an incredibly large number of agents were involved (a possibility I consider in sections 4 and 5) or only a small handful of agents were, as in Frowe's illustrative Paramedics Before Police, the case we turn to consider next.

3.2. *Saving Lives, Normative Authority, and Directional Action-Guiding Liability*

The lesson for defensive liability in cases of not saving lives, namely that violation of a claim right with corresponding normative authority is a necessary condition for directional action-guiding liability, has even greater significance for cases of saving lives. In Frowe's Paramedics Before Police, for example, the Defender clearly has a claim against both the Attacker and the Paramedic not to be killed. Equally clear, however, is that the Defender's claim right does not imply that she has authority over any and all ways in which the Paramedic might aid the Defender. What needs to be analyzed, therefore, is what the Defender's moral authorities entail in this particular case. As I argue in this subsection, a more complete analysis of the relational aspects involved demonstrates that saving lives, *in and of itself*, cannot make one agent action-guiding liable to another.

It would be prudent to begin, however, by noting explicitly that Frowe offers a nuanced and comprehensive theory of liability, distinguishing between direct causes of unjust threats and indirect causes of unjust threats, and between bystanders and observers, to name just a few. Yet as nuanced as her theory is, it is focused exclusively on the injustice of a given threat, rather than on the threatened right of the victim.⁴⁷ For the moment, let us simply grant that Frowe's intricate and detailed arguments are veridical: for individual threats outside of the context of saving lives, Frowe's nuanced account focused on injustice is appropriate—perhaps even superior to accounts focused on rights.⁴⁸ Even granting that fact, however, it becomes quickly evident that in the vast majority of individual cases, the link between violating claim rights and acting unjustly is rather tightly bound, if not complete. In the vast majority of cases in which *A* is defensively liable to *B*, *A* is *both* contributing to an unjust state of affairs *and* violating *B*'s claim right to life. Frowe herself, in fact, motivates the move

47 Frowe, *Defensive Killing*, 1–18.

48 *Defensive Killing* offers a detailed and lengthy criticism of the limitations of Judith Jarvis Thompson's account of self-defense, an account grounded in the rights of the defender.

to analyzing liability in terms of injustice by noting the tight link between injustice and rights, stating, “When the threat [a person] poses is unjust—when it threatens harm to [another] person who has a right not to suffer a harm—such moral responsibility renders the agent liable to defensive force.”⁴⁹ It might be reasonable, therefore, to move more cautiously when transitioning to other domains, especially those in which the links between claim rights and injustice are not so tightly bound.

Returning to the specific case of Paramedics Before Police, it is worth noting what Frowe says in condemning the actions of the Paramedic: “If Paramedic has indeed sworn to do no harm, she must refrain from contributing to the unjust harm that Attacker will pose. The prohibition on causing harm trumps the prohibition on allowing harm.”⁵⁰ Here, we see a consideration of the standard medical injunction of nonmaleficence, but significantly without a citation from a single biomedical theorist or even a philosophical analysis of what nonmaleficence implies in medical contexts. Given the consequent of her conditional, Frowe seems to assume that nonmaleficence requires that medical professionals never engage in activity that would contribute to an all-things-considered unjust harm. Within medical contexts, however, that assumption is simply misguided. In biomedical ethics, while broader considerations of justice always warrant some consideration, the requirement to first do no harm is universally and unequivocally read as “do no harm to *the patient*.”⁵¹ Frowe’s radical and wholesale alteration of the medical principle of nonmaleficence would not merely have implications for the real-world case she has in her sights, but for a host of other cases as well, implications one would assume Frowe would wish to avoid. If the requirement of nonmaleficence required medical professionals to never engage in activity that would contribute to an all-things-considered unjust harm, doctors for tyrants, autocrats, and even the occasional leader of democratic states may well be morally required to refuse to treat their patients because of the all-things-considered unjust harms those leaders would perform if they were to survive.⁵² Doctors of patients who ask for blood transfusions despite the objections from their large, caring family of Jehovah’s Witnesses may well have to weigh all the harms at the all-things-considered level before deciding to proceed. Those possibilities ought to strike

49 Frowe, *Defensive Killing*, 10.

50 Frowe, *Defensive Killing*, 202.

51 See, for example, Beauchamp and Childress, *Principles of Biomedical Ethics*, 155–216; and Pellegrino, “The Moral Foundations of the Patient-Physician Relationship.”

52 More precisely, doctors should refuse treatment if they knew that injustice performed by the current leader would be greater than those that would be performed by their successors.

us as problematic; and, in part, for that reason, biomedical ethicists analyze nonmaleficence, beneficence, and autonomy as ethical considerations *regarding the patient*, considerations that need to be analyzed within a framework that includes larger questions of justice for the rest of us.⁵³ That fact does not imply that larger considerations of justice can never overrule the significance of autonomy, beneficence, and even perhaps nonmaleficence.⁵⁴ It does imply, however, that in considering the ethics of saving and not saving others, one cannot merely look to questions of all-things-considered injustice.

Frowe's case is, of course, more complicated than any standard biomedical case: it involves the possibility of defensive liability, a possibility not generally considered in biomedical ethics. It would be a mistake, therefore, to believe that standard medical ethics could simply provide an answer without any further analysis. Yet it would be a similar kind of mistake to believe that since the vast majority of individual cases of self-defense can be analyzed merely by considering broader questions of injustice, this case can be analyzed that way as well. The ethics of saving lives involves complex directional, relational elements, elements not generally present in other, more standard cases of self-defense. In effect, *Paramedics Before Police* involves a tension, a tension between the action that appears to be required from Frowe's principles of liability, principles that seem extremely well suited for cases of individual self-defense outside of medical contexts, and the action that appears to be required from the dictates of medical ethics, a well-established subfield devoted to analyzing the ethical complications involved in saving lives. Frowe does not merely fail to resolve that tension—she fails to recognize it all. This approach ought to strike us not merely as misguided, but as deeply problematic: it risks a kind of epistemic colonialism, a hubristic determination that decades of debate within a distinct domain of practical ethics has nothing to teach theoretical newcomers on the scene.

How, then, to proceed? How to build off of Frowe's excellent analysis regarding individual self-defense in more traditional cases of killing without

53 One need not endorse Beauchamp and Childress's principles to endorse the conclusion that doctors must consider the harms to their patients in particular and not merely the all-things-considered impacts. See, for example, Pellegrino, "The Moral Foundations of the Patient-Physician Relationship."

54 One of the most classically cited examples of how broader considerations of justice can override patient interests involves the cost of the sixth stool guaiac test for detecting colon cancer (Neuhauser and Lewicki, "What Do We Gain from the Sixth Stool Guaiac?"). Although the analysis of Neuhauser and Lewicki has come under criticism at times (see, for example, Brown and Burrows, "The Sixth Stool Guaiac Test"), the broader philosophical point nonetheless remains: larger questions of justice will, at times, influence ethically appropriate patient care.

simply disregarding the broad consensus of medical ethics? How to determine what the Defender has the moral freedom to do in this situation if the Paramedic were to provide medical aid to the Attacker? The answer seems obvious enough: given Frowe's original analysis about threats that are unjust *and* that threaten rights, given that biomedical ethics has demonstrated that in cases of saving lives the requirements of justice placed upon those giving care do not always correlate to the claim rights of those who can demand care, and given the fact that directional action-guiding liability requires the violation of a claim right with corresponding normative authorities, one should analyze *Paramedics Before Police* not by considering if the case involves any unjust threat, but rather by considering if the Defender has directional normative authority over the Paramedic given the Defender's claim rights.⁵⁵

In terms of the Defender's right to life, it should be obvious that it prohibits the Paramedic from joining the Attacker in a joint intention to kill the Defender. Of course, saving the Attacker's life would almost never fit that description. One obvious exception would be if the Paramedic were to save the Attacker only if he could kill the Defender before the police arrive. In those kinds of rare cases, the Paramedic would be saving the Attacker not *in spite of* his plan to kill the Defender, but rather *because of* it. In that case, the Defender's claim right would give her the authority to stop the aid, justifying her use of force—even deadly force—against the Paramedic.

In more standard circumstances, however, the content of the Defender's claim right does not include prohibiting aid to the Attacker, even if that aid would be ultimately useful to the Attacker's plan to violate the Defender's rights. For if the moral principles of nonmaleficence, beneficence, and patient autonomy are to play any role at all in a medical practitioner's moral deliberations, they must be able to differentiate cases in which the harms are relatively equivalent. Notice that Frowe's case lacks the structure typically present when relational elements of morality are swamped by vastly more significant objective considerations: legal cases in which the negative impacts of violating confidentiality are vastly outweighed by the positive ones that can be achieved by violating it, medical cases in which questions of justice take precedence precisely because the benefits to the patient are so minimal and the benefits to the rest of us are so large.⁵⁶ This is not a case in which were the Rescuer to

55 Victor Tadros has related but distinct concerns about linking self-defense merely with acts of injustice. See, for example, Tadros, "Duty and Liability."

56 Consider again the sixth stool guaiac test in cases of colon cancer. Larger questions of justice take precedence there precisely because the benefits to the patient are so miniscule (the percentage of new patients that are identified with each successive test as having colon cancer decreases exponentially) and the benefits to others are so large (the cost of

forgo lifesaving care, hundreds, dozens, or even several more will be saved; it is merely one in which a more culpable life may be saved.⁵⁷

In other words, this is precisely the kind of case in which the relational aspects of medical care are meant to play a pivotal role in moral deliberation. Medical professionals are not medical professionals full stop. They also have a relational role as medical professionals *for their patients*. If one has the ability to save the life of their patient by providing standard medical care while lacking the ability to save someone else by providing standard medical care, then they should save the life of their patient. No one other than the patient or their proxy has the authority to demand that the Paramedic refrain from doing so. Given these limits of the contents of the Defender's rights, the Paramedic's actions cannot, in general circumstances, make her action-guiding liable to the Defender. In other words, it would not be appropriate for the Defender to use deadly force against the Paramedic. Saving someone's life cannot—*in and of itself*—make someone liable in the way Frowe contends. The reason ought to be straightforwardly evident: no one has a claim right against the continued life of another.

In order to better understand the claim that saving someone's life cannot—*in and of itself*—make someone defensively liable to another, however, a few important clarifications are in order. First, nothing in the argument above implies the that the Attacker is no longer liable to the Defender while the Attacker is incapacitated.⁵⁸ Second, there could be similar cases in which it

the tests does not decrease). If the impacts are anywhere near equal, then the obligation to the patient prevails.

57 One could have the strong intuition that, given the parameters of the case involving individual agents, the Defender must possess the moral freedom to save her own life. I do not share this intuition, but for those who do, one could certainly argue that for morality to demand that the Defender not have the moral freedom to do so would be to demand too much of her. Such an argument, however, would not be based on the Paramedic's liability, but rather on the limits the demands of morality can make on the Defender. That kind of argument would save the intuition that the Defender would be permitted to act to save their own life in this particular case. However, it would not imply that it is permissible for others to come to the Paramedic's aid, nor that such "defensive" actions would be justified when one's *own* personal life were not in obvious and immediate danger (as, for example, in cases in which soldiers fighting an unjust war are receiving medical attention).

58 I believe part of the difficulty with analysis in cases like these lies in conflating two different traditional prohibitions in combat: not attacking incapacitated combatants, including those currently receiving medical attention, and not attacking doctors themselves. While both of these restrictions are important, the philosophical foundations for them are quite different. Killing a culpable attacker while temporarily startled, confused, sleeping, or even while incapacitated is different in kind from stopping them from receiving medical aid, or taking the doctor who administers that aid to become liable by doing so.

would be morally appropriate to kill the Rescuer (if one could save dozens by doing so, perhaps). Third, as is true with most interesting philosophical claims, appreciating the scope the qualifier “in and of itself” is extremely important. As noted earlier, one way saving another could make an agent liable to the use of deadly force is if doing so were their part in a collective endeavor to violate another’s claim rights.⁵⁹ Some might also believe that when one agent thwarts the justified self-defense of another, the former could become action-guiding liable to the latter.⁶⁰ The central claim under contention here, that saving someone’s life cannot—*by itself*—make someone liable, has no direct relevance to possibilities like that. For, if anyone wanted to argue that an intervener could become liable, they could more carefully consider the Defender’s moral authorities in any specific case: given an Attacker’s liability to a Defender, a Defender could at times have the moral authority to prevent people from entering the fray—either on the side of her Attacker or as a neutral third party. Regardless of the ultimate judgment about cases like these, however, we can notice that in such an analysis, it is *not merely* saving another that would make an intervener liable to the use of defensive force, but rather *the manner in which* they did so. Regardless of whether a Defender may have those kinds of rights, however, that kind of justification is simply not available in Frowe’s case. The Rescuer is not thwarting the Defender’s attack. The Rescuer is not inserting themselves into a justified fight. The Defender is not even thwarted from harming the Attacker during the medical intervention.

