JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXX · NUMBER 5
September 2025

ARTICLES

SYMPOSIUM

OPEN ACCESS FREE FOR AUTHORS AND READERS ISSN 1559-3061

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXX · NUMBER 5
September 2025

ARTICLES

SYMPOSIUM

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org
ISSN 1559-3061

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

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

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

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

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · SEPTEMBER 2025
https://doi.org/10.26556/jesp.v30i5.3371

AGAINST THE RIGHT TO WORK, FOR THE RIGHT TO CONTRIBUTE

Jordan Desmond

Is there a universal right to access some form of monetary income? And if so, how should the relevant entitlement be construed? In this article, I set aside the first question in the hopes of making progress on the second. For assuming a universal right to access an income exists, it is not at all obvious how such a right ought to be responded to. More specifically, there is disagreement as to whether we ought to ensure access to income by way of a universal basic income (UBI) or a right to work in the form of remunerated employment.[1]

One reason to prefer the former option is that it presents perhaps the simplest and most direct way of ensuring universal access to income.[2] Moreover, if all persons have access to the material resources necessary to meet their basic needs, then, though they may remain vulnerable to the contingent forces of the labor market, such vulnerability no longer seems problematic in the way it otherwise might. On the other hand, ensuring universal access to income by guaranteeing opportunities for remunerated employment may seem to be an equally effective way of responding to our pecuniary interests.[3] To be sure, such a right would need to be supplemented by a basic income for those unable to exchange their labor, but among those who are capable of working, there is

1 For a particularly insightful exchange on this issue, see Thomas, "Full Employment, Unconditional Basic Income and the Keynesian Critique of Rentier Capitalism"; and Merrill and Neves, "Unconditional Basic Income and State as an Employer of Last Resort." For two attempts to demonstrate how the right to work and a policy of unconditional basic income might be complimentary, see Harvey, "The Right to Work and Basic Income Guarantees"; and Standing, "Why Basic Income Is Needed for a Right to Work."

2 For some recent defenses of establishing a universal basic income, see Den Otter, "A Constitutional Right to a Universal Basic Income"; Hemel, "Basic Income as a Human Right?"; McKinnon, "Basic Income, Self-Respect and Reciprocity"; and Van Parijs and Vanderborght, *Basic Income*.

3 For two recent and compelling cases in favor of establishing a job guarantee, see Paul et al., "A Path to Ending Poverty by Way of Ending Unemployment"; and Tcherneva, *The Case for a Job Guarantee*. Note that my concern here is distinct from related discussions about the right to meaningful work.

no obvious reason to think it would be any more or less effective than a UBI at guaranteeing access to income.

But those who defend ensuring access to income by way of a right to work rather than through a UBI often appeal to an additional benefit of pursuing the former strategy. Whereas the benefits of a UBI seem strictly pecuniary, it is often thought that work is a source of both pecuniary *and* nonpecuniary benefits.[4] And, it might be argued, these additional benefits give us decisive reason to establish not a UBI but a right to work.

The first aim of this article, then, is to evaluate the claim that the nonpecuniary benefits of work generate a decisive case in favor of establishing a right to work rather than a UBI. And my ultimate contention is that they do not. To see why, consider the following proposal. The interest theory of rights argues that rights function to protect and promote interests.[5] That is, we are justified in ascribing a right to some good or performance, on this view, just in case the subject of the right has interests of sufficient weight so as to justify imposing a duty on another party to provide the good or performance in question. And given this understanding of rights, it stands to reason that rights—and their corresponding duties—ought to be both *effective* and *efficient* in protecting and promoting the interests that purportedly ground them. They should be effective in the sense that recognizing the right would, in actuality, serve to promote or protect the interests upon which it is grounded. And they should be efficient in the sense that there should be no alternative right that would deliver a similar package of benefits to the right holder at a lesser cost or burden to those who are duty bound by it. For any given right, then, we ought to be able to demonstrate that the right passes tests of efficacy and efficiency vis-à-vis the interests purported to ground it.

As we noted above, both a UBI and a right to work seem to be effective ways of responding to our pecuniary interests. And let us assume for the purposes of this article that they are equally efficient—that they deliver equal benefits to our pecuniary interests at equal cost or burden. The question with which we are concerned, then, is whether the additional nonpecuniary benefits of the right to work generate a decisive case in its favor.

But if this is the question with which we are concerned, then the relevant comparison cannot be between establishing a right to work versus providing a UBI. We might, in providing a UBI, make possible *other* ways of enjoying nonpecuniary benefits, but it is unlikely that the provision of a UBI would in and of itself be a direct source of such benefits. Instead, we ought to ask whether

4 Schaff, "Work, Technology, and Inequality," 107.

5 Raz, *The Morality of Freedom*, ch. 7.

there might be an altogether separate right that both is capable of being packaged alongside a UBI and delivers nonpecuniary benefits more effectively or efficiently than would a right to work. For if it is the case that such a right exists, then the nonpecuniary benefits of work no longer generate a decisive case in favor of a right to work. That is, we would have decisive reason to establish not a right to work but rather some combination of a UBI and a right to this alternative way of delivering nonpecuniary benefits.

In order to carry out the first aim of the article, then, I evaluate the efficacy and efficiency of the right to work vis-à-vis those nonpecuniary interests to which its defenders appeal in asserting is supremacy over providing a UBI. In considering the efficacy of the right to work, I show that the act of *guaranteeing* employment would be a decidedly *in*effective way of responding to the relevant set of nonpecuniary interests. And in evaluating the efficiency of the right to work, I show that there is an alternative right—what I call the *right to contribute*—that would deliver similar (if not greater) benefits to the relevant set of nonpecuniary interests without taking on the set of costs and burdens attributable to the right to work.

If I am correct that the right to work fails to pass the tests of efficacy and efficiency in the case of its nonpecuniary benefits, then these benefits can no longer ground a decisive case in its favor. Rather, since we have assumed the equal efficacy and efficiency of the right to work and a UBI in the case of our pecuniary interests, the fact that a right to work would be less effective and more costly than a right to contribute in the case of the relevant nonpecuniary interests suggests that we have decisive reason to establish not a right to work but a UBI and a right to contribute.

Notice, then, that the article does not argue in any *direct* way for establishing a UBI. The upshot is rather that the debate between those who favor a right to work and those who favor a UBI must take on a new shape. The nonpecuniary benefits of work can no longer serve as grounds to license a preference for establishing a right to work. And this means that if there is a decisive case to be made in favor of one or the other option, it must be made from the perspective of the efficacy and efficiency with which each option responds to our pecuniary interests.

Such an analysis is, however, beyond the scope of this article. Instead, having made my case against establishing a right to work—or rather one reason for establishing a right to work—the remainder of the article takes up the secondary aim of justifying a right to contribute. In doing so, I begin by considering an existing attempt to justify a right to contribute that I attribute to Kimberley Brownlee.[6] Though compelling, I ultimately conclude that Brownlee's

6 See, e.g., Brownlee, "The Lonely Heart Breaks" and *Being Sure of Each Other*.

argument remains incomplete insofar as it fails to demonstrate the efficacy and efficiency of the right. I then present such a demonstration by identifying a necessary connection between social contribution and our nonpecuniary interests in self-esteem and self-realization.

The article proceeds as follows. I begin in section 1 by clarifying some of the key concepts with which we are concerned. In section 2, I motivate the requirements of efficacy and efficiency against which I assess the right to work and the right to contribute. In section 3, I list the nonpecuniary interests to which defenders of the right to work typically appeal as grounds. In section 4, I demonstrate that the right to work fails to promote these interests effectively and efficiently. In section 5, I unpack Brownlee's defense of the right to contribute and conclude that it remains incomplete. Finally, in section 6, I demonstrate the efficacy and efficiency of the right to contribute in promoting our interests in self-esteem and self-realization.

1. CLARIFICATORY REMARKS

1.1. Rights

There are three important details about the way I employ the term 'rights' in this article. First, when I speak of rights, I speak of *moral* rights, as opposed to *legal* rights. In other words, the present concern is whether a case can be made for the existence of a moral right to work. If it is indeed the case that such a moral right exists, then we may have legitimate grounds to insist upon the establishment of a codified legal right. Second, I use the term 'right' in the Hohfieldian sense of a claim right and not in the sense of a mere liberty right.[7] Thus, to say that an individual possesses a right entails that such an individual has a legitimate claim against another party (i.e., the state) to some good or performance that the other party is duty bound to provide (i.e., remunerated employment). Third and most importantly, I adopt here what is known as the *interest theory of rights*. On this view, rights function to promote and protect important interests

7 I take it that the construction of the right to work found in, for instance, the United Nations Universal Declaration of Human Rights and International Covenant on Economic, Social, and Cultural Rights (as well as the subsequent General Comments Nos. 18 and 23) is that of the *liberty* to work. Similarly, one finds an early articulation of the liberty to work in the revised Declaration of the Rights of Man and of Citizens drafted by the French National Assembly in 1793. The historical roots of the (claim-)*right* to work date back at least to the third French Revolution (the February Revolution of 1848), during which widespread support for a *droit au travail* resulted in the establishment of (ultimately temporary) national workshops designed to provide relief for the involuntarily unemployed. For a comprehensive historical survey of the right to work, see Spengler, "Right to Work."

of the right holder.[8] As such, the interest theory of rights holds that whether or not an individual has a right to some good or performance depends crucially on the interests that would be served by the recognition or exercise of such a right and whether they are sufficiently weighty as to justify regarding another party as duty bound to provide the good or performance in question.

1.2. Work and Social Contribution

As the backdrop from which this article proceeds is the increasing support one finds for policies guaranteeing employment, my use of the term 'work' is intended to be interchangeable with the term 'remunerated employment'.[9] Nevertheless, allow me to call attention to three noteworthy features of my account of work as remunerated employment. First, I take it that work involves some form of *labor*, broadly defined. Whether it be physical, cognitive, or emotional in nature, when I speak of work, I have in mind the intentional or willing performance of some particular activity.[10] Second, work is assumed here to be *productive*. That is, it is assumed that the output of those activities that constitute work contributes in some way to the fulfillment of interests. Finally and crucially, work is productive labor that intends to be *remunerated*. It is not, in other words, undertaken for frivolous purposes. Indeed, when we speak of a right to work in the context of guaranteed employment, typically what we have in mind is the right to exchange one's productive labor for an income (either in the form of pecuniary benefits or some other good of equal value).

Notice, though, that defining work as remunerated seems to narrow the sense in which work must be productive. If I spend the day gathering coconuts after having been marooned on a desert island, there is no doubt a sense in which I have engaged in a form of productive labor—I have performed an activity that serves the interests of at least one person (myself). But surely one cannot expect

8 Raz, *The Morality of Freedom*, ch. 7.

9 Note, then, that my aim here is not to define the concept of work *as such* but merely as it is employed in discussion of the right to remunerated employment. In this way, I do not mean to imply that nonremunerated forms of, say, domestic labor are not work, properly speaking, or are in any way less valuable than labor that is remunerated. Rather, my intention is merely to employ the term 'work' as a shorthand for remunerated employment, if only to remain consistent with existing literature on the right to remunerated employment.

10 There are two points worth clarifying here. First, the activity in question need not itself be active, strictly speaking. Someone who is paid to sleep by researchers conducting a study on sleep apnea is still engaging in a kind of work-constituting activity. Second, the work with which this article is concerned is done intentionally and willingly. I take it that those who defend the right to work do not have in mind anything like indentured servitude that happens to be remunerated. Thanks to an anonymous reviewer for urging me to clarify these two points.

to be remunerated for labor that serves only one's self-interest. Rather, work is remunerated precisely because it serves the interests of—and so is valuable to—at least one person other than the worker. Call those productive activities that serve the interests of others forms of *socially productive* labor.[11] Thus, when I speak of work, what I have in mind is *labor that is socially productive and remunerated*.[12] In contrast, something qualifies as a form of social contribution when it displays the first two features listed above. In other words, social contribution is labor that is socially productive. As such, all work involves some form of social contribution, but not all social contribution qualifies as work. To qualify as work, one's social contribution must additionally be the object of remuneration.

2. TWO REQUIREMENTS FOR JUSTIFYING RIGHTS

2.1. The Efficacy Requirement

The efficacy requirement, as I call it, is intended to capture the idea that rights ought to be effective at promoting the interests purported to ground them. To better understand what I mean by this, recall that the interest theory of rights holds that "a right exists where the interests of the right-holders are sufficient to hold another to be obligated" and that the relevant obligation is "to behave in a way which protects or promotes" the interests in question.[13] In light of this relationship between rights and duties (or obligations), Raz makes the following observation:

> The fact that rights are sufficient to ground duties limits the rights one has. Only where one's interest is a reason for another to behave in a way which protects or promotes it, and only when this reason has the peremptory character of a duty, and, finally, *only when the duty is for conduct which makes a significant difference for the promotion or protection of that interest does the interest give rise to a right*.[14]

The efficacy requirement is thus a different way of stating Raz's claim that we are justified in recognizing a right only when the performance of the duties it

11 I thank an anonymous reviewer for raising the desert island case and encouraging me to clarify the conditions under which productive labor can be deemed a form of social contribution.

12 The claim that work consists of these three elements—activity, productivity, and remuneration—aligns with Nickel's understanding of work in "Is There a Human Right to Employment?" 154.