While perhaps not representative of all possible cases that would fall under the broad colloquial description of “saving the life of another,” Paramedics Before Police is nonetheless an extremely significant one for considerations of liability—for it demonstrates the limits of examining unjust states of affairs to determine action-guiding liability. In standard cases of individual self-defense, the overlap between cases involving injustice and cases involving a violation of a claim right is extremely high—if not complete. In most, if not all, individual cases, if the threat *X* poses to *Y* is unjust, then *X* is also threatening *Y*’s right to life, a right with a host of corresponding normative authorities. As we have seen, however, Frowe’s case is different, and it thereby shows what liability theorists have been doing all along: in considerations of action-guiding liability, the analysis of injustice is merely a stand-in for the existence of normative authorities that generally accompany those unjust states of affairs. In individual cases of self-defense, that difference is almost always inconsequential.

59 For more analysis about the cases in which motives can be inculcating for professionals, see Skerker, *The Moral Status of Combatants*, 74–114.

60 Thanks to an anonymous reviewer for pointing out this possibility.

Nonetheless, being clear about what precisely makes one agent action-guiding liable to another is significant, for there will surely be exceptions to the tight link between injustice and claim rights, exceptions that may well be much more expansive than we might initially realize.⁶¹

There are also important practical implications of Paramedics Before Police. For in the real-world cases with which Frowe's fantastic case are rather loosely aligned, the conclusions have the implication that the corpsman medic who is explicitly treating his fellow countrymen and his fellow countrymen alone could be action-guiding liable to the enemy.⁶² The doctor, who would treat those from either side of a conflict could not be—even if at the moment they are only treating those from one side, and even if the doctor knows that by doing so, the soldiers will eventually go back to the fight. This fact about defensive liability for doctors applies regardless of what country a given doctor calls home and regardless of whether in their heart of hearts, they would prefer for one side to win over the other.

For many, the moral distinction between soldiers who happen to be providing medical care and doctors who happen to be treating soldiers from only one side aligns nicely with their preconceived moral intuitions. This distinction also happens to coincide with more than a century of international law.⁶³ Significantly, however, the alignment with our moral intuitions and laws is the consequence of the argument, rather than one of its premises. Moreover, the defense for this ethical distinction is grounded not in any kind of consequential considerations (e.g., the argument is not that war would be more gruesome if we did not follow this convention), but rather in a more nuanced appreciation of rights: what claim rights do and do not allow rights bearers to demand of others.⁶⁴

Ultimately, Frowe's contention that it could be ethical to kill a paramedic if it were the only way to save your own life may have an initial intuitive appeal for some, but the consequence of the argument's ultimate failure ought to cause a moment of reflection for us all. For if in this case a more nuanced and meticulous understanding of individual rights—the moral element that is meant to be

61 I will return to this contention in section 4.1.

62 The term "could" here is an important qualifier. Most revisionists would contend that only the corpsman medics on the unjust side will be liable to others. I consider this possibility in more detail in section 4.3.

63 While the *Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field* only went into effect in 1950, the protections for medical personnel were also central to the treaty from the First Convention in 1864.

64 Consequential considerations were certainly at the fore historically, but this argument demonstrates one can defend the practice by focusing solely on claim rights.

the centerpiece of the revisionist project—can provide a defense rather than a criticism of traditional just war conventions, then perhaps, just perhaps, there's reason to consider if there are broader lessons lurking here as well.

4. THE LIMITS OF LIABILITY IN WAR

At this point, many may suspect that regardless of how successful the argument has been so far, it offers no problems for the core revisionist project. They might simply assume that intentionally killing someone who has done nothing wrong necessarily involves violating their claim rights. As I argue in this section, however, such initial suspicions would be misguided. Recognizing that directional action-guiding liability requires violating a claim right with corresponding normative authorities rather than merely acting unjustly constitutes a serious challenge for any attempt to use liability to determine the morality of actions within war.⁶⁵ To make that case, I first consider how acting unjustly with respect to a right of bodily autonomy does not necessarily imply that one is violating a claim right. I then build off that contention, investigating how acts of normative authority can be ethically significant even in the face of grave injustice.

4.1. *Unjust Killing Need Not Violate Claim Rights*

It will be useful to begin by considering a right related to the right to life: the right to bodily autonomy. Consider, for example, the case of boxing. As noted countless times before, standard cases of boxing are not analogous to war, because, in boxing, there is typically no underlying act of injustice. In boxing, both parties come to the ring as normative equals, desiring to conditionally waive their right to be assaulted so long as the other party does so as well. In the colloquial parlance, both parties consent to box.

So let us add an underlying injustice to a standard case of boxing. While one professional fighter comes to the ring willing and excited (she would box even if there were no money on the line), her opponent is only there because of underlying issues of systemic injustice: there are simply no other opportunities for employment or meaningful activity for people of her race and socio-economic class. If she did not box, her ability to support herself and family would be severely limited. So she chooses to box rather than to beg, sell herself, or rely

65 The argument in the previous sections was based upon an analysis of why a claim rights violation is an essential element of action-guiding liability for cases of saving and not saving lives. While some elements of that argument are obviously not germane to the question of action-guiding liability for taking lives, a foundational aspect remains: that to be granted the moral ability to stop an agent's actions through force, one must first possess the moral authority to require them to stop.

on the fickle generosity of those with means. She understands the risks, but her options are severely constrained by issues of systemic injustice. She chooses to box, but only over other, even more problematic options.

Unjust Boxing: A and B agree to box one another, but B only got into boxing because of background systemic injustices.

In this case, the proper analysis appears to be that although there are injustices involved, the boxing itself has not changed its normative status. The reason is perhaps clear: the opponent of the boxer who has been forced into the ring has not herself committed any injustice. In a case like this, no plausible principle of liability would suggest that one of the boxers is liable to the other.⁶⁶

Some may initially suspect that Unjust Boxing has not changed its normative status for another reason: that in Unjust Boxing, just like in the standard case of boxing, both A and B *consent* to box. While there may be some colloquial truth to this framing, any robust analysis of consent demonstrates rather quickly that B does not, in fact, consent to box. To see why, we only need to imagine a case in which all of the injustices suffered by B are caused by a single actor. One can imagine a late-in-life, broken, and battered B chastising this miscreant for the lack of opportunities in her neighborhood, the demeaning behavior people had to endure just to survive. B will likely add to the tally of injustices the fact that she herself was forced into a life of boxing, perhaps even adding her post-boxing physical ailments to the list of injuries for which this villain is responsible. If the sole cause of all these injustices tried to reply, “Ahhh, but, B, you consented to box,” our aged, deeply wronged pugilist would almost assuredly respond with, “Consented?! It was box, beg, or steal. You gave me no choice!”

When consenting, one agent (B) sanctions another agent’s (A) ϕ -ing to, on, or with B. Consent makes A’s ϕ -ing morally permissible, and it makes ϕ -ing morally blameless for all involved. These features are clearly evident in any traditional case of consent, cases in which the same descriptive actions that would constitute theft becoming borrowing, or the same descriptive actions that would constitute assault becoming a loving embrace. In Unjust Boxing, on the other hand, while B chooses to box in a way that makes A’s actions morally permissible, B certainly does not knowingly and freely select a given course

66 Some may contend that purely objective principles of liability would entail that A is, in fact, liable to B in this case. For example, Michael Skerker seems to imply that kind of conclusion would be required of any *purely objective* principles of liability (*The Moral Status of Combatants*, 43–47). I take that implication to be less obvious, but if any purely objective principles of liability did have that implication, that would surely count as a strong reason against such a principle.

of action. Nonetheless, there is an important truth captured in the colloquial framing that *B* “consents to box.” After all, *B* does something very much like consent to box *A*. Because *A* was not responsible for the injustice, *A* gets to treat *B* in the boxing ring *as if* she has consented. *A* does not have to take the background systemic social injustices that got *B* into boxing into account while boxing. Those injustices do not make *A* guilty of assault, in part because *A* is not responsible for them. Yet those features of the case do not imply that, in actuality, *B* consents, nor do they imply a problem with the philosophical notion of consent. Rather, these features demonstrate that, even in the face of widespread injustice, agents generally retain certain normative authorities over their claim rights. They still have the authority to exercise those normative powers, and those exercises can be still be normatively significant. These exercises of normative authority can turn *A*’s potential assault into an acceptable form of sport. Our normative authority regarding our claim rights is much broader than our power to consent, a fact that turns out to be extremely important for analyzing morally complex situations involving claim rights, liability, and the use of force.

Of course, not all injustices will be the fault of others. So, let us alter the case of boxing again to one in which one boxer treats the other unjustly. Consider, for example, a title contender who is offered patently unjust terms by the champion to split the purse of a potential upcoming bout. Perhaps a fair split of the purse would be sixty percent for the champion, forty percent for the challenger. But *A* is the champion and has money already, while *B* does not. While *A* could get another fight for almost as much money as she could for fighting *B*, *B* could not get one for $\frac{1}{100}$ the size. So, *A* pressures *B* into an exploitative ninety-five percent to five percent split of the purse.

Unjust Boxer: A acts unjustly toward B before A and B get into the ring.

Does this unjust state of affairs change anything about the claim rights *A* and *B* have against one another once the bout begins? It appears it would not. Neither is violating rights in the ring itself. We do not have the case of an attacker and a defender, even if considerations of justice would dictate who fans of morality should champion. Of course, in this case, while *A* is acting unjustly toward *B*, she is not acting unjustly *qua* pugilist; her injustice has to do with finances, not fighting. So one might worry that the reason *A* has not violated *B*’s right to bodily autonomy in this case is because *A*’s injustice is tied to rights other than the right to choose to fight.

So let us combine these two cases. In this case, right after a boxer becomes the world champion, she personally goes in and destroys the only business in a particular neighborhood so that those from this neighborhood will be forced to box—perhaps because the new champion knows that bouts between her

and those from this rival neighborhood would have a built-in narrative appeal that would boost ratings, regardless of who the fighter will ultimately prove to be. She knows that people will pay more money to see her fight those from this neighborhood, perhaps because of a history of bad blood between that neighborhood and her own. Just as our villainous pugilist had hoped, someone who got laid off on their first week at that business did, in fact, take up boxing, and, several years later, is now about to fight the unjust world champion.

Unjust Boxer Unjust Boxing: *A* is personally responsible for the systemic injustices that got *B* into boxing. *A* is responsible for the underlying injustice that made *B* box at all.

What to say about this case? Those who consider defensive liability only from an agent-neutral point of view may well think the answer is simple: *A* is liable to defensive force, just as she would be in a case of unjustified assault. Equating Unjust Boxer Unjust Boxing to a standard case of assault, however, inappropriately oversimplifies the moral terrain, in part because it proves woefully inadequate for providing action-guidance for *A*. For those who might be initially skeptical, consider the advice that would be appropriate for a suddenly repentant *A*. If *A* suddenly realized the incredible brutality of their old self, if she recognized her own culpability the night before the fight and came to you saying,

I am done with all that repugnant behavior, but how can I even box *B* knowing what I know now? I am responsible for her and countless others whose lives are even worse. I should not get in the ring with *B*. She had no real choice. To box her would be nothing more than piling on a direct assault after years of indirect ones.

Now, I have no idea how I might respond to someone who realizes they are responsible for an injustice of this magnitude, but it does seem that, in some cases at least, to not honor the choices of another because past injustices so severely limited their options does not respect their autonomy, but rather serves as a way to undermine it further. If *A* were to become suddenly repentant, she would have much for which to atone, but it is at least possible that it would be *worse* to fail to respect *B*'s choice to fight, even given *A*'s past injustices.⁶⁷ *B*'s

67 Some might contend that *A* must first withdraw the coercion and ameliorate any past systemic injustice, or at least commit to doing so, before determining if *B* still wants to box. Perhaps that would be true in some cases, but once the relational aspects of the case are recognized, it becomes clear that those cases will have their limits. To see why, we can begin by noticing that *A* cannot merely make a *personal* commitment to ameliorate past systemic injustices; she is required to make that commitment *to B*. However, given the adversarial nature of the fight and the practical significance on *B* of winning (if she wins,

choices can be normatively significant, and *B*'s exercise of normative authority can be normatively significant, even if those choices do not exculpate others of their moral responsibility for placing *B* in that position to begin with. Even if *A* were not merely responsible for coercing *B* into boxing but also morally responsible for coercing *B* into boxing in this particular match, the boxing match does not become one of attacker and defender merely because *A*'s injustice caused *B* to fight. *B*'s choice to fight can be normatively significant; it can be an exercise of *B*'s moral authority even if the world is aligned such that consent (i.e., a choice that exculpates others from the moral responsibilities regarding that choice) is impossible. If *B* were to get injured in the ring or even if she were to die from her injuries, that would be a tragedy but it need not be murder, or even manslaughter. Not all acts of injustice, even those that lead to violence—even those that involve killing—are associated with the violation of claim rights.⁶⁸

4.2. Claim Rights, Moral Authorities, and War

There are many—including the author of this paper—who will worry about the moral nuances that are necessarily eliminated whenever one equates individual cases of violence with the destruction of war. The point of the boxing cases in the previous section is easy to misconstrue. As I argue more fully below, however, the previous subsection is not intended to provide a competing narrative, nor to offer yet another consent-based justification for the moral equality of combatants. Unjust Boxer Unjust Boxing is not about consent; it is about normative powers. It is intended not as a model of war but as a way of recognizing

she can make much more money on future fights than she can if she does not), there will likely be some period of time immediately before the fight in which communicating such an interpersonal commitment would be inappropriate. *B* would rightly regard such a commitment the night before the fight with deep skepticism, for example, fearing it to be disingenuous at best and one of the most perverse forms of gamesmanship at worst. The challenge then becomes how one ought to advise *A* if she has her Road to Damascus moment in which she realizes her own failures during a period in which such interpersonal, directional interactions may themselves be all-things-considered inappropriate. In some cases, to simply refuse to fight may well demonstrate an even *further* disregard for *B*'s autonomy, rather than a first of its kind moment of respect for it. Thanks to an anonymous reviewer for raising this issue.

68 I should note that this is line of argument is not intended as a criticism to any broader conception of liability. It may well be that *A* would be liable to help *B* out of their precarious situation, and that if someone has to suffer to get *B* out of that situation, it should be *A* who does so. That fact does not imply, however, that *A* is liable to *defensive force* if *A* appropriately judges the most morally appropriate course of action, even given all the past injustices, is to box *B*. *A*'s moral culpability does not remove the moral complications created by the boxing match. As I argue in the next section, that claim has much broader implications than one might initially suspect. Thanks to an anonymous reviewer for this point.

more refined requirements for any analysis of liability. The argument does not imply that combatants are blameless, but rather that their exercise of moral authority is nonetheless morally significant—a moral significance that poses a distinct, significant, and to this point unconsidered challenge for applying the concept of liability to the ethics of war.⁶⁹

Let us begin with the most likely misreading of the examples of the previous section: that they are intended to advance a competing analogy. Walzer has the domestic analogy. The revisionists have provided the self-defense analogy. This is an attempt to provide a boxing analogy: soldiers are like boxers. Unsurprisingly, proponents of revisionist theory would likely balk at any analogy to boxing. Or, more accurately, they would balk at the use of such an analogy to describe the vast majority of wars. McMahan actually considers the rare possibility of “a war in which none of the combatants on either side were compelled to fight, either by their adversaries or by their commanders. . . . Wars of this sort are perhaps analogous to situations in which two men agree to ‘step outside’ to settle a dispute by fighting.”⁷⁰ In fact, McMahan even calls this model “The Boxing Model of War.”⁷¹ As the revisionists rightly note, however, this conception of war does not come close to fitting most wars. McMahan puts the point this way, “But many wars are analogous to a different kind of individual combat, in which an unjust aggressor attacks an innocent victim, who is then compelled to defend himself or herself.”⁷² Given that background, a revisionist may read the cases in the previous subsection as advocating that war is best explained by something like the boxing model, or, perhaps more broadly, as advocating that like boxers, soldiers consent to be targeted by those on the other side. Here, I wholeheartedly agree with McMahan and the revisionists: both those possibilities should be soundly rejected.⁷³

Yet Unjust Boxer Unjust Boxing does not involve consent. As mentioned previously, informed consent requires an agent to knowingly and freely choose a given course of action. But Unjust Boxer Unjust Boxing does not do anything like that. It involves coercion and systemic injustice. In Unjust Boxer Unjust Boxing, *A* is deeply culpable for a host of wrongs. Yet even without consent, Unjust Boxer Unjust Boxing does involve the exercise of moral authority. *B*'s decision to box *A*—even if not free—is still normatively significant. It still

69 Thanks to an interaction with Joe Chapa at the International Society of Military Ethicists for elucidating these concerns.

70 McMahan, “The Morality of War and the Law of War,” 25.

71 McMahan, *Killing in War*, 51–58.

72 McMahan, “The Morality of War and the Law of War,” 25.

73 McMahan, *Killing in War*, 60.

changes the way in which one would analyze the moral complexities of the case. It still changes both where the moral misdeed is located and its moral character. The blame for *A* is high—extremely high—but that fact need not imply that what *A* is morally culpable for includes assault and battery.

Some may resist this distinction because they might assume that it necessarily involves letting a coercer off the moral hook—at least somewhat. Many would rightly want to resist that implication when considering cases in which one agent is putting the life of others in danger. But the contention that normative authority is still significant—even when coerced—does not imply that those involved in such coercion are less blameworthy than they would be in more straightforward cases involving claim rights violations. On the contrary, in many cases, it could be far worse to coerce someone to use their own normative authority against their interests than to simply violate their rights.

Consider, for example, a powerful racketeer who realizes that a peasant possesses a rare artifact that the racketeer desperately wants but that the peasant is unwilling to sell. The racketeer knows that, given the corrupt local government and police force, he could easily send his goons to steal the artifact without any fear of repercussion. But, after the racketeer learns that the peasant's sister relies on him for her merger existence, he demands that the peasant make a public display of selling the artifact to him. The racketeer tells the peasant that if he refuses to do so, his sister will be cut off from any possible means of employment and it will become publicly known that the peasant cared more for things, mere trifles, than his own family. In this case, one could argue that the act of selling the art—even if it is not free—is still normatively meaningful, for it alienates the peasant from his claims over the possession in a way stealing the artifact would not.⁷⁴ It is precisely the perversion of the peasant's moral

74 For anyone who is skeptical about this contention, consider what would happen if the government were simply overthrown and a perfectly just government were installed. That new government could very well work to restore stolen goods unpunished by its corrupt predecessor, but the return of items that had been sold would be harder to justify. There is actually an interesting historical analogue here. While much of the artwork that had stolen by Nazi officials in the mid-twentieth century has been returned to the heirs of its rightful owners, the situation is far more complicated in cases in which the art was sold rather than stolen. For example, after realizing that the World War II-era sale of the painting was not freely made, the French government has recently voluntarily returned a Klimt painting it had purchased for the Musée d'Orsay. Yet courts in the Netherlands have held that the Stedelijk Museum does not have to return a Kandinsky sold in 1940 when Holland was under Nazi occupation. Regardless of what anyone may conclude about the appropriateness of either decision, they help illustrate the significance and complexity of normative authority—even in cases in which consent may not be possible. See Breeden, "France to Return Klimt Painting to Rightful Heirs After Nazi-Era Sale"; and Siegal, "Dutch Court Rules Against Jewish Heirs on Claim for Kandinsky Work."

authorities, authorities that are meant to serve the interests of those who possess them, that makes the case so repugnant—perhaps even more repugnant than simply stealing the item would be.

Even more vividly, consider the events of Terrence McNally's *Sweet Eros*.⁷⁵ Early in the play, a disturbed young man kidnaps a complete stranger. Soon after, he rapes her. He then makes her completely dependent upon him, not sleeping with her again until she submits to him completely—that is, until she believes that doing so is what she actually wants. McNally's play was extremely controversial—even for the late 1960s at the Gramercy Arts Theatre—and it remains so to this day. One may question the appropriateness of the play for a number of different reasons, but its artistic power lies in its overt challenge, a challenge to consider where the worst evils of violence lie: in cases in which violence overcomes the most intimate parts of ourselves, or in cases in which violence becomes so pervasive that even our autonomy and moral authority no longer function to protect our own interests but instead serve to protect the interests of those who would do violence upon us. I am sure many will have competing intuitions when considering cases like *Sweet Eros* and the racketeer, but the mere fact of that disagreement ought to demonstrate that the task of determining the location of significant moral transgressions is one that can be severed from the task of determining their severity. In fact, it is only by contending that moral authority matters even in cases such as these, it is only by recognizing that these cases are distinct from more common-place cases of rape and theft that one can begin to capture all of the multifaceted elements of their moral repugnance.

Moreover, Unjust Boxer Unjust Boxing is not intended as a model for understanding war. In fact, the case stems from a deep skepticism about analogies between actions involving individual antagonists and actions within war. Regardless of whether an individual analogy is used to argue in favor or against traditional just war precepts, it must necessarily minimize, if not erase altogether, the magnitude and variation of the destruction inherent in war. War has always been and will always be more than mere fighting; it has always been and will always be more than killing. Nations spend lifetimes trying to heal those whose lives are scarred forever by injuries, both martial and moral. The argument in the previous section does not seek to minimize these distinctions. On the contrary, it seeks to shift our focus and our analyses from our incomplete models back onto war itself by demonstrating that even unjust killing need not involve a rights violation. Since action-guiding liability requires a claim rights violation, those seeking to demonstrate how a principle of liability can

75 McNally, *Sweet Eros*.

be helpful to considering the morality of war must be able to demonstrate how, within war, soldiers violate the normative authorities of soldiers on the other side.

So, Unjust Boxer Unjust Boxing is decidedly not offering a model or a metaphor, nor is it an attempt to demonstrate that soldiers killing in war is somehow less morally blameworthy merely because the killing was done in war. Rather, the case is meant to serve as an indication that the burden of proof for those who appeal to principles of liability in warfare is much higher than has generally been assumed. In order to demonstrate that someone is liable to another to be targeted in warfare, one would need to demonstrate how that person is violating the rights of others. Moreover, because action-guiding liability requires a claim rights violation, in doing so, one cannot merely point to the injustice of the war itself, or to a soldier's complicity for that *ad bellum* injustice.

The most obvious place to find that violation would be in targeting and killing others in war, but the revisionist challenge becomes even more difficult once we recognize that a threatening nation does not use its soldiers to unjustly threaten the lives of the soldiers of the threatened nation. Rather, the threatening nation uses their combatants to unjustly threaten the lives of the combatants of the threatened nation. Even more importantly, so long as the rules of *jus in bello* are followed, in war, that threat, regardless of how unjust it may be, is to combatants *de dicto* rather than combatants *de re*. An unjust mugging involves a *de re* threat, the threat is against a *particular* individual, while an unjust *ad bellum* war that seeks to follow the rules of *jus in bello* involves a *de dicto* threat; it is a threat against a class of individuals one can enter or leave at any time. The threat is still unjust, levels of magnitude more so even than in an unjust mugging, but because it threatens a chosen normative class that any particular individual can leave at any time—even in the middle of battle—it does not involve a violation of individual claim rights.⁷⁶ To notice this fact does not require an appeal to a professional role-based duty that revisionists find so problematic. Uniforms are, in this conception, targets one wears (and can take off) rather than licenses to kill.

Here, too, many would be tempted to misread the claim that war threatens a chosen class that one can leave at any time. This is not a descriptive claim, for as the revisionists rightly note, many are coerced into service with threats to themselves or those they love. Moreover, once at war, nations need not and often do not provide any and all conceivable opportunities to surrender before taking the lives of the combatants. Before an unexpected attack, for instance,

76 The threat may well involve a violation of a collective claim right, but that possibility should not impact this argument.

soldiers do not first announce their presence in order to give their adversaries one last chance to surrender, the way police are generally required to do. Nonetheless, in the right circumstances, individuals have the moral authority to enter the class of combatants, an action that often requires focusing not on the most immediate threat to them as an individual, but rather on the most significant threat to all of them together. Even more significantly, they also have the authority to leave the class of combatants at any time. They have the *moral authority* to lay down their arms, to surrender, to take off their uniforms, to disassociate themselves from any collective violent endeavor.

Neither states nor soldiers consent to war. War is not even the kind of thing to which a nation or a soldier could consent. So it is important to be clear as possible: highlighting the possibility of this kind of normative authority in war does not imply that either states or soldiers consent, nor does it entail an absolution for soldiers on the unjust side of the war who would force others to fight. Forcing others to become combatants is a grave moral misdeed, perhaps one even worse than simply violating their right to life would be. Nonetheless, soldiers may well exercise normative authority through becoming combatants, and, if so, the exercise of that normative authority would itself be normatively significant. Understanding the full complexity of normative authority allows us to recognize the fact that soldiers cannot and do not consent to war does not in any way imply that soldiers cannot waive, and therefore can only forfeit, their rights in war.

5. TWO DIFFERENT WAYS COMBATANTS CAN BE MORAL EQUALS

At this point, some might wonder what the significance might be of denying that soldiers violate claim rights by killing one another in war if the method of doing so grants that soldiers on the unjust side of a war may well commit grave moral offenses, offenses as bad as—if not perhaps even worse than—they would commit if they were violating claim rights. As I argue in this final section, the answer lies in the way this particular argument offers a distinct, and I believe important, way to (re-)consider the moral equality of combatants. Doing so allows the theoretical space to open up in which one can share with the revisionists the contention that soldiers who fight in unjust *ad bellum* wars can be morally responsible for doing so, while maintaining that individual rights often strengthen, rather than undermine, many traditional claims about morally appropriate *jus in bello* rules.