13 Raz, *The Morality of Freedom*, 182, 183.

14 Raz, *The Morality of Freedom*, 183 (emphasis added).

gives rise to can be expected to make a "significant difference" to the promotion or protection of the interests that ground the right. In other words, rights and the duties they impose on others must be *effective* at promoting the interests to which they appeal as grounds.

Why exactly do we need to make this specification? The most obvious answer is that it is the very efficacy of a right that justifies imposing potentially costly duties on others. If a right cannot be expected to be effective, then these costs become an unjustified imposition. But more importantly, as will soon become clear, the recognition of a right has the power to alter the relationship between a particular form of conduct and a set of relevant interests. Indeed, there are cases in which the very recognition of a right undermines the capacity of the conduct it demands from others to promote the interests upon which the right is ultimately grounded. For instance, many of us have interests in experiencing romantic love from others. Normally, this interest is served when others express their romantic love for us. But crucially, it is not clear that this would remain true if we regarded others as duty bound to perform such expressions. For we presumably think that in order to serve the relevant interest, expressions of romantic love must be "a freely bestowed gift, a spontaneous expression of the lover's own deepest desires, rather than something one is obligated to deliver."[15] And given this, we can expect that expressions of romantic love performed as a matter of duty would fail to make a "significant difference" to the promotion of the relevant interest. Put otherwise, recognizing a right to such expressions would prove to be an ineffective means by which to promote the grounding interest because it would undermine their capacity to do so.

2.2. The Efficiency Requirement

Perhaps less obvious are what it means for a right to be *efficient* and why we should expect them to be so. The impetus for such a requirement is that rights, despite delivering benefits to their holders, are nevertheless costly or burdensome for others. They are costly, perhaps most obviously, in the sense that they constrain the freedom of others by limiting how they are permitted to act and, in many cases, by requiring them to act in particular ways. But they can also be costly economically. To recognize a right to health care requires that a government levy taxes on its citizens to raise the funds necessary to carry out their duty to provide such health care. Precisely because rights impose costs on

15 Tasioulas, "On the Foundations of Human Rights," 59. Notably, while romantic love displays this feature, other forms of love, such as parental love, do not—that is, we typically expect that a parent's expression of love for her child will make a significant difference to her child's interest in receiving her love even if both recognize the parent as duty bound to do so. Thanks to an anonymous reviewer for calling attention to this exception.

others, then, it seems to me that a right can be (all things considered) justified if and only if it is the most efficient way of delivering benefits to the interests on which it is grounded. In other words, if a right is justified, then we should not be able to identify an alternative right that delivers the same (or greater) benefits to the relevant interests while imposing lesser costs or burdens on others.[16]

This idea that rights ought to be efficient is, I hope, an intuitive and uncontroversial expectation. Nevertheless, allow me to offer an example that helps both to clarify the concept and to show why it is a legitimate expectation of rights. Suppose that we all have an interest in having access to drinking water. Moreover, suppose that this interest is weighty enough as to recognize others to be duty bound by it. How ought we to specify the right that is grounded by such an interest? One construction of the right would be to recognize a universal right to glacial water. Counting in favor of this construction is its efficacy—the regular provision of glacial water to all would no doubt make a significant difference to the promotion of our interest in having access to drinking water. But while this construction of the right would be effective at promoting the relevant interest, it is clearly not efficient at doing so. For compare it to an alternative construction of the relevant right. On the alternative construction, we might recognize a universal right to some form of drinking water—whether it be sourced from a glacier, a spring, or a well. This alternative construction seems to deliver the exact same benefits as the initial construction but at considerably lesser cost (insofar as there are greater opportunities to discharge the corresponding duty). And since there is an alternative right that would deliver the same benefits at a lesser cost, we lack any sort of compelling reason to recognize a right to glacial water. The inefficiency of the right constitutes grounds for its rejection.

3. THE NONPECUNIARY BENEFITS OF WORK

In this section, I identify three nonpecuniary interests promoted by work: self-esteem, self-realization, and social relations.[17] Let us take them each in turn.

16 Another way of putting this thought is that rights ought not to be *unduly* burdensome—either on those they bind with duties or on affected third parties. And one way in which the burdens generated by a right might be unduly burdensome is when they are not necessary to promote the interests that ground the right. In other words, if we can identify an alternative right that promotes the relevant interests without generating similar burdens, then the burdens imposed by the original right are unduly burdensome in a way that licenses rejection of the right. Thanks to an anonymous reviewer for urging me to clarify this connection.

17 The three interests I identify are derived from an attempt to distill the many and often overlapping lists of interests appealed to by both advocates and skeptics of the right to work. For particular examples, see Collins, "Is There a Human Right to Work?"; Elster, "Is There (or Should There Be) a Right to Work?"; Forstater, "Working for a Better World"; Kildal,

3.1. Self-Esteem

The first nonpecuniary interest promoted by work is the interest we have in experiencing self-esteem.[18] The idea here is that the experiences of self-esteem tend to depend at least in part on the perception that one's "person and deeds [are] appreciated and confirmed by others."[19] When we engage in work, we are, according to my definition, engaging in an activity that makes a productive contribution to serving the interests of others. In making such contributions, we make possible an avenue through which to be esteemed and appreciated by others, which in turn forms the basis of our self-esteem.

3.2. Self-Realization

Work also serves to promote our nonpecuniary interest in self-realization—that is, the development and realization of one's full potential as a rational and purposive being.[20] According to Jon Elster, who writes extensively on the relationship between self-realization and work, self-realization is comprised of two distinct elements.[21] First, it involves self-*actualization*, or the actualization of those talents, skills, and abilities that exist in the first place as mere potentialities. Second, it involves displaying these powers and abilities in a way that can be observed and evaluated by others. This is what Elster calls *self-externalization*.

To see more clearly the connection between self-realization and work, consider how the particular features of work make possible each of its elements. The active nature of work involves actualizing our latent potentialities and,

"The Social Basis of Self-Respect"; Kirchgässer, "Critical Analysis of Some Well-Intended Proposals to Fight Unemployment"; Nickel, "Is There a Human Right to Employment?"; Schaff, "A Right to Work and Fair Conditions of Employment"; and Tool, "Employment as a Human Right." Note, however, that the provision of these nonpecuniary benefits is not exclusive to work. They are not, in other words, unavailable in a so-called "post-work" society.

18 See, e.g., Nickel, "Is There a Human Right to Employment?"; and Schaff, "A Right to Work and Fair Conditions of Employment." For arguments that dispute the claim that a right to work might be grounded in considerations of self-esteem, see Collins, "Is There a Human Right to Work?"; Elster, "Is There (or Should There Be) a Right to Work?"; and Kildal, "The Social Basis of Self-Respect."

19 Rawls, *A Theory of Justice*, 386.

20 See, e.g., Collins, "Is There a Human Right to Work?"; Forstater, "Working for a Better World"; Mundlak, "The Right to Work"; Nickel, "Is There a Human Right to Employment?"; Steinvorth, "The Right to Work and the Right to Develop One's Capabilities"; and Tool, "Employment as a Human Right." Elster offers a nuanced analysis of the relationship between self-realization and work in "Self-Realisation in Work and Politics" but nevertheless doubts that this relationship can ground a right to work in "Is There (or Should There Be) a Right to Work?"

21 Elster, "Self-Realisation in Work and Politics," 133.

ideally, honing the acuity and efficiency with which we exercise our various powers and abilities. And the productive aspect of work means externalizing our activities such that their outputs may be evaluated by managers or clients. Work thus looks to be a central means through which one might pursue an interest in self-realization.[22]

3.3. Social Relations

The final nonpecuniary interest that might be served by work is the interest in maintaining rewarding social relations with others.[23] Notably, Brownlee argues that there are in fact two interests captured by our more general interest in maintaining social relations, and indeed, they are weighty enough as to be considered basic needs.[24] (More on this later.) The first is the interest we have in maintaining meaningful social connections to others. Brownlee looks here to the work of psychologists Roy F. Baumeister and Mark R. Leary, who observe that "human beings have a pervasive drive to form and maintain at least a minimum quantity of lasting, positive, and significant interpersonal relationships."[25] The second interest is what she calls the "need to be needed."[26] Here, the idea is that we need not only to connect with others but also to be *useful* to them by way of supporting and promoting their well-being.[27]

How might work serve these needs? Importantly, Baumeister and Leary believe that satisfying the drive for social connection requires "frequent, affectively pleasant interactions with a few other people … in the context of a temporally stable and enduring framework of affective concern for each other's

22 Nevertheless, it must be noted that not all work serves to fulfill one's interest in self-realization. Certain conditions must be met if work is to play this kind of role. Elster notes, "Monotonous, repetitive tasks … are not conducive to job satisfaction," that ideally individuals ought to be placed in positions that match their skills, abilities, and interests, and that individuals must experience a sense of agency with respect to their vocation ("Self-Realisation in Work and Politics," 144). To the extent that such conditions are met, however, it is difficult to conceive of a greater site for self-realization than the workplace.

23 This is less commonly appealed to than the first two interests. Tool, "Employment as a Human Right"; Kildal, "The Social Basis of Self-Respect"; and Elster, "Is There (or Should There Be) a Right to Work?" all entertain (though Kildal and Elster ultimately reject) the possibility that the right to work might be grounded by an interest in maintaining robust social connections.

24 Brownlee, *Being Sure of Each Other*, ch. 1. See also Gordy et al., "The Missing Measure of Loneliness."

25 Baumeister and Leary, "The Need to Belong," 497.

26 Brownlee, *Being Sure of Each Other*, 16.

27 Brownlee, *Being Sure of Each Other*, 16.

welfare."[28] Thus, it might be argued, work responds to our need for social connection by providing such a framework.[29] Moreover, because work is socially productive, it is an opportunity to meet our need to be needed by making ourselves useful to others.

4. Assessing the Right to Work

We are now well placed to carry out the first aim of the article: to test the nonpecuniary benefits of work against the requirements of efficacy and efficiency. I begin by testing the efficacy with which recognizing a right to work would promote the three nonpecuniary interests listed above. Ultimately, I show that recognizing such a right would be self-defeating from the perspective of our interest in self-esteem. Moreover, I show that the capacity of such a right to make a significant difference to our interests in self-realization and social relations would require an implausibly costly construction of it. I then test the efficiency of the right by considering whether it is remuneration or merely the act of social contribution that explains work's capacity to deliver nonpecuniary benefits. For if it can be shown that these nonpecuniary benefits are a function of merely engaging in social contribution, then we will presumably be able to deliver them by recognizing a less costly right to (socially) contribute.

4.1. The Inefficacy of the Right to Work

I do not doubt that work is capable of promoting our interests in self-esteem, self-realization, and social relations. Nevertheless, I demonstrate here that recognizing a *right* to work would undermine the capacity of work to promote our interest in self-esteem. Moreover, I show that the capacity of work to make a significant difference to our interests in self-realization and social relations depends on our constructing the right in an implausibly costly way. Let us take each of these claims in turn.

The basis of the first claim—which is, admittedly, somewhat conjectural—derives from the fact that self-esteem is largely a function of the esteem one perceives from others. The issue is that coming to be employed through highly visible public entitlements (i.e., job guarantees) might actually prevent both us and others from seeing our work as worthy of esteem.[30] For others may be unlikely to see our work as worthy of esteem if it is understood to be the result

28 Baumeister and Leary, "The Need to Belong," 497.

29 To be sure, some jobs demand solitude to be performed effectively, meaning that not all forms of work serve interests in social relations.

30 This criticism is most forcefully articulated in Kildal, "The Social Basis of Self-Respect"; and Elster, "Is There (or Should There Be) a Right to Work?"

of a "government handout" and not something we have "earned."[31] Indeed, empirical evidence suggests that a different form of non–merit-based hiring—preferential selection—may have deleterious effects on the self-perceptions of its beneficiaries and may contribute to a "stigma of incompetence" in how they are perceived by others.[32] And we might think these issues will only be exacerbated if the productive contributions that result from one's guaranteed employment are minimal or invisible. It is hard to believe one's fellow citizens will come to see one as a valued contributor if the cost of providing employment outweighs one's productive contribution. In this way, grounding a right to employment in benefits to self-esteem might be self-defeating, for the very act of guaranteeing employment might serve to sever the connection between work and self-esteem.[33]

Might the relationship between the right to work and our interests in self-realization and social relations be similarly self-defeating? Ultimately, I do not think such a worry obtains here (at least to the same degree). But closer scrutiny of the relationship between the right to work and these two interests does reveal how costly such a right would be if it is to effectively serve these interests. For instance, not just any form of employment will do to serve our interests

31 Note that the conjecture on offer is thus limited to societies in which such a general attitude exists. One can, in other words, imagine a society in which there is no such general attitude, in which case this conjecture would fail to apply. Nevertheless, absent compelling reasons to think such an attitude will dissipate, this particular limitation of the objection does not strike me as problematic.

32 It is difficult to find empirical data on the relationship between guaranteed employment and self-esteem given the relative absence of existing programs. However, there is a related literature on the effects of a different form of non–merit-based hiring that seems to support the conjecture on offer: preferential selection. On the harm done to the self-perception of beneficiaries of preferential selection programs, see Heilman et al., "Intentionally Favored, Unintentionally Harmed?" On the "stigma of incompetence" displayed in others' perceptions of such beneficiaries, see Heilman et al., "Presumed Incompetent?" For a comprehensive review of this literature as it pertains to women, see Heilman, "Affirmative Action."