Although that term “the moral equality of combatants” gets bandied about in debates in a way that makes it seem that everyone who argues in favor of it and everyone who argues against it are talking about one unified and universally

recognized principle, there are, in fact, several ways in which combatants could be morally equal in war. McMahan, for instance, states the principle of the moral equality of combatants this way:

Combatants on all sides in a war have the same moral status. They have the same rights, immunities, and liabilities irrespective of whether their war is just. Those who fight in a war that is unjust (“unjust combatants”) do not act wrongly or illegally when they attack those who fight for a just cause (“just combatants”). They do wrong only if they violate the principles governing the conduct of war.⁷⁷

Yet after considering the arguments of the previous sections, it becomes apparent that there are (at least) two separate claims involved in McMahan’s unified definition.

Moral Equality of Combatants:

1. Soldiers are equally blameless for fighting in just and unjust *ad bellum* wars, so long as they follow the rules of *jus in bello*.
2. In general, whether one is fighting a just *ad bellum* war does not influence what actions *in bello* would exacerbate the moral atrocities of an immoral war (i.e., at least some questions of morality *in bello* are distinct from questions of morality *ad bellum*).

Walzer, in *Just and Unjust Wars*, famously argued in favor of both forms of moral equality.⁷⁸ The revisionists have countered by vehemently arguing against both. Recognizing that a claim rights violation is required for directional action-guiding liability, however, allows for a theoretical space to open up between these two all-or-nothing possibilities, a space in which soldiers on one side could be morally responsible for fighting in an unjust war even though the rules of *jus in bello* restrict those fighting on both sides in a relatively similar manner.⁷⁹

⁷⁷ McMahan, “The Morality of War and the Law of War,” 21.

⁷⁸ Walzer, *Just and Unjust Wars*, 34–40. It is worth noting, however, that even within Walzer’s framework, there is a significant exception to the second kind of moral equality, namely cases of supreme emergency. Although never explicitly stated as an exception to the principle of the moral equality of combatants, supreme emergencies are, by definition, going to relax the *jus in bello* requirements *only* for those on the just side who are fighting to maintain their political independence.

⁷⁹ Some may worry that this possibility is self-contradictory; they may believe that it is incoherent to ask what actions could be permissible in an impermissible war. A full consideration of this objection and a reasonable reply would take us too far afield from the current investigation, but we could offer three quick responses to those who may be so concerned. First, some soldiers could be excused rather than justified for participating in unjust wars (a situation to which revisionists frequently appeal—see, for example, McMahan, “The

It is worth noting that like Walzer's argument in favor of both kinds of moral equality, and like the revisionist rejection of them, the possibility of being able to blame soldiers for participating in an unjust war while maintaining that the rules within war possess an important kind of moral equality is also one with a rich and storied theoretical tradition.⁸⁰ In *De Indus*, for example, Vitoria grapples with questions of complicity and permissibility when the military is used for a host of complex reasons, many of which were unjustifiable.⁸¹ Despite its many faults, the Lieber Code signed by President Lincoln, a document that dictates how the Union forces should conduct themselves in a war against their fellow citizens, is essentially framed around trying to come to terms with this essential moral tension.⁸² Perhaps the most memorable example can be found in Cannon Three of Saint Basil, in which he advised that those returning from morally complicated wars, "whose hands are not clean . . . abstain from communion for three years," a recommendation followed for centuries by the Eastern Orthodox Church.⁸³

Morality of War and the Law of War," 26, 42). Second, if some soldiers could be excused rather than justified for participating in unjust wars, then a theory of morality for war ought to be able to provide them with moral guidance, both about how to deliberate (e.g., whether the justness of the war should factor into their deliberations) and what actions would still be morally impermissible even though their participation in the war is excused. Third, that case would be strengthened if one also believed (as many do) that one could morally participate in a war in which there were some element of doubt about whether the war was, in fact, just.

80 Walzer's historical analysis of a defense of Moral Equality of Combatants 1 and 2 can be found *Just and Unjust Wars*, 34–40. For a historical analysis of the rejection of Moral Equality of Combatants 1 and 2, see Kahn, "Liability to Deadly Force in War," 13–32.

81 Vitoria, "De Indus."

82 Hartington, *Military Rules, Regulations, and the Code of War*.

83 Basil of Caesarea, *Canonical Letters (Letters 188, 199, and 217)*—more specifically, Letter 188, #12. Of course, St. Basil was not advancing the kind of third way considered here; he was not contending that soldiers could bear some responsibility for participating in an unjust *ad bellum* war without changing the *in bello* actions that would be permissible within war. This particular provision may well stem instead from the Old Testament laws that equate contamination with bodily fluids with being ritually impure. (See, for example, Thomas, "Unjust War and the Catholic Soldier.") Regardless of the foundations for the dictum, depriving someone of communion is a significant restriction, an acknowledgement that the penitent remains unworthy in the eyes of the church. Therefore, regardless of its ultimate purpose, St. Basil's requirement provides a helpful reminder of a venerable, historical tradition that held that while soldiers were not murderers, their hands could still be dirty. It is up to more modern scholars to determine if there's good reason to maintain that tradition focused on a more figurative, rather than a more literal, interpretation of the moral issue at hand. Thanks to Michael Skerker for this clarification.

While much more analysis than space allows would be required for a full-throated defense of the possibility that Moral Equality of Combatants 2 obtains while Moral Equality of Combatants 1 does not, it would be helpful before closing to briefly consider a few reasons why some might find such a possibility worthy of further analysis.⁸⁴ First, denying Moral Equality of Combatants 1 implies soldiers are morally required to do all they can to avoid fighting in an unjust war, for their participation would be a moral taint of unimaginable proportions. One can thereby capture the underlying moral truth of Voltaire's famous quip, "It is forbidden to kill; therefore, all murderers are to be punished unless they kill in large numbers and to the sound of trumpets."⁸⁵ Denying Moral Equality of Combatants 1 captures a significant and fundamental truth of the revisionists project: it is folly to believe that the mere mention of the term "war" somehow magically absolves those who participate in it from deep moral inquiry about the tragedies they help unleash on the world.

Yet recognizing that questions about soldiers' responsibility for unjust *ad bellum* wars need not rise and fall together with questions of who is a legitimate target within war allows theorists to accept that central revisionist insight without committing to all revisionist *jus in bello* conclusions.⁸⁶ While there is a wide variety of revisionist views regarding *in bello* rules, many revisionists have argued that the concept of liability creates a significant problem for the traditional distinction between combatants and noncombatants.⁸⁷ Some have

84 While it is my hope that this paper can serve as a sketch of such an argument, I do not believe that one can both argue why this option is possible and why it is preferable in a single paper.

85 Voltaire, *Philosophical Dictionary*, 322. This passage is quoted by Kahn, "Liability to Deadly Force in War," 18.

86 To be clear: if Moral Equality of Combatants 2 obtains while Moral Equality of Combatants 1 does not, that would not imply that those fighting on the unjust side get to consider those fighting on the just side to be liable to being killed. On the contrary, this analysis implies that while those fighting a just *ad bellum* war are not liable, killing them would nonetheless not exacerbate the moral atrocities of engaging in an immoral war because the soldiers on the just side of the war exercise their own moral authority by joining a war that seeks inherently collective ends, rather than merely fighting for themselves as individuals. In short, this possibility requires a return to considerations of who can be treated as a combatant rather than considerations of who is liable in order to determine what actions would be morally permissible *in bello*.

87 For an excellent analysis of the wide variety of revisionist views regarding *in bello* rules, see Lazar, "Evaluating the Revisionist Critique of Just War Theory." For an example of the analysis of the tension created by liability on the distinction between combatants and noncombatants, see McMahan, "The Morality of War and the Law of War," 21–22. Even those revisionists who deny this principle in practice, however, will still have issues of parsimony. For more on this issue, see note 91.

argued more specifically, for instance, that the list of those who can be targeted by those fighting a just war ought to be radically expanded.⁸⁸ Now, many will take avoiding these possibilities as a benefit in and of itself.⁸⁹ Others will counter that being less revisionary cannot, by itself, be considered an advantage when the question under consideration is whether those traditional practices are, in fact, justified.⁹⁰ Yet even for those who would take the latter position, the possibility of avoiding the revisionists' *ius in bello* conclusions still offers at least four important potential theoretical advantages.

The first advantage is parsimony. If one grants that the current international legal regime regarding the prohibitions of actions during war has some moral benefit, then, all things being equal, it would be preferable to be able to defend both that legal regime and our ethical judgments about liability for defensive force with a single, unified philosophical framework.⁹¹ The second advantage

88 See previous note. With those caveats, however, a good example here is Frowe, "Intervening Agency and Civilian Liability."

89 A good example here would be Skerker, *The Moral Status of Combatants*.

90 See Frowe, "CEPPA Chats - Helen Frowe Talks Just War Theory."

91 The antecedent of this condition is itself debatable for some revisionists. Nonetheless, many will grant the usefulness of some current legal protections. See, for example, McMahan, *Killing in War*, 203–13. Even those who grant the truth of the antecedent, however, may object to this characterization, contending that there's a straightforward sense in which McMahan is using a single framework focused on minimizing rights violations and injustice. To do so, the framework distinguishes between simple, first-order moral problems (which can be framed without worrying about how to build institutions to handle them), and second-order institutional moral problems where more complex factors of evidence, reasonable disagreement, methods for solving those disagreements, allocating authority to make tough calls, and worries about the misuse of authority have to be taken into account. While that is a reasonable characterization of McMahan's account, the framework remains less paraspinous and unified than the alternative being proposed at both the theoretical and practical level. In McMahan's account, at a theoretical level, the authority of intuitions is indirectly and epistemically grounded. Any framework with that structure would have to do more—much more—to demonstrate the grounding for that very different kind of authority, epistemic authority. That is an activity that, to be done appropriately, would require much more analysis into much deeper questions of epistemology and more empirical work into investigating under what conditions that kind of epistemic authority turns out to have any kind of veracity. All things being equal, it would be better to ground that authority, if possible, on the kinds of moral authorities that are a significant element of claim rights themselves. Regardless of the parsimony on the theoretical level, however, an approach that seeks to demonstrate how Moral Equality of Combatants 1 and Moral Equality of Combatants 2 need not rise and fall together has an immense advantage in simplicity at the practical, action-guiding level. After all, McMahan's theory holds that soldiers often should act as if the distinction between combatants and noncombatants matters—even in some cases in which it does not. This would place an extreme epistemic moral burden on those already susceptible to immense moral decision fatigue. All things being equal, it would be better to

lies in avoiding the possibility that morality would have to conceal itself. If there is evidence to suggest that it is easier to convince soldiers that there ought to be moral limits on their behavior even in a just war than to convince them that their country is engaged in an unjust war even when it is, then in order to avoid even greater moral tragedies, revisionism runs the risk of having to become self-effacing.⁹² In other words, even if the revisionists are right that there is no difference between combatants and noncombatants, morality may nonetheless require us to tell soldiers that there is, in fact, such a difference. A moral framework that could justify the central revisionist insight about soldiers' moral responsibility for engaging in unjust wars without having to do so offers another theoretical advantage. Third, the conception of moral authority that underlies the distinction between the Moral Equality of Combatants 1 and the Moral Equality of Combatants 2 can capture the normative significance of countless activities that men and women fighting against obviously unjust wars of aggressions, such as those fighting for Ukraine in 2022, take to be morally meaningful: their military training (regardless of how truncated or rudimentary), their uniforms (regardless of how haphazard), their placement within a military hierarchy (regardless of how federated), and, most particularly, their authorization to fight *for* their nation, not merely in pursuit of its collective ends but as part of its authorized and justified defense force.⁹³

Finally, and perhaps most notably, a distinction between different types of moral equality could offer its own important insight about a particular subset of *jus in bello* rules. Consider, for example, the numerous moral prohibitions against targeting soldiers who are not *currently* combatants even though they have been in the past: one should not target soldiers in hospitals, soldiers who are surrendering, soldiers who are prisoners of war, etc.⁹⁴ Recognizing that war ought to target only combatants, a category that agents must maintain the moral authority to enter or leave, helps elucidate why the commando order, why crying havoc and releasing the dogs of war, why the pardoning of war crim-

be able to present a unified moral framework that can be presented to soldiers as a unified, action-guiding theory for recognizing their actual epistemic and moral responsibilities. Thanks to an anonymous reviewer for helping me make these distinctions more precise.

92 For an excellent analysis of how and why this happens, see Luban, "Integrity."

93 I am thankful for an anonymous reviewer for this point.

94 If these prohibitions are justified merely because of instrumental reasons, they cannot maintain their necessary deontic force, for soldiers will rightly recognize countless exceptions when targeting prisoners or hospitals as both necessary and proportionate. Although it would take too long to demonstrate, I suspect that for this very particular subset of *in bello* rules, those kinds of problematic, instrumental kinds of justifications are the only ones available to both the revisionists and those with a Walzer-inspired interpretation of the just war tradition.

inals who kill prisoners, all strike us as so deeply morally abhorrent—regardless of whether they are done in support of a just cause.⁹⁵

Many have analyzed the ways in which these rules make war less gruesome, the ways in which nations have agreed to follow these rules, and the ways in which the codified nature of these rules makes it easier to find soldiers culpable of any wrongs they may commit. Yet none of these moral considerations highlight the deeper, underlying moral issue: why do so many believe not merely that those rules and agreements make war less horrific, but also that they capture an inherent underlying moral requirement of warfare, such that any system that gave combatants the ability to simply “Cry Havoc,” publicly announcing that they were no longer following war conventions, and that therefore their adversaries were no longer required to do so either, could be criticized morally and not merely prudentially?⁹⁶

An argument based on the moral authority of combatants can answer that challenge. As argued in the previous section, one way to avoid the revisionist claim that actions that contribute to an unjust war would necessarily violate claim rights is to highlight the moral authority agents possess to leave the category of those threatened by an unjust war. Given that argument, if enough combatants were to deny their adversaries this moral authority, they would thereby irrevocably taint all those who fight alongside them. If those soldiers were on the unjust *ad bellum* side, they would *all*—not merely those who violate the rules of *ius in bello*—*all* become no better than thugs, mobsters, and murderers, turning their uniforms and medals into nothing more than ignominious window dressing.

6. CONCLUSION

This paper began by considering how saving or not saving another could make an agent liable to the use of deadly force. That was not some extraneous endeavor, for it is in these cases that important but generally overlooked moral requirements for defensive liability become recognizable and distinct; it is in

95 The Commando Order was a 1942 communiqué from the Nazi government ordering its soldiers to kill any captives caught behind enemy lines. For more, see “Commando Order” in *The Oxford Companion to World War II*, edited by I. C. B. Dear and M. R. D. Foot (New York: Oxford University Press, 2014). “Cry Havoc and let slip the dogs of war” is a famous line from Shakespeare’s *Julius Caesar*, act 3, sc. 1, l. 273. In the Middle Ages, the order “Havoc” was given to soldiers to encourage them to pillage, create chaos, and—most relevant for our purposes—take no prisoners. Finally, for a powerful criticism of President Trump’s pardoning of Chief Petty Officer Eddie Gallagher, see Kaurin and Strawser, “Disgraceful Pardons.”

96 I am thankful for an anonymous reviewer for this point.

these cases that the difference between acting unjustly and violating a claim right becomes most vividly clear. Only after recognizing the varied and rich complexity contained within those distinctive cases of defensive force can we consider the even more complicated case of killing in war, armed with the recognition that the cases in which an agent has the moral authority to physically compel another to take a given course of action must necessarily be a subset of the cases in which they have the moral authority to demand compliance. Those findings do not merely imply that saving a life can never make one agent liable to another, they also highlight a significant limitation for any theory that seeks to employ the concept of liability for analyzing the ethics of war.

Like actual revolutions, this revolution in the ethics of war has been exciting for many, troubling for others. Perhaps the greatest reason for excitement, the greatest promise of this particular insurgency, has been its challenge to the soldiers who actively participate in war, a challenge for them to reconsider their own complicity in the massive, often incalculable injustices of war. Yet this very promise lies in tension with its overidealized approach. Sadly, one of first casualties in the revisionist rebellion has been the ability for theorists and practitioners to listen to and to engage with one another, a wound that is doubly tragic given the fact that until very recently, that kind of engagement was a genuine strength of considerations in the ethics of war.⁹⁷

My lasting hope is that this paper has done enough to demonstrate that this promise of revisionism can be achieved while avoiding the schism it has created between practitioners and theorists. One need not simply assume that the duties of the soldier are necessarily distinct from the duties of a civilian to believe there is a need for taking the moral complications of war—as they are experienced—as an important input to the theoretical work of unpacking the ethics of war.⁹⁸ One need not beg the question on role responsibilities to analyze the way warfare often requires the exercise of normative authorities in ways that individual killing does not.

While the arguments in this paper have been, in many ways, critical of the revisionist project, my hope is that it can nonetheless be read as deeply respectful of one of its central aims. The just war debate cannot and should not return

97 Notably, this is a strength that has been admired and even attempted to be replicated in most of the successfully robust subfields of practical ethics (e.g., biomedical ethics, environmental ethics, etc.).

98 It is certainly true that this kind of analysis has been sadly missing from this paper as well, but I believe that to best demonstrate the limitations of thought experiments and individual cases to discern the ethics of war, it was necessary to highlight the limitations of that approach through a kind of *reductio*, taking on board as many of the assumptions, methods, and theoretical maneuvers of the revisionists as possible.

to where it was in when our Just War Rip Van Winkle first dozed off a little over a decade ago. There is another way forward: one can hold on to one of the core aims of the revisionist project, its weary skepticism of the blank pass given by some traditional just war theorists to those who fight in unjust wars, while maintaining that individual rights often strengthen, rather than undermine, many of the traditional *in bello* rules.⁹⁹

US Naval Academy
marcushedahl@gmail.com

REFERENCES

- Alexander, Larry. "Recipe for a Theory of Self-Defense: The Ingredients and some Cooking Suggestions." In *The Ethics of Self Defense*, edited by Christian Coons and Michael Weber, 20–50. New York: Oxford University Press, 2016.
- Anscombe, G. E. M. "War and Murder." In *Nuclear Weapons: A Catholic Response*, edited by Walter Stein, 43–62. New York: Sheed and Ward, 1961.
- Ashford, Elizabeth. "The Inadequacy of Our Traditional Conceptions of the Duties Imposed by Human Rights." *Canadian Journal of Law and Jurisprudence* 19 (2006): 217–35.
- Basil of Caesarea. *Canonical Letters (Letters 188, 199, and 217)*. Translated by Andrew Raddle-Gallwitz. In *The Cambridge Edition of Early Christian Writings*, vol. 2, edited by Ellen Muehlberger, 143–67. Cambridge: Cambridge University Press, 2017.
- Bazargan, Saba. "Killing Minimally Responsible Threats." *Ethics* 125, no. 1 (October 2014): 114–36.

99 This paper has benefited more from critical feedback than any I have written, in part because of the length of time it has been in development. The genesis for this analysis stems from 2003, when I was serving on active duty and teaching military ethics during the commencement of an unjust *ad bellum* war. I am thankful for my colleagues at the Air Force Academy and those at the Joint Services Conference on Professional Ethics in 2006, both of whom helped highlight the complex ways in which individuals might be responsible for unjust *ad bellum* wars. I am also extremely grateful for engagement with Blake Hereth, without whom this paper would not exist in this current form. Finally, I am indebted to those who helped improve this paper over the last several years, including those at the US Naval Academy, especially Mitt Reagan, David Luban, Ed Barrett, and Michael Skerker, those at the 2019 International Society of Military Ethics, especially Joe Chapa, those at the 2019 Rocky Mountain Ethics Congress, especially Ryan Mott and Patrick Smith, and several anonymous reviewers at *JESP* who made significant contributions to the final product.

- Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*. 8th ed. New York: Oxford University Press, 2019.
- Beauchamp, Tom L., and Robert M. Veatch. "Truth Telling with Dying Patients." In *Ethical Issues in Death and Dying*, 2nd ed. New York: Pearson Press, 1995.
- Benner, Patricia. "Honoring the Good Behind Rights and Justice in Healthcare When More Than Justice Is Needed." *American Journal of Critical Care* 14, no. 2 (March 2005): 152–56.
- Bomann-Larsen, Lene. "Licence to Kill? The Question of Just vs. Unjust Combatants." *Journal of Military Ethics* 3, no. 2 (2004): 142–60.
- Breeden, Aurelien. "France to Return Klimt Painting to Rightful Heirs after Nazi-Era Sale." *New York Times*, March 15, 2021. <https://www.nytimes.com/2021/03/15/arts/design/france-klimt-painting-restitution.html>.
- Broome, John. *Climate Matters: Ethics in a Warming World*. New York: W. W. Norton and Company, 2012.
- Brown, Kaye, and Colin Burrows. "The Sixth Stool Guaiac Test: \$47 Million That Never Was." *Journal of Health Economics* 9, no. 4 (1990): 429–45.
- Buxton, Rebecca. "Reparative Justice for Climate Refugees." *Philosophy* 94, no. 2 (April 2019): 193–219.
- Daniels, Norman, and James Sabin. "Limits to Health Care: Fair Procedures, Democratic Deliberation, and the Legitimacy Problem for Insurers." *Philosophy and Public Affairs* 26, no. 4 (October 1997): 303–50.
- Darwall, Stephen. "Bi-Polar Obligation." *Oxford Studies in Metaethics*, vol. 7, edited by Russ Shafer-Landau, 333–67. Oxford: Oxford University Press, 2012.
- Fabre, Cécile. *Cosmopolitan War*. Oxford: Oxford University Press, 2012.
- . "Guns, Food, and Liability to Attack in War." *Ethics* 120, no. 1 (October 2009): 36–63.
- Ferzan, Kim. "Forfeiture and Self-Defense." In *The Ethics of Self-Defense*, edited by Christian Coons and Michael Weber, 233–53. New York: Oxford University Press, 2016.
- Frowe, Helen. "CEPPA Chats - Helen Frowe Talks Just War Theory." University of St. Andrews, April 29, 2020. YouTube video, 14:23. <https://www.youtube.com/watch?v=bhy8Za7fhpA>.
- . *Defensive Killing*. New York: Oxford University Press, 2014.
- . "Intervening Agency and Civilian Liability." *Criminal Law and Philosophy* 16 (2022): 181–91.
- . "A Practical Account of Self-Defence." *Law and Philosophy* 29, no. 3 (May 2010): 245–72.
- Guerrero, Alexander. "Don't Know, Don't Kill: Moral Ignorance, Culpability,

- and Caution." *Philosophical Studies* 136, no. 1 (October 2007): 59–97.
- Hart, H. L. A. *Essays on Bentham*. Oxford: Clarendon Press, 1982.
- Hartington, Richard. *Military Rules, Regulations, and the Code of War: Francis Lieber and the Certification of Conflict*. New York: Routledge University Press, 2010.
- Hedahl, Marcus. "Directional Climate Justice: The Normative Relationship between Moral Claim Rights and Directed Obligations." In *Choosing a Future: Social and Legal Aspects of Climate Change*, edited by Anna Grear and Conor Gearty, 35–49. Northampton, MA: Edward Elgar Publishing, 2014.
- . "The Significance of a Duty's Direction: Claiming Priority rather than Prioritizing Claims." *Journal of Ethics and Social Philosophy* 7, no. 3 (September 2013): 1–28.
- Hereth, Blake. "Saving Lives, Taking Lives." Unpublished manuscript.
- Hohfeld, Wesley Newcomb. *Fundamental Legal Conceptions, As Applied in Judicial Reasoning*. Edited by Walter Wheeler Cook. New Haven: Yale University Press, 1919.
- Kahn, Leonard. "Liability to Deadly Force in War." In *Who Should Die: The Ethics of Killing in War*, edited by Ryan Jenkins, Michael Robillard and Bradley Jay Strawser, 13–32. New York: Oxford University Press, 2018.
- Kaurin, Pauline, and Bradley Strawser. "Disgraceful Pardons: Dishonoring Our Honorable." *War on the Rocks*, November 25, 2019. <https://warontherocks.com/2019/11/disgraceful-pardons-dishonoring-our-honorable/>.
- Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.
- Kramer, Mathew. "Getting Rights Right." In *Rights, Wrongs and Responsibilities*, edited by Mathew Kramer, 28–95. London: Macmillan, 2001.
- Lazar, Seth. "Evaluating the Revisionist Critique of Just War Theory." *Daedalus* 146, no. 1 (Winter 2017): 113–24.
- . "In Dubious Battle: Uncertainty and the Ethics of Killing." *Philosophical Studies* 175, no. 4 (April 2018): 859–83.
- Luban, David. "Integrity: Its Causes and Cures." *Fordham Law Review* 72, no. 2 (2003): 279–310.
- MacCormick, Neil. *Legal Right and Social Democracy*. Oxford: Oxford University Press, 1982.
- McAdam, Jane. *Climate Change, Forced Migration, and International Law*. Oxford: Oxford University Press, 2012.
- McMahan, Jeff. *Killing in War*. New York: Oxford University Press, 2009.
- . "The Morality of War and the Law of War." In *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, edited by David Rodin and Henry Shue, 19–43. New York: Oxford University Press, 2008.