33 Kavka is suspicious of the claim that the publicity of entitlements to work would interfere with the capacity of work to bring about self-esteem ("Disability and the Right to Work"). Against it, he raises two objections. First, he notes that even if it is true that jobs received through public entitlements are less effective a source of self-esteem than jobs won through merit, surely jobs received through public entitlements are a better source of self-esteem than no job at all. Second, though how one comes to be employed may be among the relevant determinants of how one's employment is perceived by others, it is unlikely to be the sole or even most important determinant. Rather, it is more likely that the esteem of others reflects perceptions of one's contributions while one is on the job. Neither of these observations offers a decisive reason to recognize a right to work, however, because they fail to respond to the objection from inefficiency, which is articulated below.

in social relations.[34] To serve this interest, employment must be enduring and stable, which might make an effective implementation of the right to work considerably more costly. The state would not be able to merely direct labor resources where they are most needed. Instead, if the supposed benefits to our interest in social relations are to be achieved, the state may have to sacrifice its ability to flexibly respond to labor shortages in particular domains in order to ensure stable employment of the kind that permits enduring relationships.

Similar considerations hold for our interest in self-realization. Employment that contributes to self-realization cannot amount to mere drudgery. Indeed, self-realization in the domain of work often requires that our employment aligns in some way with our particular interests or plans of life and that it offers a degree of complexity.[35] But surely providing employment that meets these conditions would be a costly endeavor. Even moderate scarcity requires that we emphasize employing persons in positions that create more resources than they consume. For this reason, we cannot, for instance, guarantee the opportunity to "direct epic color films," even if this is one's preferred (or sole) path to self-realization.[36] To dedicate resources in this way would certainly interfere with the state's capacity to carry out other important functions. Thus, the right to work appears either to be incapable of making a significant difference to our interests in social relations and self-realization or else to require imposing what might seem to be unjustifiable costs on taxpayers and the state more generally.

4.2. The Inefficiency of the Right to Work

To assess the efficiency of the right to work at promoting the relevant set of nonpecuniary interests, I propose the following test. What we want to know is whether these benefits might be generated by the recognition of a similar but less costly or burdensome right. The obvious point of comparison here is a right to (socially) contribute without remuneration, in contrast to a right to remunerated employment. Thus, we can test the efficiency of the right to work, as opposed to a right to contribute, by considering the extent to which its purported benefits derive from the specifically remunerative aspect of work. If the benefits are derived primarily from this aspect of work, then we should expect them to be exclusive to the right to work and therefore a legitimate basis upon which to justify the recognition of this right rather than a right to contribute. In contrast, if it is the case that the benefits in question are best explained in

34 Elster, "Is There (or Should There Be) a Right to Work?" 64–66, 76.

35 For a more thorough analysis of the ways in which work must be structured if it is to allow for self-realization, see Elster, "Self-Realisation in Work and Politics."

36 Elster, "Is There (or Should There Be) a Right to Work?" 77.

terms of making social contributions, then we should expect a right to contribute to deliver similar benefits without taking on the costs associated with remuneration. In other words, at least from the perspective of the relevant nonpecuniary interests, we would seem to lack any grounds upon which to prefer the recognition of a right to work rather than a right to contribute. As such, in what remains of this section I return to each of the interests offered as grounds for the right to work and consider what aspect of work best explains its ability to promote them.

Beginning with self-esteem, let us retrace the connection to work. We noted earlier that self-esteem is largely resultant of the esteem one perceives from others. In this way, if the remunerative aspect of work is to be the basis of its capacity to promote and protect our interest in self-esteem, it must be because our contributions are held in esteem by others only to the extent they are remunerated. Clearly this is implausible. Indeed, I take it that in all but the most exceptional of cases, the esteem we afford to others is, in the first place, a function of their social contributions and the value we attach to them. In other words, it is the contributive element, not the remunerative one, that does the work here, suggesting that a right to contribute would fare just as well (if not better) at promoting our interest in self-esteem as the right to work.

Similar considerations apply to our interest in self-realization. I see no reason why we should understand the capacity for the right to work to promote or protect such an interest to be at all a function of its remunerative guarantee. Rather, it is the two elements that work shares with social contribution that make sense of its connection to self-realization. Work allows us to realize our potential powers and abilities if and because it requires us to engage in activities that call on us to develop these powers and abilities, and it allows us to externalize our exercise of these powers and abilities so that they may be recognized and evaluated by others. That we are paid in exchange for the exercise of such powers and abilities does not seem to play any direct or necessary role in our pursuit of self-realization.

So too for our interests in social relations. If work contributes to the protection and promotion of our social connection interests, it is indirectly as a result of the fact that the activities that make up work are typically done alongside or in cooperation with others—remuneration plays no role here. And more directly, it is obvious that the need to be needed is served by contributing to the promotion and protection of others' interests. Whether or not we are remunerated for such contributions is beside the point.

It seems clear, then, that the capacity for work to promote the relevant set of nonpecuniary interests is a function of its involving social contribution rather than the mere fact of its being the object of remuneration. And assuming that

the guarantee of remuneration renders the right to work a more costly way of delivering benefits to these interests than a right to contribute, the right to work fails to meet the efficiency requirement—at least in the case of our nonpecuniary interests.

At this point, I take myself to have successfully demonstrated the inefficacy and inefficiency of the right to work vis-à-vis the relevant set of nonpecuniary interests. What this seems to entail is that defenders of construing the right to access an income in the form of a right to work rather than a UBI cannot appeal to the nonpecuniary benefits of work to defend this preference. For these nonpecuniary benefits might be met equally well (and in a less costly way) by providing opportunities to engage in nonremunerated social contribution. What must be shown, then, if the supremacy of the right to work is to be defended, is that construing the right to an income in the form of a UBI would be a less effective and/or less efficient way of promoting the relevant *pecuniary* interests.

For what remains of the article, however, I set this question aside in an effort to pursue an alternative line of inquiry. My hope is to show that where the nonpecuniary benefits identified above have failed to make a decisive case for recognizing a right to work, they are well placed to ground the right to contribute.

5. BROWNLEE ON SOCIAL CONTRIBUTION

As noted above, I set aside for the remainder of the article the question of whether we ought to recognize a right to a basic income and ask instead whether the nonpecuniary benefits of social contribution detailed throughout the article ground a right to contribute. Importantly, it is not enough to merely demonstrate that social contribution does indeed deliver these benefits. For we additionally want to know whether a right to contribute would deliver these benefits effectively and efficiently. I carry out the assessment of the right to contribute along these dimensions in two steps. In this section, I further motivate the need to demonstrate the efficacy and efficiency of the right to contribute. I do so by rehearsing Brownlee's argument for recognizing the right to contribute and revealing that it remains incomplete precisely by virtue of lacking such a demonstration.[37] Given this, I return in the following section to the nonpecuniary interests surveyed above and demonstrate that a right to contribute would both effectively and efficiently generate benefits for at least two of these interests.

Central to Brownlee's argument for recognizing a right to contribute is the concept of a basic need. Basic needs, as I understand them here, represent a

37 Brownlee, *Being Sure of Each Other*, 75.

particularly weighty subspecies of interests. They are those "noncontingent" interests the fulfillment of which is "essential to a minimally good human life."[38] As a result, we typically treat basic needs as interests that command a particular kind of moral urgency. Indeed, we might think the idea of an associated right is internal to the concept of a basic need—that the term 'basic needs' merely describes a category of "deep, rights-grounding interests."[39] And even if this is not the case, certainly our basic needs are of sufficient weight that, though not essential to the concept, we have good reasons to recognize at least *prima facie* moral rights to their fulfillment.

Given this understanding of basic needs, then, Brownlee's argument is quite straightforward. As we earlier noted, Brownlee holds that as human beings, we have at least two basic social needs: the need for social connection and the need to be needed. We are needed, in turn, when we have the capacity to support the well-being of others. In other words, our need to be needed is satisfied only when we have opportunities to perform activities in service of others—that is to say, opportunities for social contribution. Thus, to the extent that our basic needs ground associated rights (either as a matter of necessity or as a matter of fact), our need to be needed suggests that we ought to recognize a corresponding right to contribute.[40] Because Brownlee's argument moves directly from the recognition of a basic need to the recognition of a corresponding right, to evaluate the argument is essentially to ask whether the need to be needed is, properly speaking, a basic need. On what grounds might she be making this assertion?

Brownlee seemingly makes two claims to support the assertion that the social needs she identifies are *basic* needs. Our social needs are basic needs, on her view, because they are "constitutive elements of a minimally good human life and … *preconditions for securing our other interests*."[41] Ultimately, however, it is this second claim that appears to do the heavy lifting—at least in the case of the need to be needed. That is, Brownlee's demonstration of the basicness of the need to be needed proceeds by identifying three antecedent interests served by social contribution. Let us take them in turn.

First, Brownlee suggests that the need to be needed can be explained in part by "narrow self-interest."[42] When we contribute to the survival or well-being of others, we make ourselves valuable to them and, in doing so, generate for

38 Gordy et al., "The Missing Measure of Loneliness," 12.

39 Brownlee, *Being Sure of Each Other*, 3.

40 Indeed, Brownlee refers interchangeably to "the need to be needed" and "social contribution needs" (*Being Sure of Each Other*, 16).

41 Brownlee, *Being Sure of Each Other*, 2 (emphasis added).

42 Brownlee, *Being Sure of Each Other*, 76.

ourselves a secure sense of belonging in the group or community within which we are recognized as a valued contributor. In contrast, when we lack this sense of security and belonging, our lives are severely impoverished.[43] Second, social contribution allows us to promote those interests we have in the well-being and survival of *specific* others.[44] When we stand in certain kinds of close personal relationships to others, we often care as much for their well-being and survival as we do our own. When we have opportunities to engage in social contribution, we can contribute directly to the interests of those close to us in important ways. Finally, Brownlee claims that our social contribution needs can be grounded in part by the interests of the groups and communities to which we belong and the ways in which group interests interact with our narrow self-interest and our interests in the well-being of those near and dear to us.[45] The fulfillment of social contribution needs provides us opportunities to contribute directly to the well-being of our groups, since such groups typically rely on the contributions of a great many of their members in order to carry out their central functions.[46] When we are able to contribute to groups such as these, we ensure the well-being of not only the group itself but also its individual members (including ourselves and those near and dear to us).

If one accepts Brownlee's claim that we have a basic need to serve as social contributors, the argument for a corresponding right follows naturally. Basic needs, as those interests the fulfillment of which is necessary to live a minimally worthwhile life, seem plausibly to be paradigmatic cases of the sorts of interests that reliably ground rights. If social contribution is indeed among such needs, as Brownlee argues, then clearly such needs ought to be responded to by recognizing a right to serve as a social contributor—the efficacy and efficiency of the right would appear to be indisputable. Thus, we seem to have grounds to recognize at least a *prima facie* right to opportunities for social contribution.

Yet we might still wonder how successful Brownlee's argument actually is. It is, for instance, slightly puzzling that she attributes the status of basic need to opportunities for social contribution only because the fulfillment of this

43 In addition, Brownlee makes the insightful observation that among our most important interests is an interest in being able to fulfill our moral responsibilities (*Being Sure of Each Other*, 76). The ability to contribute in important ways to the lives of others often makes possible a kind of moral agency through which we might take on and fulfill a variety of moral obligations.

44 Brownlee, *Being Sure of Each Other*, 77.

45 Brownlee, *Being Sure of Each Other*, 77.

46 For instance, Sangiovanni observes that states that regulate and provide important goods such as security and a stable system of property rights and entitlements rely on the contributions of citizens to do so ("Global Justice, Reciprocity, and the State").

"need" is instrumentally valuable for promoting a set of antecedent interests.[47] For we might thereby wonder whether it is really our narrow self-interest, interests in the well-being of specific others, and interests in the well-being of the communities to which we belong that represent our basic needs, and whether the opportunity to socially contribute represents just one way of serving these more fundamental needs rather than a *basic* need itself. As Cheshire Calhoun puts it, Brownlee's argument "tends to imply that the right to contribute is not a basic right but is merely instrumental to securing other more basic rights."[48]

Allow me to spell out the problem a bit more clearly. On one construction of Brownlee's argument, we proceed directly from the claim that we have a basic need for opportunities to socially contribute to the claim that we have a corresponding right to such opportunities. As such, whether we accept the conclusion of the argument ultimately depends on whether or not such a basic need indeed exists (since we can be relatively certain that if it does, a right to opportunities for social contribution would be effective and efficient in responding to it). But to demonstrate the existence of such a need, Brownlee merely shows how opportunities for social contribution serve to promote three kinds of interests that seem to stand separate from the need to be needed. And this means that her argument no longer proceeds directly from the identification of a basic need to the recognition of a corresponding right. For it now seems as though social contribution is merely instrumental to the promotion of these three *other* interests rather than being an essential feature of their content.

Given this, we might further doubt that there is any sort of *necessary* connection between social contribution and feelings of belonging, the security of our loved ones, or the well-being of our communities. In other words, it remains an open question whether the instrumental value of social contribution vis-à-vis *these* interests is sufficient to justify recognizing a right to contribute. This means we need to engage in the same kind of analysis to which we subjected the right to work. We need to know whether the instrumental relation between opportunities for social contribution and these supposedly rights-grounding interests is effective and efficient. For whereas we can be reasonably certain of the efficacy and efficiency of a right to contribute vis-à-vis a basic need to contribute, we cannot make this same assumption if the relevant interests are the more diffuse interests to which Brownlee ultimately appeals.