- . “Self-Defense and the Problem of the Innocent Attacker.” *Ethics* 104, no. 2 (January 1994): 252–90.
- McNally, Terrance. *Sweet Eros: Acting Edition*. New York: Dramatist Play Service, 1968.
- Neuhauser, Duncan, and Anne Lewicki. “What Do We Gain from the Sixth Stool Guaiac?” *The New England Journal of Medicine* 233, no. 5 (July 31, 1975): 226–28.
- Otsuka, Michael. “Killing the Innocent in Self-Defense.” *Philosophy and Public Affairs* 23, no. 1 (Winter 1994): 74–94.
- Parsi, Kayhan, and William Winslade. “Why Dax’s Case Still Matters.” *American Journal of Bioethics* 19, no. 9 (2019): 8–10.
- Peel, Michael. “Human Rights and Medical Ethics.” *Journal of the Royal Society of Medicine* 98, no. 4 (April 2005): 171–73.
- Pellegrino, Edmund D. “The Moral Foundations of the Patient-Physician Relationship.” In *Military Medical Ethics*, vol. 1, edited by Thomas E. Beam and Linette R. Sparacino, 3–22. Washington, DC: Office of the Surgeon General, 2003.
- Quong, Jonathan. “Proportionality, Liability, and Defensive Harm.” *Philosophy and Public Affairs* 43, no. 2 (Spring 2015): 144–73.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.
- Rodin, David. “Justifying Harm.” *Ethics* 122, no. 1 (October 2011): 74–110.
- . *War and Self Defense*. New York: Clarendon Press, 2005.
- Siegel, Nina. “Dutch Court Rules against Jewish Heirs on Claim for Kandinsky Work.” *New York Times*, December 16, 2020. <https://www.nytimes.com/2020/12/16/arts/design/kandinsky-stedelijk-museum-restitution.html>.
- Singer, Peter. “Why We Must Ration Health Care.” *New York Times Magazine*, July 15, 2009. <https://www.nytimes.com/2009/07/19/magazine/19healthcare-t.html>.
- Skerker, Michael. *The Moral Status of Combatants: A New Theory of Just War*. New York: Routledge Press, 2020.
- Steiner, Hillel. *An Essay on Rights*. Oxford: Blackwell, 1994.
- Strawser, Bradley. “Defensive Interrogational Torture and Epistemic Limitations.” *Public Affairs Quarterly* 27, no. 4 (October 2013): 311–34.
- . “Revisionist Just War Theory and the Real World: A Cautiously Optimistic Proposal.” In *The Routledge Handbook of Ethics and War: Just War Theory in the 21st Century*, edited by Fritz Alhoff, Nicholas G. Evans, and Adam Henschke, 76–90. New York: Routledge, 2013.
- Tadros, Victor. “Duty and Liability.” *Utilitas* 24, no. 2 (June 2012): 259–77.
- . “Walking the Tightrope of Just War.” *Analysis* 71, no. 3 (2011): 533–44.
- Thomas, Ward. “Unjust War and the Catholic Soldier.” *Journal of Religious Ethics*

- 35, no. 3 (September 2007): 509–25.
- Thomson, Judith Jarvis. “Self-Defense.” *Philosophy and Public Affairs* 20, no. 4 (Autumn 1991): 283–310.
- Thomson, Michael. “What Is It to Wrong Someone? A Puzzle about Justice.” In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 333–84. New York: Clarendon Press, 2004.
- Tomlin, Patrick. “Subjective Proportionality.” *Ethics* 129, no. 2 (January 2019): 254–83.
- Veatch, Robert M. “Physicians and Cost Containment: The Ethical Conflict.” *Jurimetrics* 30, no. 4 (Summer 1990): 461–82.
- Vitoria, Francisco de. “De Indus.” In *Vitoria: Political Writings*, edited by Anthony Pagden and Jeremy Lawrance, 231–99. New York: Cambridge University Press, 1992.
- Voltaire. *Philosophical Dictionary*. Edited and translated by Theodore Besterman. London: Penguin Press, 1984.
- Walzer, Michael. *Just and Unjust Wars*. New York: Basic Books, 1977.
- Wellman, Carl. *A Theory of Rights*. Totowa, NJ: Rowman and Allanheld Press, 1985.
- Williams, Bernard. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, by J. J. C. Smart and Bernard Williams, 77–150. Cambridge: Cambridge University Press, 1973.

IT'S A FINE LINE BETWEEN SADISM AND HORROR

Scott Woodcock

HORROR FILMS now occupy a prominent position in popular culture. They are available on most streaming services, promoted enthusiastically in mainstream media, and viewed by fans from increasingly varied demographic groups. It is an exciting time for horror, in fact, with new voices creating films from perspectives not previously included within the genre.¹ For some, however, the violence in horror films deserves scrutiny, and the need for such scrutiny is all the more pressing given their increasing status and availability. Horror films may have gained widespread acceptance, but are there not risks involved for those who enjoy simulations of intense violence for entertainment?

In answer to this question, Ian Stoner has provided a thought-provoking defense of the kinds of violent horror films that raise concerns for those who are suspicious of how they might affect our moral capacities.² Specifically, Stoner offers reasons to reject a type of argument that Gianluca Di Muzio and I have each put forward that watching horrific violence as a form of entertainment risks harm to the reactive attitudes required for agents to exhibit a well-functioning moral psychology.³ Stoner makes persuasive points in his defense of horror films, and he is correct that most horror films are examples of morally permissible forms of entertainment. Yet my aim will be to argue that he has overstated his case for the permissibility of viewing extreme violence as entertainment. Thus, my position remains that at least *some* instances of creating or viewing horror films ought to be considered morally problematic. The details of particular cases can be complicated; nevertheless, I argue that we ought to

- 1 Jordan Peele (*Get Out*, *Us*, *Nope*), for example, has added a prominent African American voice to the typically white landscape of horror films, but other filmmakers, such as Ana Lily Amirpour (*A Girl Walks Home Alone at Night*), Jeff Barnaby (*Blood Quantum*), and Remi Weekes (*His House*), have also gained critical acclaim presenting horror themes from the perspectives of Iranians; Indigenous Canadians; and African refugees in the United Kingdom.
- 2 Stoner, "Barbarous Spectacle and General Massacre."
- 3 Di Muzio, "The Immorality of Horror Films"; Woodcock, "Horror Films and the Argument from Reactive Attitudes."

remain mindful of risks to our reactive attitudes when we consider these details and ask ourselves why we enjoy viewing horrific imagery.

To begin, it is important to identify the argument at stake in this discussion, since there are a number of objections one might have to films categorized within the horror genre. One might claim, for example, that horror films exhibit misogynistic tropes, or that they contribute to the stigmatization of persons with disabilities.⁴ These objections are important, but they are not the principal concern that will be the focus of this discussion. The principal concern at stake here is the impact that fictional scenes of violence might have on the psychology of those who view this material as a form of entertainment. It is a concern about how deriving enjoyment from fictional depictions of suffering might affect an agent's underlying dispositions for sympathy, compassion, and so on. Importantly, it is not the concern that those who view horror violence will become more likely to engage in this same type of behavior. Instead, the concern is that the moral psychology of agents who take pleasure in violent imagery can be harmed in more subtle ways that are difficult to measure via concrete empirical methods.

The argument I presented to defend a precautionary approach with respect to this concern is the *Argument from Reactive Attitudes* (ARA). Stoner summarizes it as follows:

1. It is *prima facie* morally wrong to view, or to facilitate viewing, those features of a work of art or entertainment that encourage the corruption of reactive attitudes that are necessary for human agents to develop and maintain a well-functioning moral psychology.
2. Taking pleasure in murder, torture, dismemberment and other acts of horror violence can threaten the proper functioning of our sympathetic attitudes.
3. Therefore, it is *prima facie* morally wrong to view, or to facilitate viewing, horror films.⁵

The ARA articulates the hypothesis that agents who enjoy watching fictional acts of horror violence risk damaging their capacities for understanding and

4 Prominent discussions of gender and horror include Mulvey, "Visual Pleasure and Narrative Cinema"; Clover, *Men, Women and Chain Saws*; Freeland, "Feminist Frameworks for Horror Films"; and Harrington, *Women, Monstrosity and Horror Film*. Discussions of disabled persons in horror films include Smith, *Hideous Progeny*; Sutton, "Avenging the Body"; and Hall, "Horrible Heroes."

5 Stoner, "Barbarous Spectacle," 513. This presentation of the ARA is helpful as a starting point; however, it does not accurately reflect the argument that I put forward as the ARA. The premises are quoted from my work, but the conclusion is a *prima facie* claim about all horror films that I do not assert. My conclusion is that the potential for harm to our reactive attitudes is a *useful standard* for evaluating horror films (Woodcock, "Horror Films," 311, 320, 323).

sympathizing with suffering in the real world. This may strike some as alarmist, but the concern driving the ARA is worth considering if the violence at stake is presented in ways that invite or encourage sadistic responses from viewers. Consider, for example, a hypothetical film that contains intense, realistic scenes of torture and extreme violence. If this film is created in ways that encourage viewers to enjoy the cruelty depicted or to set aside sympathy for its victims, then it is reasonable to consider the possibility that the film is inviting harm to its viewers by affecting their reactive attitudes in ways that persist after the film is over. Given this possibility, the ARA claims that we have moral reasons to avoid contributing to this type of harm. It suggests that we ought to take a precautionary approach to creating, promoting, or viewing films like this hypothetical example if other, less risky options are available for our enjoyment. Stoner is not persuaded that this precautionary approach should be applied to horror films. He argues that *whatever* horror film is proposed for analysis, the ARA offers no compelling reason to believe that watching violence for enjoyment carries any potential to negatively affect our moral capacities. I am not as optimistic about this conclusion.

Yet before we dig into the details of this disagreement, it is worth emphasizing that Stoner and I agree on some important points related to the ARA and horror films. First, Stoner acknowledges that we should consider the implications of choices that might compromise our moral capacities. For example, he argues that one ought to avoid working in a slaughterhouse if one has reasonable evidence that this type of work results in damage to one's reactive attitudes.⁶ To add recreational examples, I would argue that one ought to renounce a kung fu academy or a hockey league if it encouraged aggressive "us versus them" thinking and seemed to be undermining one's capacity for fairness. Moreover, I would argue that this is an intuitive precautionary outlook that most of us would adopt even if we lacked concrete empirical data *proving* that our kung fu academy or hockey league was having a negative effect on our character. In these kinds of practical situations, decisive empirical evidence is not required to adopt the precautionary view that immersing oneself in toxic norms is not worth the risk of these norms spilling over into other aspects of one's life.⁷

Second, Stoner is correct that the ARA should not apply uniquely to horror films. Other genres, such as black comedies and action movies, can similarly

6 Stoner, "Barbarous Spectacle," 513.

7 It will certainly be difficult in these kinds of practical situations to specify what counts as reasonable grounds to believe that our reactive attitudes are at risk. For now, I set aside this question and note only that clear empirical evidence is not necessary when recreational choices are at stake. Clear empirical evidence *is* important for making decisions about censorship or consumer boycotts, but a much lower standard of epistemic warrant is required to take a precautionary approach to entertainment choices among easily available alternatives.

have negative effects on our reactive attitudes if they invite harmful responses from viewers. This is important to emphasize, since Stoner presents the ARA as a comparative argument that singles out horror films as *prima facie* wrong compared to other types of films. My presentation of the ARA explicitly denies this claim.⁸ The ARA is best understood as a wide-ranging argument about the potential damage to our moral faculties that can be applied to films of any particular genre.⁹ Applying the ARA to horror films is especially interesting because the violent imagery they present can reach such disturbing extremes, but films in this genre are not the only ones with the potential to undermine our moral capacities.

Finally, Stoner is correct that many, if not most, horror films are morally permissible according to the ARA. Films involving monsters, for example, invite their viewers to have reasonable reactions to the horrific imagery presented—for example, disgust, curiosity, suspense, or anticipation. In fact, many of these films encourage viewers to engage with their content in ways that promote compelling meditations on trust (*The Thing*), faith (*The Exorcist*), consumerism (*Dawn of the Dead*), adolescent love (*Let the Right One In*), grief (*The Babadook*), identity (*Annihilation*), and corporate greed (*Alien*). Others examine the bonds of friendship and family in ways that promote curiously wholesome social norms (*The Conjuring; It; A Quiet Place*). Even so-called slasher films are normally designed to create identification with the victim protagonists so that viewers are not vicariously invited to endorse the violence presented in the films (*Halloween; You're Next; Hush*). The victim-oriented appeal of such films is puzzling, as so much work examining the aesthetics of horror has revealed.¹⁰ Yet it is clear that the ARA should not be interpreted such that all horror films are condemned as having disruptive effects on our moral faculties. Rather, the ARA is best understood as an argument that targets a particular subset of horror films.

With this much decided, where is the controversy? If a sensible version of the ARA suggests that some horror films are permissible, if not praiseworthy, while others are problematic if they invite harm to our reactive attitudes, it is difficult to see what is left for debate. Yet disagreement remains because Stoner

8 Woodcock, "Horror Films," 316. Stoner acknowledges this point in a footnote ("Barbarous Spectacle," 524).

9 For example, military-themed films arguably contribute to xenophobic tendencies in human nature that blunt our sympathy for those we perceive as outsiders—through, e.g., racist characterizations of those depicted as enemies (*The Deer Hunter*), failures to depict enemies as full persons (*Black Hawk Down*), or complacency about the permissibility of torture (*Zero Dark Thirty*).