This limitation of Brownlee's argument suggests a need to demonstrate the efficacy and efficiency of the right to contribute vis-à-vis some set of grounding

47 The instrumental focus of Brownlee's argument as it concerns social contribution is documented and critiqued in Calhoun, "Social Connections, Social Contributions, and Why They Matter."

48 Calhoun, "Social Connections, Social Contributions, and Why They Matter," 460.

interests. It is to this end that the remainder of the article is dedicated. Note, however, that I pursue this demonstration not by way of the interests Brownlee identifies but rather by way of the set of nonpecuniary interests described in section 3.

6. Assessing the Right to Contribute

Recall the earlier observation that the capacity for work to deliver benefits to our interests in self-esteem, self-realization, and social relations is a result of its involving exercises of social contribution. Nevertheless, we cannot thereby conclude that a right to contribute exists. Given there will no doubt be costs associated with recognizing this kind of right, we want to be sure that such a right would effectively promote the interests in question and would do so in a way that could not be replicated by recognizing a less costly construction of the right. In what remains of this section, I show that the kind of efficacy and efficiency we were unable to identify in the case of the right to work can indeed be found in the case of the right to contribute. In particular, my strategy is to demonstrate that the very pursuit of our interests in self-esteem and self-realization depends on our having access to opportunities for recognized forms of social contribution—that there is a necessary connection between them.

To see the connection between self-esteem and social contribution, recall once more the connection between self-esteem and how we are perceived by others (or how we think we are perceived by others).[49] Notably, it is not just how we are perceived by others as persons that matters. According to Rawls, self-esteem, or self-respect, depends on the perception that our *deeds* are esteemed and appreciated by others.[50] That is to say, the satisfaction of our interests in self-esteem depends on our engaging in some activity the results of which admit of the possibility of esteem. And to admit of the possibility of esteem, it would seem, requires that the activity in question serves interests of some sort. Whether these interests are the self-interests of the actor, the interests of the appraiser, the interests of an unrelated party, or some combination of the three, it is difficult to conceive of an activity that admits of esteem but does not contribute in any way to the interests of some person or set of persons. In this way, because self-esteem itself seems to depend on the opportunity to act in ways that admit of appreciation by others and because such appreciation is largely a reflection of the degree to which one's activities contribute to the

49 Rawls, *A Theory of Justice*, 386.

50 Rawls, *A Theory of Justice*, 386–87.

fulfillment of interests, it is difficult to see how this fundamentally social aspect of self-esteem might be realized absent opportunities for social contribution.[51]

Similar considerations apply in the case of self-realization. Though definitions of self-realization range, and the term 'self-realization' itself may be substituted with similar terms such as 'self-development' or 'self-actualization', it nevertheless seems plausible to assume that self-realization consists to some degree or another in developing, honing, and exercising the capacities that make one *who one is*. In this way, self-realization seems to be at once a good *and* an activity. But what exactly does this activity consist of? What is the activity of developing, honing, and exercising one's capacities? I want to suggest that one develops, hones, and exercises the capacities that make one who one is by engaging or participating in acts of social contribution. I take this to be the case because of what Elster calls the self-externalizing component of self-realization.[52] For self-realization to occur, we must make the exercise of these capacities available to be evaluated by others. In turn, others tend to evaluate the acuity with which we exercise these capacities in terms of their contributive outcomes. Thus, to engage in the very activity of self-realization, in which we, among other things, develop and exercise capacities in a way that makes them available for evaluation, we must possess opportunities to exercise our capacities in those socially contributive ways that allow others—and ourselves—to appraise them.

Thus, it appears that fulfilling our interests in self-esteem and self-realization depends crucially on having opportunities to engage in social contribution, which in turn suggests that the right to socially contribute not only is effective but passes the test of efficiency.[53] What I hope to have demonstrated in this

51 It might be objected that the self-esteem derived from guaranteed opportunities for social contribution is vulnerable to the same conjecture I raise against the right to work in section 4. However, I take it that much of the stigma associated with non–merit-based hiring practices results from an attitude of resentment toward the idea of persons being remunerated for positions they did not "earn," so to speak. Because the right to contribute does not involve a right to remuneration, it strikes me as considerably less vulnerable to the conjecture raised against the right to work (if at all).

52 Elster, "Self-Realisation in Work and Politics," 133.

53 It might be wondered whether the right to contribute is vulnerable to a similar worry about efficiency as the right to work. That is, it might be the case that we could promote these interests in a less costly way by systematically misleading people into thinking social value attaches to a set of activities that are in fact without value but for which the cost of creating opportunities is low. There are two lines of response to such a worry. First, it is not clear to me that the scenario described is indeed a realistic one. It strikes me that it will always be reasonably apparent whether a given activity serves the interests or contributes to the well-being of another. Second and perhaps more importantly, even if such a state of affairs could be brought about, this merely suggests that the relevant right might be the right to contribute in whatever ways one's society *deems* valuable rather than the right to

section, then, is that despite failing to ground a right to work, we should not conclude that the nonpecuniary interests surveyed throughout this article are incapable of grounding any rights whatsoever. Where they fail to establish a decisive case for recognizing a right to work, they succeed in demonstrating the need for a right to contribute.

7. CONCLUSION

How should we interpret the universal right to access an income, assuming such a right exists? Two constructions of the right have seen increasing support in recent years. On the one hand, there is support for construing the right to access income in the form of a universal basic income. But others remain unconvinced that this is the most promising way to deliver on the right to access income. Instead, they argue that the relevant entitlement ought to be provided in the form of guaranteed opportunities for remunerated employment. In support of this position, it is observed that work provides not only income but a host of nonpecuniary benefits to our interests in self-esteem, self-realization, and social relations.

But if the nonpecuniary benefits of work are to play a decisive role in the case for establishing a right to work rather than a UBI, we ought to be able to demonstrate the efficacy and efficiency of the right to work in providing these benefits. I argue that no such demonstration is possible. Rather, the right to work is likely to prove both a decidedly ineffective way of promoting the relevant set of nonpecuniary interests, and inefficient in its promotion of these interests when compared against a less costly right to contribute. As such, the nonpecuniary benefits of work cannot serve as grounds upon which to assert the supremacy of the right to work as a way of construing the right to access income.

Nevertheless, where a demonstration of the efficacy and efficiency of the right to work vis-à-vis the relevant set of nonpecuniary interests proves unsuccessful, I argue that a successful demonstration *is* possible in the case of the right to contribute. For we have good reason to think that the fulfillment our interests in self-esteem and self-realization depends in some essential way on our having opportunities to engage in social contribution. Thus, having laid the theoretical grounds upon which such a right may be justified, future research efforts are well placed to pursue more practical questions such as how

make an *actual* contribution. For it is not obvious to me that the social basis of our interests in self-esteem or self-realization depends in any way on our making a contribution of objective value. All that is required is that one's contributions be esteemed by others or that they are available for their appraisal, both of which seem possible under the scenario described. Thanks to an anonymous reviewer for raising this worry.

opportunities for social contribution might be generated for or fairly distributed to those who currently lack them.[54]

Queen's University at Kingston
13jpd@queensu.ca

REFERENCES

Baumeister, Roy F., and Mark R. Leary. "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation." *Psychological Bulletin* 117, no. 3 (1995): 497–529.

Brownlee, Kimberley. *Being Sure of Each Other: An Essay on Social Rights and Freedoms*. Oxford University Press, 2020.

———. "The Lonely Heart Breaks: On the Right to Be a Social Contributor." *Aristotelian Society Supplementary Volume* 90, no. 1 (2016): 27–48.

Calhoun, Cheshire. "Social Connections, Social Contributions, and Why They Matter: Comments on *Being Sure of Each Other*." *Criminal Law and Philosophy* 17, no. 2 (2023): 453–62.

Collins, Hugh. "Is There a Human Right to Work?" In *The Right to Work: Legal and Philosophical Perspectives*, edited by Virginia Mantouvalou. Hart Publishing, 2015.

Den Otter, Ronald C. "A Constitutional Right to a Universal Basic Income." *New Political Science* 46, no. 2 (2024): 171–92.

Elster, Jon. "Is There (or Should There Be) a Right to Work?" In *Democracy and the Welfare State*, edited by Amy Gutmann. Princeton University Press, 1988.

———. "Self-Realisation in Work and Politics: The Marxist Conception of the Good Life." In *Alternatives to Capitalism*, edited by Jon Elster and Karl O. Moene. Cambridge University Press, 1989.

Forstater, Matthew. "Working for a Better World: Cataloging Arguments for the Right to Employment." *Philosophy and Social Criticism* 41, no. 1 (2015): 61–67.

Gordy, Ariel, Helen Han Wei Luo, Margo Sidline, and Kimberley Brownlee. "The Missing Measure of Loneliness: A Case for Including Neededness in Loneliness Scales." *International Journal of Environmental Research and Public Health* 19, no. 1 (2022): 1–18.

54 This article draws on research supported by the Social Sciences and Humanities Research Council. I owe a special thanks to Will Kymlicka, Jaclyn Rekis, the audience of the Queen's University Philosophy Graduate Colloquium, and two anonymous reviewers for their helpful feedback on an earlier draft of this paper.

Harvey, Philip. "The Right to Work and Basic Income Guarantees: Competing or Complementary Goals?" *Rutgers Journal of Law and Urban Policy* 2, no. 1 (2005): 8–59.

Heilman, Madeline E. "Affirmative Action: Some Unintended Consequences for Working Women." In *Research in Organizational Behavior*, edited by Barry M. Straw and L. L. Cummings. JAI Press, 1994.

Heilman, Madeline E., Caryn J. Block, and Jonathan A. Lucas. "Presumed Incompetant? Stigmatization and Affirmative Action Efforts." *Journal of Applied Psychology* 77, no. 4 (1992): 536–44.

Heilman, Madeline E., Michael C. Simon, and David P. Repper. "Intentionally Favored, Unintentionally Harmed? Impact of Sex-Based Preferential Selection on Self-Perceptions and Self-Evaluations." *Journal of Applied Psychology* 72, no. 1 (1987): 62–68.

Hemel, Daniel J. "Basic Income as a Human Right?" In *Tax, Inequality, and Human Rights*, edited by Philip G. Alston and Nikki R. Reisch. Oxford University Press, 2019.

Kavka, Gregory. "Disability and the Right to Work." In *Philosophy and the Problems of Work*, edited by Kory Schaff. Rowman and Littlefield, 2001.

Kildal, Nanna. "The Social Basis of Self-Respect: A Normative Discussion of Politics Against Unemployment." *Thesis Eleven* 54 (1998): 63–77.

Kirchgässner, Gebhard. "Critical Analysis of Some Well-Intended Proposals to Fight Unemployment." *Analyse und Kritik* 31, no. 1 (2009): 25–48.

McKinnon, Catriona. "Basic Income, Self-respect, and Reciprocity." *Journal of Applied Philosophy* 20, no. 2 (2003): 143–58.

Merrill, Roberto, and Catarina Neves. "Unconditional Basic Income and State as an Employer of Last Resort: A Reply to Alan Thomas." *Basic Income Studies* 16, no. 1 (2021): 169–90.

Mundlak, Guy. "The Right to Work: Linking Human Rights and Employment Policy." *International Labor Review* 146, nos. 3–4 (2007): 189–215.

Nickel, James W. "Is There a Human Right to Employment?" *Philosophical Forum* 10, nos. 2–4 (1978–79): 149–70.

Paul, Mark, William Darity Jr., Darrick Hamilton, and Khaing Zaw. "A Path to Ending Poverty by Way of Ending Unemployment: A Federal Job Guarantee." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 4, no. 3 (2018): 44–64.

Rawls, John. *A Theory of Justice*. Rev. ed. Harvard University Press, 1999.

Raz, Joseph. *The Morality of Freedom*. Oxford University Press, 1988.

Sangiovanni, Andrea "Global Justice, Reciprocity, and the State." *Philosophy and Public Affairs* 35, no. 1 (2007): 3–39.

Schaff, Kory P. "A Right to Work and Fair Conditions of Employment." In *Fair*

Work: Ethics, Social Policy, Globalization, edited by Kory P. Schaff. Rowman and Littlefield International, 2017.

———. "Work, Technology, and Inequality: A Critique of Basic Income." In *The Future of Work, Technology, and Basic Income*, edited by Michael Cholbi and Michael Weber. Routledge, 2019.

Spengler, Joseph J. "Right to Work: A Backward Glance." *Journal of Economic History* 28, no. 2 (1968): 171–96.

Standing, Guy. "Why Basic Income Is Needed for a Right to Work." *Rutgers Journal of Law and Urban Policy* 2, no. 1 (2005): 91–102.

Steinvorth, Ulrich. "The Right to Work and the Right to Develop One's Capabilities." *Analyse und Kritik* 1 (2009): 101–13.

Tasioulas, John. "On the Foundations of Human Rights." In *Philosophical Foundations of Human Rights*, edited by Rowan Cruft, S. Matthew Liao, and Massimo Renzo. Oxford University Press, 2015.

Tcherneva, Pavlina R. *The Case for a Job Guarantee*. Polity Press, 2020.

Thomas, Alan. "Full Employment, Unconditional Basic Income and the Keynesian Critique of Rentier Capitalism." *Basic Income Studies* 15, no. 1 (2020): 1–38.

Tool, Marc R. "Employment as a Human Right." In *Employment, Technology, and Economic Needs*, edited by Jonathan Michie and Angelo Reati. Edward Elgar Publishing, 1998.

Van Parijs, Philippe, and Yannick Vanderborght. *Basic Income: A Radical Proposal for a Free Society and a Sane Economy*. Harvard University Press, 2017.

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · SEPTEMBER 2025
https://doi.org/10.26556/jesp.v30i5.4092

RECLAMATION AND THE EPISTEMIC OBJECTIONABILITY OF SLURS

Pasi Valtonen

Reclamation is a phenomenon whereby social groups, for their own ends, use oppressive language that otherwise targets them. Reclamation aims to neutralize the derogatory power of slurs by transforming their meaning. Here, I argue that by transforming the meaning of slurs, the members of oppressed groups challenge epistemic distortions that target them. To support this claim, I discuss a specific view of slurs. The introduced epistemic view holds that slurs are not only morally objectionable but also epistemically objectionable.

Before introducing the epistemic view, I discuss expressivism. I think that expressivism offers a strong explanation for the meaning of slurs and for reclamation. Expressivism holds that in addition to their truth-conditional content, slurs express a negative moral evaluation of the target. For example, the meaning of 'Boche' is something like "Boo the Germans!" In the context of reclamation, expressivism becomes a polarity-reversing view. The negative evaluation is transformed into a positive evaluation: "hooray" rather than "boo." The polarity reversal is a simple and effective explanation for reclamation. It is no surprise that the polarity-reversing views of reclamation are popular.[1]

However, expressivism is not the whole story. I argue that expressive views do not fully explain the meaning of slurs. This also hinders the chance of exhaustively explaining reclamation. To elaborate on this claim, I mainly discuss Robin Jeshion's expressive account of reclamation. She emphasizes two things in relation to reclamation: *self-labeling* and *self-definition*. The expressive polarity-switch view explains self-labeling well. Expressivism also offers an explanation for self-definition, but I argue that the expressive account of self-definition may not be a full picture of self-definition. To put it a bit crudely at this stage, I argue that expressivism does not have enough resources to account for a full picture of self-definition. A full picture of self-definition requires more than just "boo" and "hooray." Hence, I propose an alternative view of reclamation

1 See Cepollaro, "The Moral Status of the Reclamation of Slurs," 678, and "Let's Not Worry About the Reclamation Worry," 181–93. See also Ritchie, "Social Identity, Indexicality, and the Appropriation of Slurs," 155–80.

that stems from an inferentialist view of the meaning of slurs. According to the proposed view, derogation derives from negative stereotypes. The view also holds that the meaning of slurs is not only morally objectionable but also epistemically objectionable. The meaning of 'Boche' is epistemically objectionable because it enables the transition from a slurring statement like "Hans is a Boche" to a conclusion that Germans are cruel. From this perspective, I formulate a view of reclamation as a way to challenge epistemic distortions. It should be emphasized that the proposed view is not incompatible with expressivism. Rather, it details the picture. It can make sense of the polarity switch, and it can also make sense of challenging the negative stereotype in reclamation.

1. RECLAMATION AND SELF-DEFINITION

1.1. Identity and Self-Definition

Reclamation takes place in a complex social surrounding. I have discussed the wider social surrounding in "Reclamation and Authority." There, I discussed the authority over self-definition in reclamation, arguing that the authority remains with the target group even though sometimes members of out-groups can join reclamation. The in-group authority over reclamation enables a view according to which reclamation is analogous to other social phenomena like gender ascriptions. Recently, Quill Kukla and Mark Lance have argued that gender ascriptions are not just descriptions but rather "social negotiations over how someone will be positioned within social normative space."[2] Kukla and Lance emphasize that gender ascriptions should not be assessed just as true or false descriptions but as performatives with the right kind of authority. Similarly, reclamation depends on the authority of the target group. Crucially, both gender ascriptions and reclamation involve identity. So the question regarding the resources to express that identity turns out to be very important. I argue here that the epistemic view expands decisively the resources to investigate reclamation as an expression of self-identity.

The proposed view and expressivism adhere to the idea that the meaning of a slur changes in reclamation. According to Katherine Ritchie, this idea is rather popular. She notes that at least Mark Richard, Christopher Potts, Robin Jeshion, Christopher Hom, and Paul Saka adhere to the change in meaning.[3] However, there are also views according to which reclamation need not involve a change in meaning. For example, Christopher Davies and Elin McCready argue that reclamation need not involve a change in meaning. Reclamation as

2 Kukla and Lance, "Telling Gender," 1135.

3 Ritchie, "Social Identity, Indexicality, and the Appropriation of Slurs," 156–57.

an expression of solidarity within the in-group can be achieved with the conventional meaning of slurs. It is more likely to depend not on the meaning of slurs but on the identity of the speaker and the audience and on the intentions behind the usage.[4] I agree with these claims. These claims about the identity of speakers, audience, and intentions emphasize the authority of the in-group. Therefore, the claims are not incompatible with the idea that the reclamation involves a change in meaning.[5] Next, I discuss self-definition in more detail through the ideas presented by Jeshion, but before that I discuss another notion related to reclamation.

1.2. Self-Labeling

According to Jeshion, reclamation is about weapons control. The aim of reclamation is to take the slur word from the bigot's arsenal by neutralizing its derogatory force. Her expressive strategy proceeds to neutralize slurs *via* the polarity reversal. To use a real example, the polarity of 'queer' was, in the initial stage, reversed from negative to positive, and then in the long run, the term became the neutral term it nowadays is.[6] Jeshion's insightful view highlights two aspects of reclamation: self-labeling and self-definition. Let us start with self-labeling. Self-labeling simply means that the oppressed group starts to apply a derogatory term to themselves, thereby taking the sting out of the term. I agree that self-labeling is an effective way to diminish the negative power of slurs.

4 Davis and McCready, "The Instability of Slurs," 70.

5 In fact, the claims help to overcome an objection against the change-in-meaning view. In reclamation, slurs become polysemous, according to the change-in-meaning views. Slurs have a negative meaning and a reclaimed non-offensive meaning. Luvell Anderson and Ernie Lepore formulate an objection against this view ("Slurring Words"). According to them, polysemy "fails to explain why non-members cannot utilize a second sense. If it were just a matter of distinct meanings, why can't a speaker opt to use a slur non-offensively?" (42). See also Anderson, "Calling, Addressing, and Appropriation," 10. They continue that the change-in-meaning views need an explanation that details the rules governing the access to the appropriated meaning. I have argued that the in-group has the authority to grant and deny access to the non-offensive meaning (Valtonen, "Reclamation and Authorization, 463–73). Cassie Herbert quite rightly points out that reclamation has a hazardous nature. Unsuccessful reclamation can in fact reconstitute the mechanism of oppression ("Precarious Projects," 132). The distinction between reclaimed meaning and the standard derogatory meaning of slurs and the idea that the in-group has the authority over the reclaimed meaning bring significant relief to Herbert's worries about the hazards of reclamation. The distinction, coupled with the idea of authority, sets the conditions for successful reclamation. It has to be the nonderogatory meaning, and it has to be authorized by the in-group.

6 See Jeshion, "Pride and Prejudiced," 108–15, especially 113–15.

According to Jeshion's expressive explanation of the phenomenon of self-labeling, when speakers reclaim a slur, they switch the negative evaluation to a positive evaluation.[7] Adam Galinsky and his colleagues have studied the effects of self-labeling. The results of the study are very encouraging to proponents of views that rely on the combination of polarity reversal and self- labeling.[8] As Bianca Cepollaro points out, the study shows that self-labeling has important empowering effects.[9] Ritchie comments that the study found that "self-labeling with a slur increases an individual's sense of power and increases an observer's evaluation of both the self-labeler's power and the power of the target group. . . . Self-labeling led to decreased perceptions of negativity in the slur that was used to self-label."[10] The study is particularly important because it shows the measurable effects of self-labeling, as Ritchie points out. Self-labeling not only increases the sense of power within the group but also decreases the negative perception outside the group. In other words, the overall effect is that self-labeling takes the sting out of slurs.

It should be noted that the work of Galinsky et al. does not exclusively support polarity-reversing views. Rather, the study supports the idea that self-labeling works, and hence, the results derivatively support any effort to explain the phenomenon. For example, Mihaela Popa-Wyatt combines Galinsky's study with her game-theoretic view, according to which agents in a dialogue are assigned roles based on their power. Reclamation radically changes these power relations.[11] Claudia Bianchi also appeals to the study to support her echoic account, according to which reclamation echoes derogatory content with a dissociative effect.[12] However, as explained in section 2.3, the epistemic view introduced here has the necessary resources to accommodate the expressive explanation for self-labeling. That is why the focus here is on the expressive explanation of self-labeling.

1.3. Self-Definition

As important as self-labeling is, Jeshion argues that more is at stake in reclamation, self-definition. In her view, polarity reversal also affects the representation of the target group. The reversed polarity is used as "an identity-label, as a means

7 Jeshion, "Pride and Prejudiced," 121–22.

8 Galinsky et al., "The Reappropriation of Stigmatizing Labels," 2020–29.

9 Cepollaro, "The Moral Status of the Reclamation of Slurs," 676.

10 Ritchie, "Social Identity, Indexicality, and the Appropriation," 164.

11 Popa-Wyatt, "Reclamation," 159–76.

12 Bianchi, "Slurs and Appropriation," 35–44.

to self-define as a group on their own terms."[13] She connects self-definition with oppressive norms that affect social representations. Jeshion points out that social representations involve norms. *Mutatis mutandis*, the representations of oppressed groups involve oppressive norms. Jeshion argues that reclamation as self-definition aims to break these oppressive norms. She says that "acts of reclamation strive for a reversal of social norms," and reclamation does so "by breaking established linguistic conventions on an expression or representation that manifests and whose use reinforces those norms."[14] Indeed, there are norms that systematically oppress and stigmatize people, and reclamation strives to counter these norms.

Still, the proposed view offers a different take on self-definition. Just because the epistemic view has richer resources, it can challenge the negative stereotypes *and* the oppressive norms that derive from the stereotypes. While expressivism challenges the oppressive norms, the epistemic view goes straight to the source by challenging the stereotypes that are responsible for the oppressive norms. In section 2.3, it is elaborated that competent speakers (including the reclaimers) may not be aware of the specific stereotypes associated with the target group. In these cases, the polarity switch is very likely the right analysis. But often the users of slurs and, more importantly, the reclaimers are aware of the stereotypes. The epistemic view provides insight into these cases. According to the proposed view, oppressive norms derive from stereotyping, and the view explains how this specific way of thinking based on stereotypes can be challenged. For example, oppressive norms regarding women derive from the misogynist view that women are promiscuous. At the same time, the derogation associated with the slur 'slut' derives from this stereotype. This means that both the derogation associated with slurs and the oppressive norms stem from negative stereotypes. By challenging the stereotypes, the derogatory power of slurs *and* the oppressive power of norms are mitigated. According to the epistemic view, the goal of reclamation is neutralization. As Jeshion points out, reclamation is a process, and when the stereotypes are challenged, the aim is to purge the meaning of a slur from this objectionable element. Objectionability has a very special meaning in the context of the epistemic view, as detailed in the next section.

The proposed view also contrasts with Lauren Ashwell's view. While her view is not an expressive one, she still adheres to the idea that the derogation associated with slurs stems from oppressive norms. She argues that while 'sluts' refers to a subset of women—to those who are deemed promiscuous—the

13 Jeshion, "Pride and Prejudiced," 122.

14 Jeshion, "Pride and Prejudiced," 111.

oppressive constraint stemming from the stereotype affects all women.[15] Hence, the stereotypical conception leads to an oppressive norm that all women should be monitored. Ashwell also argues that the derogation associated with 'slut' derives from the oppressive norm, not from the stereotype.[16] I disagree. To me, it seems fairly obvious that the derogation derives from the stereotype.[17] According to the epistemic view, by challenging the meaning of 'slut', reclamation challenges the stereotypical conception and thereby the basis of the oppressive norm.

2. RECLAMATION AND EPISTEMIC OBJECTIONABILITY

2.1. The Epistemic View of Slurs

The proposed view of reclamation is based on the epistemic view of the meaning of slurs. Hence, before discussing the epistemic view of reclamation, we need to explicate the epistemic view of the meaning of slurs. I have previously developed the epistemic view of the meaning of slurs.[18] In the following, I briefly explain the features that are relevant to reclamation.

The central claim of the epistemic view is that slurs are not only morally but also epistemically objectionable. The idea that slurs are epistemically objectionable dates back to the work of Michael Dummett. According to his inferentialist view, the inferential rules for terms determine the semantics for language. Conversely, the referential direction goes the other way around; valid inferential rules are determined on the basis of semantics. According to Dummett, the meaning of slurs is epistemically objectionable because the meaning of slurs allows one to infer unwarranted information. He says that the condition for the application of 'Boche' is that the target is German, but the consequence of the application is that the target is "barbarous and more prone to cruelty

15 I agree that slurs are often used to refer only to a subset of the target group. In section 2.1, I argue that the stereotype associated with the target group is attributed generically, allowing exceptions. Elsewhere I have argued that the interpretation of the generic element is the following. While only some women actively behave promiscuously, all women are disposed to such behavior, according to the misogynist mindset. This accommodates the subset usage.

16 Ashwell, "Gendered Slurs," 236–39.

17 Justina Diaz Legaspe points out that Ashwell fails to account for the full derogatory force of slurs. Even though yelling, "You are not acting according to what is expected from Black folks!" will most likely lead to a heated discussion about how exactly Black people should behave, it is certainly not as derogatory as yelling the n-word. See Diaz Legaspe, "Normalizing Slurs and Out-Group Slurs," 244. Given this, the negative stereotype provides a more plausible source for the derogation.