10 Noteworthy examples include Carroll, *The Philosophy of Horror*; Feagin, "Monsters, Disgust and Fascination"; Gaut, "The Paradox of Horror"; Morreall, "Fear without Belief"; Smuts, "The Paradox of Painful Art"; Bantinaki, "The Paradox of Horror"; and Strohl, "Art and Painful Emotion."

denies that *any* horror films are problematic when it comes to ARA-derived concerns. As he sets up his discussion of the ARA, he boldly asserts that: “Whatever gore films you take to pose the challenge most starkly . . . , those are the films whose graphic depictions of violence I will defend.”¹¹ Why does Stoner argue for such an extreme position? His view is that whatever horrific violence is depicted in a horror film, its aesthetic appeal is premised on viewers engaging with horrific material in ways that are morally appropriate—for example, ways that reflect our desires to experience fear, suspense, disgust, and sympathy. He therefore denies that viewers of horrific violence are drawn to this content to vicariously live out sadistic impulses or enjoy the suffering of victims from a detached perspective that could reinforce negative impulses in our reactive attitudes. Even in the face of extreme hypothetical cases such as Di Muzio’s *Nazi Cruelty Film*, Stoner argues that creating or viewing such films would be morally permissible if not for the fact that most of us cannot get enough emotional distance from the Holocaust to enjoy fictional depictions of those atrocities presented as entertainment.¹² Thus, a film presenting equivalent levels of horrific suffering devoid of any meaningful plot, character, or subtext would be acceptable for Stoner if it lacked historical associations, because the appeal for viewers could only be based on paradoxical, victim-oriented interests. In short, Stoner presumes that any realistic prospective horror film will be enjoyed by viewers for innocent reasons; therefore, he believes that “it isn’t clear that a movie that invited vicarious sadism—instead of paradoxical enjoyment of fear and disgust—could ever be a candidate for a horror film.”¹³

I would like to believe that Stoner is right about the appeal of any and all horror films. However, I am not nearly as optimistic about the complexity of our motivations for viewing horrific suffering, and I do not think we ought to be complacent about the various examples of films, or certain features of films,

11 Stoner, “Barbarous Spectacle,” 512. Note that Stoner stipulates that the depictions of violence must be fictional and that the objection at stake must not be extraneous to horrific violence—e.g., based on sexism or ableism.

12 Di Muzio, “The Immorality of Horror Films,” 281–84. Stoner addresses this example in “Barbarous Spectacle,” 520–22.

13 Stoner, “Barbarous Spectacle,” 519. Stoner is not alone here. In *Horror Film and Affect*, Xavier Aldana Reyes claims that horror films invite a “masochistic contract” with viewers that necessarily aligns identification with their victims (164–66). By contrast, S. Evan Kreider carefully leaves open the possibility that the ARA might gain traction for certain films, or given sufficient empirical evidence, despite the reservations he expresses about the way Di Muzio applies the ARA to slasher films (“The Virtue of Horror Films”). Similarly, Marius A. Pascale claims that our macabre fascination with horrific content can lead either to harmful or beneficial results depending on the details of individual cases (“Morality and Morbidity”).

that seem to invite sadism rather than sympathy. It is at this point that one might expect an analysis of the narrative details of some of the most conspicuous examples of shocking horror films, such as *Wolf Creek*, *The Loved Ones*, or *The Devil's Rejects*. Indeed, I think certain parts of these films invite viewers to identify with those inflicting suffering rather than their victims, so it is tempting to dive into the aesthetic details and argue case by case that they cannot plausibly be interpreted in the optimistic way Stoner requires for viewers to enjoy their content for morally innocent reasons. However, instead of working through a list of examples, I will proceed with a *reductio* strategy to highlight the extreme nature of Stoner's position when it comes to (a) the creation of horror films and (b) what makes these films appealing to their intended audiences.

Consider first what it would mean for the writers and directors of horror films if Stoner were correct that viewers are never drawn to horrific imagery for anything but innocent reasons. It would imply that the creators of this content do not ever need to ask themselves whether their work invites identification with the agents who commit the violence depicted nor whether the narrative or aesthetic features of their work invite morally compromised responses. They may need to worry about historical associations or other factors that might decrease how favorably their work is received if viewers respond atypically, but they can rest assured that ordinary viewers will not engage with their work in morally compromised ways *no matter how they script and film horror violence*. Stoner's view would imply, therefore, that interpretive discussions of the morality of scripting and filming violence are ill-conceived and that directors such as Karyn Kusama, Jennifer Kent, and Sophia Takal are mistaken if they take the time to present horrific material in specific ways they consider to be socially responsible compared to other alternatives that would generate just as much revenue and acclaim.¹⁴

This implication strikes me as a deeply implausible view of what is at stake when artists make the effort to consider the subtleties of how they depict violence in their work. The choices they make are certainly not straightforward. As the title of this paper suggests, it can be very difficult to know which narrative choices, camera perspectives, or editing decisions will create appropriately horrific responses in viewers without inviting them to vicariously enjoy the content out of sadistic inclinations. Moreover, reasonable people can disagree about many of these choices and their implications. Yet my title is also meant as a reference to the film *This Is Spinal Tap* when David St. Hubbins acknowledges the difference between an album cover with a rock star protagonist tied down by powerful women and one that presents a naked woman on all fours held by

14 See Dobbs, "Karyn Kusama Interview"; Juzwiak, "Jennifer Kent's Brutal New Film *The Nightingale* Spits on Your Colonialism"; and Morse, "Black Christmas Remake Director Sophia Takal Talks What Sets Her Version Apart."

a leash and dog collar with a glove pushed in her face. The humor in this scene is the deadpan earnestness with which Hubbins recognizes what is obvious: details in the way art is presented can significantly affect the way audiences are invited to engage with its content, and there are at least *some* clear cases where art invites immoral responses. To presume otherwise is naïve, and the details of at least some horror films, such as *The Devil's Rejects*, would, I think, require Hubbinsesque naïvety to defend as innocuous.

To emphasize this point, it is worth returning to Di Muzio's *Nazi Cruelty Film*. Even if we grant Stoner the claim that we might not judge a depiction of Holocaust atrocities as problematic if we could get enough emotional distance from the history at stake, it is still vital to ask how the film is scripted and edited in terms of inviting viewers to identify with the victims or the perpetrators of the atrocities that it presents. It may be correct to reject Di Muzio's claim that a film is necessarily compromised if it presents horror violence for no other purpose than to provide viewers with heightened emotional intensity, yet it is critical to specify whether the hypothetical film under consideration is victim oriented. In other words, it is one thing to bite the bullet and accept that a grim depiction of atrocities could be permissible if it invites viewers to engage in the *paradoxical* pleasures that horror can provide, but it is entirely different to claim that *any* formulation of a Nazi cruelty film with sufficient distance from historical associations is innocuous. If this is not already clear, consider Di Muzio's other hypothetical example: a child torture film.¹⁵ Viewers need not have historical associations with this content to judge that a film is problematic if it invites viewer identification with the torturer rather than the child. Of course, a sadistic version of the child torture film is not a fair analog for most horror films, because even those that display harm to children are normally victim oriented (*Jaws*; *Dr. Sleep*; *The Innocents*).¹⁶ However, it is important to remember the burden of proof in play after Stoner denies that any horror film risks compromising our reactive attitudes by inviting sadistic pleasures. If we accept the variety of explanations of the appeal of horror violence that have been presented in the literature, why think that no horror film invites viewers to align with the perpetrators of violence toward innocent victims? For example, if Aaron Smuts is correct that a "rich experience theory" explains some of the appeal of horrific violence to viewers, is it not *possible* that some horror films provide visceral experiences that are not victim oriented?¹⁷ If this is a realistic

15 Di Muzio, "The Immorality of Horror Films," 281–82.

16 Marius A. Pascale makes this point in "Art Horror, Reactive Attitudes, and Compassionate Slashers."

17 See Smuts, "Art and Negative Affect."

possibility, it places pressure on Stoner's categorical claim that no horror films inviting viewers to share in vicarious sadism exist. Thus, while the appeal of horror is still an open area of inquiry, the burden of proof remains on Stoner to guarantee that no horror films engage viewers in ways that invite sadistic responses and potentially compromise their reactive attitudes.

This leads us to the viewer side of Stoner's claim that every horror film is defensible. For we can anticipate the following reply: it is admittedly *possible* for a horror film to invite sadistic responses from viewers, but in practice, no actual horror film operates this way, because viewers are just not attracted to horror films to satisfy sadistic urges. Even in cases such as *The Devil's Rejects*, this anticipated reply argues, viewers only identify with perpetrators of violence out of a meta-level interest in challenging established normative conventions. Like more overt attempts to force viewers to reflect on the limits of what they find stimulating (e.g., *A Clockwork Orange*; *Funny Games*; *A Serbian Film*), viewers are drawn to apparent endorsements of cruelty from a perspective that will not threaten their reactive attitudes in the way the ARA predicts.¹⁸ Thus, one might claim that the ARA never gains traction in the real world since ordinary viewers are not motivated by sadistic impulses when they enjoy even the grisliest of horror films.

Again, I would genuinely like to believe this claim about the appeal of horrific content. It would be comforting to know that it is only unusual sociopaths who watch horror films to satisfy their sadistic impulses rather than their paradoxical interests in horror, disgust, suspense, and so on. Yet it would be foolish for those of us who enjoy horror films to believe that our interests in watching violence are always pure and noble compared to those of some separate class of sociopaths who share none of our psychological traits. It is too convenient for the complexity of human desires to play out so straightforwardly, and we cannot rely on our introspective experience to reassure us that our reasons for watching violent content are always defensible. If recent work in empirical psychology on implicit bias and cognitive dissonance has taught us anything, it is that we ought to follow Kant's warning to avoid being overconfident about

18 Jeremy Morris provides something like this type of justification for *The Devil's Rejects*. He acknowledges that the film encourages vicarious sadism by inviting viewers to identify with its killers yet claims that this is the point of its metalevel horror: "it transforms the source of fear from a distant other to something familiar in ourselves. The terror of the victim is supplanted by the delight of the torturer, which is being consciously shared by the audience: *that* is the source of horror" ("The Justification of Torture-Horror," 51). I agree with Morris that *The Devil's Rejects* uses humor, music, and narrative structure to encourage viewers to identify with its torturers; I do not agree that the film exhibits metalevel objectives to redeem the sadism it invites.

the true impulses that are responsible for our choices.¹⁹ Thus, I maintain that when certain films appear to invite viewers to take pleasure in horrific violence, we ought to seriously consider the possibility that audiences of these films are at least partly taking pleasure in that violence. We should not avoid this possibility by seeking refuge in ironic distance or metalevel analysis, for there is no excluded middle here. Instead, it might be that viewers are sometimes drawn to horror because of paradoxical aesthetic tastes *and* the sadistic impulses that allow the ARA to gain traction.²⁰

Stoner, it must be noted, attempts to accommodate this concern in his defense of horror films from the ARA. In addition to recognizing that many horror films are complicit in misogyny and ableism, he acknowledges that we ought to adopt a precautionary attitude toward films that invite sadism: “if a particular [horror film] somehow did, in contravention of genre expectations, invite vicarious sadism, I would support a precautionary attitude toward that specific film.”²¹ This is an odd claim from someone who has promised to defend *whatever* horror films one selects to advance the ARA. How can Stoner advance his sweeping defense of every horror film with this prominent concession? The key is a rigid definition of genre conventions such that any film that invites sadism is no longer functioning *qua* horror film. Stoner defends all possible candidates for horror films from the ARA because as soon as any film exhibits traits that allow the ARA to gain traction, the film in question is no longer a genuine example of a horror film after Stoner defines the genre in strictly victim-oriented terms.

I think this way of defending horror films from the ARA is less than satisfying. First, it attempts to settle a longstanding question about the moral status of horror films by terminological stipulation, and in doing so, it commits a

19 Kant, *Groundwork of the Metaphysics of Morals*, 19–20. To scratch the surface of the literature on implicit bias and cognitive dissonance, see Brownstein and Saul, *Implicit Bias and Philosophy*; Levy, “Implicit Bias and Moral Responsibility”; Harmon-Jones, *Cognitive Dissonance*; and Brownstein, Madva, and Gawronski, “Understanding Implicit Bias.”

20 This point about the dual appeal of horror violence is a key part of what makes Clover’s *Men, Women and Chain Saws* so influential in the literature on horror aesthetics. Clover observes victim-oriented viewer identification in slasher films, and she challenges the presumption that these films are unambiguously grounded in misogynistic sadism. Yet she simultaneously acknowledges features of the films that invite identification with killers who prey on young women in varying states of undress. What makes her work so compelling is the way she illuminates the complexity of the shifting and sometimes contradictory perspectives that make horrific violence appealing to its viewers (see esp. *Men, Women and Chain Saws*, 182). Stoner, however, emphasizes only one side of this complexity by referring to Clover’s “final girl” thesis as if it exonerates all slasher films from vicarious sadism (“Barbarous Spectacle,” 519). I think that is a mistake. Clover’s work is fascinating because she so artfully describes how horror films invite identification with both victims and their sadistic killers.