18 Valtonen, "Generic Inferential Rules for Slurs," 6533–51, and "Gendered Slurs and the Subset Argument," 762–79.

than other Europeans."[19] In the inferentialist framework, introduction rules specify the conditions for the introduction of a term. Here, the condition for the introduction of 'Boche' is that *x* is German. Then, elimination rules specify the consequences of introducing the term. In this case, the consequence of introducing 'Boche' is that Germans are cruel. For clarity, Dummett's proposed rules can be explicated as follows:

Boche-I: *x* is German; therefore, *x* is a Boche.

Boche-E: *x* is a Boche; therefore, *x* is cruel.

Inferentialism imposes a *harmony constraint* on the inferential rules. Briefly, the constraint says that the introduction rules and the elimination rules must match. Crucially, the Boche-I and Boche-E rules do not match. To use Ian Rumfitt's terminology, the rules are lacunose because Boche-E unpacks more than Boche-I packs in.[20] This results in an inferential gap, so the conclusion in Boche-E is unwarranted. If inferentialism is right, and the inferential rules explicate the meaning of language, then the *meaning* of 'Boche' is objectionable; it permits the attribution of cruelty to Germans without any actual evidence. The objectionability is based on an intuitive conception of deductive practice. In deductive reasoning, one should not add information at any step of the inference. One should only, so to speak, manipulate information already in the premises.[21] Boche-E clearly adds information, and the objectionability stems from that.[22]

The epistemic objectionability is the main target of Timothy Williamson's objection. According to him, just because of the epistemic objectionability, Dummett's view cannot explain how slurs are actually used. To illustrate, when we look at the semantics of the Boche-I and Boche-E rules, the extension of 'Boche' is a union of the German people and cruel people. In other words, 'Boche'

19 Dummett, *Frege*, 454–55.

20 Rumfitt, "'Yes' and 'No'," 785–89.

21 Dickie, "Negation, Anti-Realism, and the Denial Defence," 164.

22 The harmony constraint and epistemic objectionability have become the cornerstone of inferentialism since Buel Belnap formulated it as a response to Arthur Prior's objection. Prior argues that inferentialism does not have the resources to rule out bad inferential rules ("The Runabout Inference-Ticket," 38–39), but Belnap shows that, with the idea of harmony, inferentialism does have the necessary resources ("Tonk, Plonk and Plink," 130–34). Concerning the importance of harmony, see also Williamson, "Reference, Inference, and the Semantics of Pejoratives," 137–41; and Valtonen, "Generic Inferential Rules for Slurs," 6534–36. There are inferentialist views of slurs that do not rely on epistemic objectionability. Robert Brandom famously rejects the notion (*Articulating Reasons*, 66–79). Lynne Tirrell develops the "Brandomian" line of inquiry in her seminal work (e.g., "Derogatory Terms," 41–79). However, in this article, the aim is to motivate the idea that epistemic objectionability can be a useful explanatory notion regarding reclamation.

can refer either to the set of German people or to the set of cruel people. However, Williamson argues that this is not how slurs are used. Even though Stalin was undoubtedly cruel, he was not a "Boche"; he was Russian. Thus, 'Boche' is applicable first and foremost to German people. This reflects the neutral counterpart assumption, according to which slurs have neutral counterparts: 'Boche' is coextensional with 'German'; 'Frog' is coextensional with 'French'; and so on. In the next section, I discuss why the counterpart assumption is important for the epistemic view. However, for now, I concentrate on how the epistemic view can accommodate the counterpart assumption despite Williamson's objection.

Williamson thinks that the counterpart assumption is incompatible with epistemic objectionability. He goes on to argue that the rules can be easily made compatible with the counterpart assumption. One needs only to modify the elimination rule to a reversal of the introduction rule: "*x* is a Boche; therefore, *x* is German." However, the problem is that even though these rules now accommodate the counterpart assumption, the modified rules no longer support the claim that the meaning of slurs is objectionable. The introduction rule and the modified elimination rules match.[23] In the following, I aim to preserve the idea that slurs are epistemically objectionable *and* still accommodate the counterpart assumption.[24] To this purpose, I propose the following rules for 'Boche'. The conditions for the introduction of 'Boche' are the same as before ("*x* is German; therefore, *x* is a Boche"), but the elimination rule adds the information that Germans are typically cruel. The elimination rule can be more formally stated as

> Boche-E(Gen): *x* is a Boche; therefore, *x* is German $\wedge$ Gen *x* [German(*x*)] [cruel(*x*)]

This is a standard way to represent the structure of a generic statement. The stereotypical conception is attributed to Germans generically.[25] 'Gen' is a

23 Williamson, "Reference, Inference, and the Semantics of Pejoratives," 145–48.

24 In his response to Williamson, Daniel Whiting agrees with Williamson, and he formulates the rules broadly in the way Williamson suggests. He goes on to say, just like Williamson, that the derogatory part is fleshed out with a conventional implicature that Germans are cruel. See Whiting, "Conservatives and Racists," 375–88, especially 384–87. See also Williamson, "Reference, Inference, and the Semantics of Pejoratives," 149–52. By doing this, Whiting loses the objectionability but gains an explanation for the projection behavior of slurs. It is often thought that the projection behavior requires some kind of nontruth conditional element, and conventional implicature is just that—nontruth conditional. Nevertheless, I do think the epistemic objectionability is worth saving. For one thing, as I argue below, it offers a fresh and unique perspective on reclamation. However, I also have to address the issue of projection, which I do at the end of the next section.

25 The suggested analysis applies also to gendered slurs. The rules for 'slut' are roughly that 'slut' refers to a woman, and women are typically promiscuous; and 'bitch' refers to a

generic operator comparable to quantifiers, but unlike quantifiers, it does not specify the exact relationship between sets within the scope of the operator. In natural language, the generic operator can be substituted with terms like 'usually', 'generally', or 'typically', although it is often omitted altogether—as in "Tigers are striped."[26]

As I have explained elsewhere, an important part of this proposal is that the generic element receives a Leslie-style interpretation. Sarah-Jane Leslie argues that generics do not have truth conditions in the traditional sense.[27] The Gen operator that appears in Boche-E(Gen) does not contribute to the truth conditions of generics in the way that, for example, quantifiers do. Rather, generics are based on a psychological mechanism, and hence, they have much looser *worldly truth-makers*, as she calls them. The relevant truth-maker is the one related to striking-feature generics. According to Leslie, the psychological mechanism is regularly triggered in the presence of a striking feature, and often, the feature is something horrific or appalling.[28] In the case of "Germans are cruel," the mechanism is indeed triggered, but it goes without saying that the worldly truth-makers do not support this generalization. The most important point here is that the Leslie-style interpretation allows one to adhere to the counterpart assumption. The assignment of reference is the set of German people because the generic part does not contribute to the truth conditions of the conjunction in Boche-E(Gen). Hence, Boche-E(Gen) does not contribute to the assignment of reference. Assignment is done solely on the basis of Boche-I(Gen). To be clear, if the generic component did contribute to the assignment of the reference, it would result in a rather odd situation. The reference of 'Boche' would be indeterminate since it could be one of two sets: the set of German people assigned by the introduction rule or an empty set assigned by the Boche-E(Gen) since the false generic component falsifies the whole conjunction.

2.2. Counterpart Assumption and Objectionability

One of the virtues of the counterpart assumption is that it provides a simple and effective explanation of the semantics of slurs. For example, 'Boche' is a bad word for Germans, 'Frog' is a bad word for the French, and so on. According to the counterpart assumption, the truth-conditional contribution of slurs is the same as that of their neutral counterparts. Semantically speaking, the

woman, and typically women are overbearing. See Valtonen, "Gendered Slurs and the Subset Argument," 762–79. In section 2.5, I argue that the analysis also applies to slurs involving body shaming.

26 Leslie, "Generics," 1–6.

27 Leslie, "Generics and the Structure of Mind," 386–88.

28 Leslie, "Generics and the Structure of the Mind," 383–86.

epistemic view is compatible with expressivism. They both rely on the counterpart assumption. Although the proposed view is semantically compatible with expressivism, the view has something in common with what might be called "semantic views," like Christopher Hom and Robert May's semantic innocence.[29] According to the proposed view, there is something wrong with the meaning of slurs. Namely, the meaning is epistemically objectionable. Expressivism does not have this feature. According to expressivism, there is nothing semantically wrong with slurs. The term 'Boche' works just as it should when it expresses hostility toward a German. Rather, it is the expressed racist attitude that is objectionable. In contrast, Hom and May argue that the meaning of slurs is morally objectionable, and this objectionability seeps into the truth-conditional contribution of slurs, as unnegated slurring statements are systemically false. So based on their view of the moral objectionability of slurs, Hom and May reject the counterpart assumption.[30]

Eleonore Neufeld argues that slurs involve essentialization with a similar effect. She claims that slurs "designate an essence that is explanatorily connected to a set of negative stereotypical features of a social group."[31] As a consequence, she comes to a similar conclusion as Hom and May: the extension of slurs is empty. Leslie also says that generalization involves essentialization, which can have pernicious effects concerning social kinds since it can lead to generalizations like "Germans are cruel" and "Muslims are terrorists."[32] The comparison between Neufeld and the current proposal emphasizes Leslie's interpretation of generics and the claim that generics do not have a truth-conditional contribution in Boche-E(Gen). According to the proposed view, slurs are epistemically objectionable while the counterpart argument is maintained. Thereby, the view takes on board the appealing intuitions of expressivism and of semantic views. On the one hand, the epistemic view maintains the idea that slurs are bad names for groups. This is the contribution of the counterpart assumption. On the other hand, the epistemic view holds that there is something wrong with the meaning of slurs, just like semantic views argue. In fact, the epistemic objectionability only emphasizes the connection between the derogation and the counterpart assumption. Slurs are bad because they derogate targets on the basis that the targets belong to a certain demographic group. Despite your personal moral integrity and accomplishments, with slurs, you are reduced to a criminal, a slacker, or a vulgarian on the basis of a demographic

29 See Hom and May, "Moral and Semantic Innocence."

30 Hom and May, "Moral and Semantic Innocence," 294–300.

31 Neufeld, "An Essentialist Theory of the Meaning of Slur," 2.

32 Leslie, "The Original Sin of Cognition," 393–421.

marker.[33] The proposed view adds that this is done in an epistemically objectionable way. (The idea of epistemic objectionability is revisited in section 2.5, where the epistemic view of reclamation is detailed.)

The final note concerning the differences between expressivism and the epistemic view is about the projection behavior that is a much-discussed feature of slurs. This means that slurs project out of the scope of truth functional operators.[34] For example, the conditional "If Macron is a Frog, then so is his partner" is derogatory even though neither of the constituents are actually asserted. Similarly, slurs project out of indirect speech reports. In "John said Macron is a Frog," it is not clear whose attitude the slur expresses. Is it the speaker's or John's?[35] It has often been thought that to explain the projection, one needs to introduce nontruth conditional content like expressivism does. Since the derogatory content is nontruth conditional, it is not a big surprise that it escapes the scope of truth conditional operators.[36] Admittedly, the epistemic view does not attribute a separate nontruth conditional content to slurs. Hence, at first sight, one might think that it struggles to explain projection behavior. Nevertheless, there are alternative explanations for projection. One prominent is the *taboo* status of slurs associated originally with Luvell Anderson and Ernest Lepore's prohibitionism. According to prohibitionism, taboo words are not to be used or even mentioned in any context. This then explains the projection. Nonetheless, as Anderson and Lepore point out, embargoes are not absolute but often contain some caveats, and according to them, in-group use of slurs is one of them.[37] This gives enough room for reclamation. Similarly, the epistemic view argues that the derogatory nature of slurs leads to the taboo status, which then explains the projection behavior. Crucially, the explanation does not require nontruth conditional resources.[38]

33 Similarly, Deborah Mühlebach observes that slurs differ from ordinary swear words in that while a pejorative like 'asshole' might be derogatory, the derogation might be warranted. 'Asshole' targets individuals based on their actions or character, not on their membership in a specific social or demographic group. Mühlebach, "A Non-Ideal Approach to Slurs," 15–16. See also Diaz-Legaspe, "What Is a Slur?" 1399–422.

34 See McCready, "Varieties of Conventional Implicature," 12.

35 In comparison, in "John said that the moon is made of cheese," there is no doubt who has the silly belief.

36 See, e.g., Potts, "The Expressive Dimension," 167–76.

37 Anderson and Lepore, "Slurring Words," 25–48. See also Berkovski, "Slurs, Synonymy, and Taboo."

38 It seems to me that it is a very natural idea that the taboo status derives from the bad meaning of slurs. However, Anderson and Lepore disagree on this issue ("Slurring Words"). Their view is highly deflationary. They think that slurring terms and their neutral counterparts are synonymous in every way. It is just that at one point, the other one was declared

2.3. Self-Labeling and the Competence Objection

As noted, expressive views explain self-labeling well. The reason for this is the expressive conception of the meaning of slurs. At the same time, it is argued here that the epistemic view can tap into the expressive resources to explain self-labeling. In this section, I explain how the epistemic view can use the resources of expressivism. It is important to note that this discussion does not bring out any advantages concerning the epistemic view. In fact, in this section, it appears that expressivism has the upper hand as it is admitted that the competence objection is a powerful objection against the stereotype views. However, I argue that as a stereotype view, the epistemic view can handle the competence objection. Moreover, the offered solution does lead to an advantage that is detailed in the next section. In short, the advantage is the following. While expressivism and the epistemic view can both explain self-labeling, the epistemic view has more resources to explain self-definition; therefore, it provides a more detailed picture of reclamation as self-definition.

First, let us look at the dispute between expressive views and stereotype views. A stereotype view, like the epistemic view, holds that derogation stems from a stereotypical conception of the target. In contrast, according to expressivism, derogation is not based on any specific stereotypes; rather, derogation is based on an expression of a negative attitude toward the target. Instead of a specific stereotype, the meaning of 'Boche' adds a hostile attitude toward the Germans, similar to 'Boo the Germans!' As Jeshion points out, the source of the derogation is a nonpropositional attitude, like contempt. Even though contempt can be "a highly structured affectively and normatively guided moral attitude," it is still nonpropositional and nondescriptive. The contempt is not propositionally encoded.[39] In her paper "Slurs and Stereotypes," Jeshion rejects the stereotype view in the following way:

> I challenge its most fundamental claim—that a speaker who uses a slur thereby expresses and endorses a stereotype … of the group that the slurring term references.[40]

---

a prohibited word, and it became taboo. To emphasize, this declaration has nothing to do with the meaning or any other feature of the word (39). In other words, Anderson and Lepore argue that slurs are bad *because* they are prohibited. The current view argues that it is the other way around. Slurs are prohibited because they are bad. Recently, Stefano Predelli has presented similar thoughts, exploring the connection between taboo and various aspects of meaning ("Unmentionables," 726–44).

39 Jeshion, "Expressivism and the Offensiveness of Slurs," 316–18.

40 Jeshion, "Slurs and Stereotypes," 320.

From this, Jeshion goes on to formulate what might be called the *competence objection*. She points out that the "rationale for requiring the encoding of stereotypes" to the meaning of slurs "is questionable."[41] The ordinary competence in the use of slurs does not support it. Competence does not require knowledge of any specific stereotype; for example, a speaker can use 'Frog' competently without knowledge of any specific stereotype associated with French people.[42]

The competence objection is a genuine challenge for the stereotype view. Slurs can be used competently to express hostility even if one is completely ignorant about the stereotypes associated with the target group. This is just a fact about the *use* of slurs. However, it need not be a fact about the *meaning* of slurs. Therefore, in order to defend the stereotype view, I have previously distinguished between the use of slurs and the meaning of slurs. According to the stereotype view, derogation is based on a negative stereotype associated with the target group. I argue that stereotypes are part of the meaning of slurs. For example, 'Frog' involves the idea that French people are vulgar, and 'Boche' involves the idea that Germans are cruel. The meaning and the use of slurs are then two different things. It seems to me that the linguistic fact that a speaker can *use* 'Frog' to express general hostility toward French people derives from the meaning of 'Frog', which can include the attribution of a negative stereotype. With the distinction between meaning and use, the current view adheres to Dummett's conception of the social character of language. According to Dummett, none of us have perfect knowledge even of our native tongues. Even if a speaker is a native speaker of English, for example, they might be ignorant about the English vocabulary of theoretical physics. Similarly, Dummett says that one can have only partial knowledge concerning individual words and yet be a competent user of those words: "we constantly use words whose meanings we do not fully know, but we use them with confidence that what we are saying is true, and that we are therefore transmitting correct information."[43] He illustrates this with an example. Consider someone who knows that 'chess' refers to a board game played by two people but does not know the rules of chess. In this case, it is fair to say that the speaker has only partial knowledge of what 'chess' means. Yet the speaker can use the term correctly when the speaker says two of their friends play chess in the park every Sunday. Dummett also gives an example of complete ignorance. He says that his knowledge of the term 'gasket' is nonexistent. Still, he can pass on the information that his car does not work because the gasket is leaking. The guy at the garage said so. In this

41 Jeshion, "Slurs and Stereotypes," 320.

42 Jeshion, "Expressivism and the Offensiveness of Slurs," 321–23.

43 Dummett, *The Logical Basis of Metaphysics*, 84.

case, Dummett says that this is not real knowledge, and he does not know what 'gasket' means.[44] It seems to me that Dummett is merely quoting what the guy at the garage told him.

The crucial question for us is which example resembles the expressive use of slurs (without the knowledge of the stereotype), and it seems to me that the expressive use displays genuine partial knowledge. In fact, it seems to me that an expressive user of 'Frog' knows quite a lot. The speaker knows that 'Frog' applies to French people, and the speaker also knows that the term conveys contempt toward the French. As an upshot, I disagree with Jeshion's characterization of the "fundamental claim" of the stereotype view. The central claim is not that the competent use of a slur is always an expression and an endorsement of a particular stereotype. In light of previous discussion, this claim is too strong. Rather, the main claim is more likely on the lines of: even though the meaning of slurs includes stereotypes, partial knowledge can also result in competence. Just like the term 'chess' can be used competently without the knowledge of the rules of chess, slurs can be used competently without the knowledge of the stereotype. The adherence to Dummett's notion of partial knowledge enables the current stereotype view to explain the expressive use and also the meaning of slurs, which involves stereotypes. When slurs are used expressively, the view can attribute competence to the speaker on the basis of partial knowledge just because the knowledge of stereotypes is not required, but still, the full meaning of slurs involves stereotypes. One way to illustrate the proposed view is to compare it with Leopold Hess's recent view. He argues that in his inferentialist account, the truth conditional content of "Hans is a Boche" is that Hans is German. He goes on to argue that there is also explicit and implicit nontruth conditional content. The explicit nontruth conditional content is the speaker's contempt toward Germans. Additionally, the implicit nontruth conditional element is the stereotype that Germans are cruel.[45] While there are obvious similarities between my development and Hess's view, there are also slight differences. In my view, the meaning does not involve three components. My view involves only the rules that assign the reference of 'Boche' to Germans. Furthermore, the rules suggest generically that Germans are cruel. Given Dummett's view of competence, the latter part enables the expressive usage.

The outcome of the previous is that the epistemic view is able to borrow the expressive resources to explain self-labeling. Reclamation as self-labeling happens exactly like expressivism describes. Hence, expressivism provides a satisfactory explanation for self-labeling. As Ritchie points out, self-labeling is about

44 Dummett, *The Logical Basis of Metaphysics*, 83–84.

45 Hess, "Inferentialist Semantics for Lexicalized Social Meanings," 358.

changing the evaluation of the targeted group: "Galinsky [and others] found that self-labeling with a slur . . . increases an observer's evaluation of . . . the power of the target group."[46] Crucially, this change in evaluation leads "to decreased perceptions of negativity in the slur that was used to self-label." If we consider the epistemic view, the response to the previous objection regarding competence has a rather fortunate outcome. As a result of the distinction between the use and the meaning of slurs, the epistemic view can borrow the expressive explanation for self-labeling. Not only are expressivism and the epistemic view semantically compatible because they both adhere to the counterpart assumption, but the epistemic view also acknowledges the expressive use of slurs. So when slurs are used to express hostility toward the target, reclamation works as a polarity switch, as expressivism predicts. Nevertheless, taking a cue from Jeshion, I argue that in addition to self-labeling, reclamation also involves self-definition. The epistemic view argues that the meaning of slurs involves more than just hostility toward the target. The full meaning of slurs also involves stereotypes, and this becomes crucial in order to provide a detailed picture of reclamation and self-definition, as seen next.

2.4. Self-Definition and Public Representations

Jeshion argues that the point of reclamation is to break oppressive norms, but all descriptive elements are left out because the derogation associated with slurs—and hence the reclamation of slurs—does not contain any descriptive elements. In contrast, the epistemic view holds that the derogation associated with slurs and the oppressive norms both derive from stereotypes. An example of this situation is the oppressive conception that women must be monitored and controlled because they are prone to promiscuity, alongside the gendered slur 'slut', which captures this stereotypical conception. By challenging the meaning of 'slut', reclamation challenges the vehicle of the stereotypical conception of promiscuity and therefore undermines the basis for the oppressive norm. Since, according to the stereotype view, slurs encode stereotypes, the *primary* aim of reclamation as self-definition is to challenge the meaning of a slur that puts forward these stereotypes. As a consequence, the effects of the oppressive norms are mitigated. This view takes a cue from Elisabeth Anderson, who argues that stereotypes are part of a public representation, even if all parties reject it. According to Anderson, just because they are part of the public representation of the targeted group, harmful stereotypes hang on like a "cloud."[47] To illustrate her point, she describes an encounter she had in Detroit

46 Ritchie, "Social Identity, Indexicality, and the Appropriation of Slurs," 164.

47 Anderson, *The Imperative of Integration*.

with a Black man who helped when she was having car trouble. The encounter started with the stranger's assurance that he was there to offer help, not to rob Anderson. She goes on to point out:

> This man suffered a harm of racial stigmatization in this interaction. He was harmed, regardless of how he felt about it, and notwithstanding the fact that I refused to apply the stereotype of the criminally violent black male to him.... The harm consists in the fact that he walks under a cloud of suspicion in ... encounters with strangers. To gain access to cooperative interactions, he must assume the burdens of dispelling this cloud, of protesting and proving his innocence of imagined crimes.[48]

The proposed view suggests that reclamation is a way to protest and challenge the meaning of slurs as vehicles of harmful stereotypical conceptions, thereby contributing to the redefinition of the public representation of the target group. To connect reclamation to the two different conceptions of the meaning of slurs, you might say that reclamation of the n-word, according to expressivism, is something like "We are Black and we are proud!" The n-word refers to the same set as its neutral counterpart, but the negative evaluation, which is similar to contempt, is switched to a positive evaluation that is similar to pride. The epistemic view adds, "The n-word links criminality with Black people, and we need to change that!" The epistemic view sets the stage to view reclamation this way as it allows one to view the act of reclamation as a challenge to the meaning of a slur that unwarrantedly attributes a harmful stereotype to the target.

2.5. Challenging the Meaning

Reclamation concerns the meaning of slurs and aims to transform the meaning of the contested word. Expressivism aims to transform a pejorative word into a positive one. The epistemic view also aims to transform the meaning of slurs by challenging the epistemically objectionable part. Expressivism explains self-labeling with a polarity switch: a negative attitude toward a target is changed to a positive one. In contrast, the proposed view does not work like that. I propose that the epistemic challenge should not be understood as switching from a negative stereotype to a positive stereotype. Rather, the challenge should be understood as a challenge to the way a slur word captures a negative stereotype. The aim of reclamation is to *neutralize* the slur by transforming its meaning. Cepollaro notices that the polarity-switch strategy is potentially problematic for

48 Anderson, *The Imperative of Integration*, 53.

expressivism.[49] Specifically, if it is unwarranted to condemn someone merely on the basis of a demographic marker, then surely it is equally unwarranted to praise someone on the basis of that same demographic marker. She goes on to argue that this potential worry can be dealt with by comparing reclamation to affirmative action. Just like in affirmative action, a short-term imbalance can be justified with long-term benefits, such as "turning the slur to a less powerful weapon in the long run."[50] It seems to me that this is a good response to the potential worry. In a similar fashion, one can object against the proposed view that if the adding of the negative stereotyping makes slurs epistemically objectionable, then surely the attribution of positive stereotypes is equally objectionable. (This is discussed in detail below, but at this point, I just highlight that objectionability does not depend on the content of what is added; it derives from the fact that something is added in the first place.) One can counter this objection with a similar train of thought as Cepollaro defends expressivism. The long-term benefits of the attribution of positive stereotypes outweigh the objectionability of meaning. However, it is questionable whether positive stereotypes have long-term benefits.

Alexander M. Czopp and his colleagues provide a comprehensive meta-analysis of studies of positive stereotypes. Their meta-analysis suggests that positive self-stereotyping may have similar effects to what Galinsky and others found in regard to self-labeling. Czopp and others point out that positive self-stereotyping can act as "a compensatory coping strategy in response to the stigmatization associated with their group's negative stereotypes."[51] Self-definition with positive stereotypes may protect against negative stereotyping, and hence, defining oneself with positive stereotypes can help "to preserve a positive self-concept." Furthermore, the meta-analysis suggests that others' impressions of the targeted group are improved.[52] Given these two points, at least initially, it seems like positive stereotyping has similar effects to self-labeling. It not only insulates the group from negative stereotypes and helps to improve the group's self-image but also improves others' impressions of the group. However, Czopp and others go on to suggest that the phenomenon of positive stereotyping is more complex, and positive self-stereotyping often comes with a high price.

First, positive stereotyping can result in resentment because the targets feel that despite the flattery, positive stereotyping is depersonalizing. It is based on demographic markers rather than personal merits.[53] Even though Czopp and

49 Cepollaro, "The Moral Status of the Reclamation of Slurs."

50 Cepollaro, "The Moral Status of the Reclamation of Slurs," 672–88.

51 Czopp et al., "Positive Stereotypes Are Pervasive and Powerful," 453.

52 Czopp et al., "Positive Stereotypes Are Pervasive and Powerful," 454.

53 Czopp et al., "Positive Stereotypes Are Pervasive and Powerful," 456.

others do not discuss the issue of accuracy in relation to self-stereotyping, it seems to me that self-stereotyping, even positive self-stereotyping, is on shaky ground if one does not believe that the attribution of the stereotype is merited. Second, positive stereotypes often maintain a complementary relationship to negative stereotypes. To put it another way, positive stereotypes can exclude other positive stereotypes. For example, Asian people are stereotypically perceived as smart but cold. Conversely, when women are stereotyped as warm and friendly, they are perceived as less competent than their male counterparts.[54] Needless to say, stereotyping yourself as warm and friendly but less competent than your colleagues is not good for your career. Third, I started the discussion of self-definition by noting that the purpose of self-definition is to break oppressive norms concerning the target group. Ironically, Czopp and others suggest that positive stereotypes might possess even more normative force than negative ones. They argue that positive stereotypes can create expectations that are "more likely to encourage and reinforce stereotype-consistent behaviors than … negative stereotypes."[55] To illustrate, Czopp and others point out that people generally do not think that Black people *ought* to be uneducated and aggressive, even though that might be how Black people are stereotyped. In contrast, a Black person might be under the pressure to excel in sports, and an Asian person might feel the pressure to excel in math just because those are positive stereotypes.