21 Stoner, “Barbarous Spectacle,” 520.

no-true-Scotsman fallacy. Any potential counterexample put forward as morally problematic can be dismissed according to this strategy, because as soon as it seems persuasive that a film invites harms to our reactive attitudes, Stoner can claim that the film is not properly defined as a horror film. It is a foolproof strategy, to be sure, but it does not meaningfully advance discussion of the ARA for those interested in the darker elements of what are normally considered horror films.

Moreover, as a matter of classification, it is not persuasive to draw the boundaries of the horror genre so narrowly that its films cannot possibly include elements that invite vicarious sadism. Returning to *The Devil's Rejects*, it is implausible to think that the film loses its status as a horror film because its main characters (who commit murder, torture, and sexual assault) are presented as protagonists. Similarly, if one asks whether films in the *Saw* franchise present feeble pretexts for the characters to be punished in sadistic ways that viewers can vicariously enjoy, it is not as if one is asking *whether the films are horror films*, no matter how one answers the question. We also want the flexibility to apply the ARA to specific features of films that might otherwise be morally defensible. The original *Texas Chain Saw Massacre*, for example, is a mix of grisly victim-oriented scenes followed by a final dinner scene that surely veers into vicarious sadism. Di Muzio unfairly condemns the film as uniformly immoral, but he is not wrong about the dinner scene when he observes that

although the audience began watching the scene from Sally's point of view, it is now drawn irresistibly to the side of the table where Leatherface and the hitchhiker are sitting. The accumulation of disturbingly entertaining sights and the high level of stress induced by Sally's piercing screams have won the spectator over to the killers' side.²²

If this evaluation is correct, is this final part of the film suddenly no longer properly described as part of the horror genre? It would seem odd for a specific scene from a classic horror film to be exempt from the genre to which the film as a whole is clearly a member. This classification system is not incoherent, but it is misleading to rely on it to defend the otherwise bold thesis that all horror films are immune to ARA-derived concerns.

Finally, the classification system Stoner's argument relies on can only set up a contrast with advocates of the ARA if the argument is interpreted such that victim-oriented genre conventions in horror films are alleged to be morally hazardous. Yet my presentation of the ARA (which Stoner primarily draws on in his summary of the argument) explicitly denies this claim and makes

22 Di Muzio, "The Immorality of Horror Films," 287.

considerable efforts to discuss the complexity of what might trigger the ARA when it comes to the controversial appeal of horrific content for viewers.²³ This makes it odd for Stoner to try to ensure that discussion of the ARA occurs within the confines of a presumption that true horror films invite nothing but paradoxical, victim-oriented responses. Stoner can legitimately set up a contrast with Di Muzio's position that all slasher films are problematic, but if an advocate of the ARA allows, as I do, for an open-ended framework to consider whether films invite sadistic responses, then it is not helpful to stipulate that any case where reactive attitudes are at risk is no longer a genuine horror film.

In the end, what we find is that neither side of the ARA debate is truly seeking to defend a cut-and-dried thesis about the moral status of every film or part of a film that would commonly be classified as belonging to the horror genre. My defense of the ARA does not argue for a precautionary claim that all horror films are morally problematic, and Stoner ultimately allows for concerns about our moral capacities by excluding invitations to vicarious sadism from his definition of what counts as a horror film. I hope it is clear, however, that Stoner's artful classification strategy does not help advance a sensible, case-by-case approach to horror violence. Some films are problematic for reasons described by the ARA; others are not. This middle ground is not usefully captured by a thesis that any horror film is morally defensible ... unless it is not defensible, in which case it is no longer a horror film. Instead, we should apply the ARA to whatever films invite sadistic responses, and some of these films will be horror films. The difference is important, because we ought to remain open to the possibility that horror films draw some of their appeal from the darker aspects of human nature—for example, a sadistic fascination with suffering that potentially disrupts our reactive attitudes if it is promoted as a source of enjoyment. That may not be our predominant interest in horrific content, but neither is it a possibility that can be dismissed via genre definitions. In conclusion, then, I hope that consideration of the ARA should encourage us to balance the paradoxical pleasures of horror films with a willingness to confront the possibility that some of the films are morally problematic if they invite us to compromise our moral psychology. This may be a frightening possibility for us to contemplate, but fans of horror films are presumably up to the task.²⁴

University of Victoria
woodcock@uvic.ca

23 Woodcock, "Horror Films," 316–20.

24 Thanks to Jason Anderson, David Boutillier, Sam Cowling, Colin Macleod, and two anonymous referees for their helpful comments in the writing of this paper.

FILMOGRAPHY

- Alien*. Directed by Ridley Scott. Written by Dan O'Bannon and Ronald Shusett. 1979.
- Annihilation*. Directed and written by Alex Garland. 2018.
- A Serbian Film*. Directed by Srdjan Spasojevic. Written by Aleksandar Radivojevic and Srdjan Spasojevic. 2010.
- The Babadook*. Directed and written by Jennifer Kent. 2014.
- Black Hawk Down*. Directed by Ridley Scott. Written by Mark Bowden and Ken Nolan. 2001.
- Blood Quantum*. Directed and written by Jeff Barnaby. 2019.
- A Clockwork Orange*. Directed by Stanley Kubrick. Written by Stanley Kubrick and Anthony Burgess. 1971.
- The Conjuring*. Directed by James Wan. Written by Chad Hayes and Carey W. Hayes. 2013.
- Dawn of the Dead*. Directed and written by George A. Romero. 1978.
- The Deer Hunter*. Directed by Michael Cimino. Written by Michael Cimino, Deric Washburn, and Louis Garfinkle. 1978.
- The Devil's Rejects*. Directed and written by Rob Zombie. 2005.
- Dr. Sleep*. Directed by Mike Flanagan. Written by Mike Flanagan and Stephen King. 2019.
- The Exorcist*. Directed by William Friedkin. Written by William Peter Blatty. 1973.
- Funny Games*. Directed and written by Michael Haneke. 1997.
- Get Out*. Directed and written by Jordan Peele. 2017.
- A Girl Walks Home Alone at Night*. Directed and written by Ana Lily Amirpour. 2014.
- Halloween*. Directed by John Carpenter. Written by John Carpenter and Debra Hill. 1978.
- His House*. Directed by Remi Weekes. Written by Remi Weekes, Felicity Evans, and Toby Venables. 2020.
- Hush*. Directed by Mike Flanagan. Written by Mike Flanagan and Kate Siegel. 2016.
- The Innocents*. Directed and written by Eskil Vogt. 2021.
- It*. Directed by Andy Muschietti. Written by Chase Palmer, Cary Joji Fukunaga, and Gary Dauberman. 2017.
- Jaws*. Directed by Steven Spielberg. Written by Peter Benchley and Carl Gottlieb. 1975.
- Let the Right One In*. Directed by Tomas Alfredson. Written by John Ajvide Lindqvist. 2008.

- The Loved Ones*. Directed and written by Sean Byrne. 2009.
- A Quiet Place*. Directed by John Krasinski. Written by Bryan Woods, Scott Beck, and John Krasinski. 2018.
- Saw*. Directed by James Wan. Written by Leigh Whannell and James Wan. 2004.
- The Texas Chain Saw Massacre*. Directed by Tobe Hooper. Written by Kim Kenkel and Tobe Hooper. 1974.
- The Thing*. Directed by John Carpenter. Written by Bill Lancaster and John W. Campbell Jr. 1982.
- This Is Spinal Tap*. Directed by Rob Reiner. Written by Christopher Guest, Michael McKean, and Harry Shearer. 1984.
- Wolf Creek*. Directed and written by Greg McLean. 2005.
- You're Next*. Directed by Adam Wingard. Written by Simon Barrett. 2011.
- Zero Dark Thirty*. Directed by Kathryn Bigelow. Written by Mark Boal. 2012.

REFERENCES

- Aldana Reyes, Xavier. *Horror Film and Affect: Towards a Corporeal Model of Viewership*. New York: Routledge, 2016.
- Bantinaki, Katerina. "The Paradox of Horror: Fear as a Positive Emotion." *The Journal of Aesthetics and Art Criticism* 70, no. 4 (November 2012): 383–92.
- Brownstein Michael, Alex Madva, and Bertram Gawronski. "Understanding Implicit Bias: Putting the Criticism into Perspective." *Pacific Philosophical Quarterly* 101, no. 2 (June 2020): 276–307.
- Brownstein, Michael and Jennifer Saul, eds. *Implicit Bias and Philosophy*. 2 vols. Oxford University Press, 2016.
- Carroll, Noël. *The Philosophy of Horror; or, Paradoxes of the Heart*. New York: Routledge, 1990.
- Clover, Carol. *Men, Women and Chain Saws: Gender in the Modern Horror Film*. Princeton, NJ: Princeton University Press, 1992.
- Di Muzio, Gianluca. "The Immorality of Horror Films." *International Journal of Applied Philosophy* 20, no. 2 (Fall 2006): 277–94.
- Dobbs, Sarah. "Karyn Kusama Interview: xx, Women in Horror, Scary Movies in Times of Adversity." *Den of Geek*, May 7, 2017. <https://www.denofgeek.com/movies/karyn-kusama-interview-xx-women-in-horror-scary-movies-in-times-of-adversity>.
- Feagin, Susan L. "Monsters, Disgust and Fascination." *Philosophical Studies* 65, nos. 1–2 (February 1992): 75–84.
- Freeland, Cynthia A. "Feminist Frameworks for Horror Films." In *Post-Theory: Reconstructing Film Studies*, edited by David Bordwell and Noël Carroll, 195–218. Madison, WI: University of Wisconsin Press, 1996.

- Gaut, Berys. "The Paradox of Horror." *The British Journal of Aesthetics* 33, no. 4 (October 1993): 333–45.
- Hall, Melinda. "Horrible Heroes: Liberating Alternative Visions of Disability in Horror." *Disability Studies Quarterly* 36, no. 1 (Winter 2016). <https://dsq-sds.org/index.php/dsq/article/view/3258/4205>.
- Harmon-Jones, Eddie, ed. *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology*. 2nd ed. Washington, DC: American Psychological Association, 2019.
- Harrington, Erin. *Women, Monstrosity and Horror Film: Gynaehorror*. London: Routledge, 2018.
- Juzwiak, Rich. "Jennifer Kent's Brutal New Film *The Nightingale* Spits on Your Colonialism." *Jezebel*, February 8, 2019. <https://jezebel.com/jennifer-kents-brutal-new-film-the-nightingale-spits-on-1836899139>.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1998.
- Kreider, S. Evan. "The Virtue of Horror Films: A Response to Di Muzio." *International Journal of Applied Philosophy* 22, no. 1 (Spring 2008): 149–57.
- Levy, Neil. "Implicit Bias and Moral Responsibility: Probing the Data." *Philosophy and Phenomenological Research* 94, no. 1 (January 2017): 3–26.
- Morreall, John. "Fear without Belief." *The Journal of Philosophy* 90, no. 7 (July 1993): 359–66.
- Morris, Jeremy. "The Justification of Torture-Horror: Retribution and Sadism in *Saw*, *Hostel*, and *The Devil's Rejects*." In *The Philosophy of Horror*, edited by Thomas Fahy, 42–56. Lexington, KY: The University Press of Kentucky, 2010.
- Morse, Chris. "Black Christmas Remake Director Sophia Takal Talks What Sets Her Version Apart." *Dead Entertainment*, November 27, 2019. <https://deadentertainment.com/2019/11/27/black-christmas-remake-director-sophia-takal-talks-what-sets-her-version-apart>.
- Mulvey, Laura. "Visual Pleasure and Narrative Cinema." *Screen* 16, no. 3 (Autumn 1975): 6–18.
- Pascale, Marius A. "Art Horror, Reactive Attitudes, and Compassionate Slashers: A Response to Di Muzio's 'The Immorality of Horror Films.'" *International Journal of Applied Philosophy* 33, no. 1 (2019): 141–59.
- . "Morality and Morbidity: Semantics and the Moral Status of Macabre Fascination." *The Journal of Value Inquiry* 53, no. 4 (December 2019): 551–77.
- Smith, Angela M. *Hideous Progeny: Disability, Eugenics, and Classic Horror Cinema*. New York: Columbia University Press, 2011.
- Smuts, Aaron. "Art and Negative Affect." *Philosophy Compass* 4, no. 1 (January 2009): 39–55.

- . “The Paradox of Painful Art.” *Journal of Aesthetic Education* 41, no. 3 (Fall 2007): 59–76.
- Stoner, Ian. “Barbarous Spectacle and General Massacre: A Defence of Gory Fictions.” *Journal of Applied Philosophy* 37, no. 4 (August 2020): 511–27.
- Strohl, Matthew. “Art and Painful Emotion.” *Philosophy Compass* 14, no. 1 (January 2019): e12558.
- Sutton, Travis. “Avenging the Body.” In *A Companion to the Horror Film*, edited by Harry M. Benshoff, 73–89. Malden, MA: Wiley-Blackwell, 2014.
- Woodcock, Scott. “Horror Films and the Argument from Reactive Attitudes.” *Ethical Theory and Moral Practice* 16, no. 2 (April 2013): 309–24.