As a consequence of these three points, Czopp and others conclude that positive stereotypes provide "a uniquely powerful mode of perpetuating inequality" and that positive stereotypes reinforce "hierarchies in which certain groups are consistently disadvantaged."[56] Taking a cue from Anderson, reclamation is intended to challenge the vehicles of harmful stereotypes, not to load the vehicles with other harmful stereotypes. The above discussion suggests that although self-stereotyping and self-labeling may have similar, initially positive effects, self-stereotyping might ultimately be more harmful than beneficial to the targeted group. Positive stereotypes, whether self-imposed or imposed by others, can be restrictive. If believed, they can restrict career and education opportunities, perhaps even more than negative stereotypes. If they are not believed, it is difficult to see how they can help the self-representation of the targeted group.

Doubt about the accuracy of stereotypes raises another separate but equally important question: Can there be accurate stereotypes? Discussion of this question reveals an important point about what is challenged in reclamation

54 Czopp et al., "Positive Stereotypes Are Pervasive and Powerful," 456; and Czopp, "When Is a Compliment Not a Compliment?" 414.

55 Czopp et al., "Positive Stereotypes Are Pervasive and Powerful," 456.

56 Czopp et al., "Positive Stereotypes Are Pervasive and Powerful," 457.

according to the epistemic view. One of the most convincing arguments against stereotyping derives from the unwarranted nature of the stereotypes, which seems closely connected to depersonalization. According to the epistemic argument, stereotypes are unwarranted because they do not take into account the personal traits of a target. Rather, some (stereotypical) traits are inferred solely on the basis of skin color or some other demographic marker that connects the targeted individual with a group. However, Erin Beeghly questions the objectionability of these kinds of inferences. She asks whether there are any good epistemic arguments for the claim that stereotyping is always objectionable. She does not find existing arguments convincing and concludes, "Stereotyping could be … epistemically permissible in some cases."[57] Similarly, Uwe Peters distinguishes between *neutral* stereotypes and *loaded* stereotypes: a loaded stereotype can be "an epistemic injustice" to the target; neutral stereotyping, on the other hand, is "a tendency to treat certain noticeable markers of social identity automatically as accurate predictors of certain beliefs."[58] Admittedly, I sympathize with the idea that stereotypes are *always* unwarranted. However, I do not want my view to be a hostage to the possibility of neutral stereotypes—that is, to the possibility that some stereotypes are warranted. Hence, it needs to be emphasized how exactly the challenge works according to the proposed view.

The epistemic view holds that the objectionability of slurs stems from the meaning of slurs, not from the epistemology of stereotypes. Slurs are objectionable because it is the meaning of slurs, not actual evidence, that allows one to attribute cruelty to Germans and vulgarity to French people. Even though it is a fine distinction, there is a difference. Even if the evidence for a stereotypical conception, *S*, makes *S* warranted, the meaning of a slur that attributes *S* to a group, *G*, is still objectionable because the attribution is based not on evidence but on meaning. Whether or not there are neutral or accurate stereotypes does not change the objectionability of slurs. The meaning of 'Boche' is objectionable because it allows the attribution of cruelty to Germans without any evidence. This affects reclamation, which first and foremost challenges the vehicle of the stereotype: the meaning of a slur.

The previous point can be elaborated with Dummett's criticism of classical logic. Initially, this might seem like a rather far-fetched analogy, but a closer inspection reveals that it too hinges on the notion of epistemic objectionability. Dummett criticizes classical logic on the grounds that the rules for classical negation are lacunose and therefore proof theoretically objectionable. Dummett himself adheres to intuitionistic logic. The crucial difference between classical

57 Beeghly, "What Is a Stereotype?" 675–91, especially 688.

58 Peters, "Hidden Figures," 33.

and intuitionistic logic is the divergent conceptions of negation. Intuitionistic and classical logic share the same introduction rule for negation, *reductio ad absurdum*, but the dispute is over different elimination rules. In intuitionistic logic, the elimination rule is just "A and not-$A \vdash \perp$"—which says that together A and not-A lead to absurdity. In classical logic, the negation elimination rule is double negation elimination rule, "not-not-$A \vdash A$." Dummett criticizes the classical rule because it smuggles in the idea of bivalence, which is not present in the introduction rule. So the classical rule, according to Dummett, is objectionable because it does not preserve warrants. Rather, the classical negation relies on a tacit additional assumption of bivalence. To use Rumfitt's phrase again, the elimination rule unpacks more than the introduction rule packs in. From this, the dispute usually expands to the nature of truth. For the proponents of classical logic, truth is evidence-transcendent, and the classical rule for negation is saved with this realistic notion of truth. If realism is right about the bivalent nature of truth, then the classical conception of negation is vindicated. Nevertheless, Dummettian intuitionists disagree. It is true that if there is a guarantee of decidability (that every sentence or its negation in the discourse is effectively provable), then all bivalent rules are justified even within intutionistic logic, but that does not yet vindicate the classical conception of negation. Rather, the justification derives first from the demonstration of decidability to bivalence, then from bivalence to the law of excluded middle; and finally, the law of excluded middle warrants the use of the double elimination rule in inferences. However, that does not change the objectionability of the classical negation. It is still lacunose. Hence, for an intuitionist, the idea that negation involves bivalence even before we know that the discourse in question is decidable is unwarranted.[59]

I argue that in reclamation, the target group takes the intuitionistic attitude and challenges the meaning of slurs on similar grounds. Even if there are neutral and accurate stereotypes, that does not change the objectionability of slurs. The meaning of 'Boche' is objectionable because it allows the attribution of cruelty to Germans regardless of the evidence concerning the cruelty of German people.

Needless to say, there is a caveat concerning the analogy between the attitudes of a logical intuitionist and someone who reclaims a slur. The idea that people engage in reclamation because they know that the meaning of a slur is proof theoretically objectionable seems unrealistic. But I do not think they need to be aware of the proof theoretic badness. As it is already covered, mere self- labeling is beneficial for the targeted group. Furthermore, while reclaimers may not be aware of the proof theoretical badness, they are often aware that

59 E.g., Tennant, *Taming of the True*, 175–76. My exposition here relies on Neil Tennant's discussion of the dispute mainly because of the clarity of his discussion but Dummett himself expresses very similar thoughts in "'Yes', 'No' and 'Can't Say'," 289–95.

reclamation is a way to challenge slurs as vehicles of stereotypical conceptions. For example, a body positivity activist, Raisa Omaheimo, was asked about her use of the term 'fat'. Her answer perfectly captures the connection between reclamation and the epistemic view:

> For many years, it has been my project to reclaim this word 'fat'. Traditionally, it has been associated with [an] awful lot of negative things. 'Fat' is stupid and lazy and … has no self-control, but I want to empty the word from these meanings. I think that 'fat' means that there is more fat in a [particular] body than in the average body, period. It describes a certain type of body, and it does not mean anything more. It has been very liberating to reclaim the term and take it away from the offensive usage.[60]

Omaheimo emphasizes that reclamation is about the meaning of slurs. More specifically, the aim of reclamation is to empty slur words of the negative and harmful stereotypes and thereby to neutralize the words. From the point of view of epistemic objectionability, when a word is neutralized, it is no longer objectionable. If the word 'fat' no longer inferentially takes you from a certain body type to stupidity or laziness, then it is no longer lacunose or objectionable. To illustrate, let us take Omaheimo's view of the meaning of 'fat' and formulate the inferential rules for 'fat':

> Fat-I(Gen): *x* has more body fat than the average body; therefore, *x* is fat.
>
> Fat-E(Gen): *x* is fat; therefore, *x* has more body fat than the average body, and typically, fat people are stupid, lazy, etc.

Given Omaheimo's recipe for reclamation, in reclamation, the rules for 'fat' are transformed into

> Fat-I(Gen): *x* has more body fat than the average body; therefore, *x* is fat.
>
> Fat-E(Gen)*: *x* is fat; therefore, *x* has more body fat than the average body.

As a consequence, the elimination rule no longer unpacks more than the introduction rule packs in. It just takes you back to where you started. The meaning of 'fat' is no longer objectionable. And to repeat the earlier point, the switch from negative stereotype to positive stereotype would be equally objectionable because the elimination rule would still unpack more than the introduction

60 Omaheimo talked about her new book *Ratkaisuja Läskeille* (*Solutions to Fats*, my translation) on the Finnish daytime show *Puoli Seitsemän* (*Half Past Six*, on Finnish broadcasting channel Yle TV1) on November 17, 2022. The interview was in Finnish; I have translated the section quoted here. There is also an article based on the interview (https://yle.fi/a/74-20005131), which contains the relevant passage in a slightly edited form.

rule packs in. The goal of reclamation is to *neutralize* slur words—to empty them of their lacunose content. This is what the epistemic view takes Jeshion's weapons control to mean. At the same time, reclamation contributes to self-definition. Being fat does not mean that one is stupid or lazy. 'Fat' just describes a certain type of body and nothing more.

3. CONCLUSION

Taking a cue from Jeshion's account of reclamation, there are two key aspects of reclamation: self-labeling and self-definition. As has been shown, the proposed epistemic view can account for both of these aspects. Although the epistemic view can accommodate only self-labeling, it has been shown that it offers a unique perspective on self-definition.

The epistemic view borrows expressive resources to explain self-labeling. This is possible because of the distinction between the meaning and the use of slurs and because of Dummett's idea that even with partial language knowledge, speakers can achieve successful communication. Notably, the epistemic view acknowledges that slurs can be used to vent hostility toward the target group without any knowledge of stereotypes associated with the target.

Nevertheless, the epistemic view holds that the expressive usage of slurs stems from the meaning of slurs, which includes stereotypical conceptions of the targets. These conceptions are harmful because they impose oppressive norms on the targets. According to the epistemic view, reclamation challenges the meaning of slurs that convey harmful stereotypes. By doing so, reclamation neutralizes slur words by transforming their meaning from derogatory to neutral.[61]

Tampere University
pablo.valtonen@gmail.com

61 I am very grateful to the members of the Research Council of Finland project Social and Cognitive Diversity in Science (Inkeri Koskinen, Jaakko Kuorikoski, Renne Pesonen, Samuli Reijula, and Kristina Rolin), who commented and vastly improved the manuscript. This work was also funded by the project. I also received funding from the Finnish Cultural Foundation and Kone Foundation. I thank the reviewers for their insightful comments. I would also like to extend my gratitude to the editors at *JESP* for making the effort to find good reviewers.

REFERENCES

Anderson, Elizabeth. *The Imperative of Integration*. Princeton University Press, 2010.

Anderson, Luvell. "Calling, Addressing, and Appropriation." In *Bad Words*, edited by David Sosa. Oxford University Press, 2018.

Anderson, Luvell, and Ernest Lepore. "Slurring Words." *Noûs* 47, no. 1 (2013): 25–48.

Ashwell, Lauren. "Gendered Slurs." *Social Theory and Practice* 42, no. 2 (2016): 228–39.

Beeghly, Erin. "What Is a Stereotype? What Is Stereotyping?" *Hypatia* 30, no. 4 (2015): 675–91.

Belnap, Nuel, D. "Tonk, Plonk and Plink." *Analysis* 22, no. 6 (1962): 130–34.

Berkovski, Sandy, Y. "Slurs, Synonymy, and Taboo." *Australasian Journal of Philosophy* 101, no. 2 (2022): 423–39.

Bianchi, Claudia. "Slurs and Appropriation: An Echoic Account." *Journal of Pragmatics* 66 (2014): 35–44.

Brandom, Robert. *Articulating Reasons*. Harvard University Press, 2000.

Cepollaro, Bianca. "Let's Not Worry About the Reclamation Worry." *Croatian Journal of Philosophy* 17, no. 2 (2017): 181–93.

———. "The Moral Status of the Reclamation of Slurs." *Organon F* 28, no. 3 (2021): 672–88.

Czopp, Alexander M. "When Is a Compliment Not a Compliment? Evaluating Expressions of Positive Stereotypes." *Journal of Experimental Social Psychology* 44, no. 2 (2008): 413–42.

Czopp, Alexander M., Aaron C. Kay, and Sapna Cheryan. "Positive Stereotypes Are Pervasive and Powerful." *Perspectives on Psychological Science* 10, no. 4 (2015): 451–63.

Davis, Christopher, and Elin McCready. "The Instability of Slurs." *Grazer Philosophische Studien* 97, no. 1 (2020): 63–85.

Diaz Legaspe, Justina. "Normalizing Slurs and Out-Group Slurs: The Case of Referential Restriction." *Analytic Philosophy* 59, no. 2 (2018): 234–55.

———. "What Is a Slur?" *Philosophical Studies* 177, no. 5 (2020): 1399–422.

Dickie, Imogen. "Negation, Anti-Realism, and the Denial Defence." *Philosophical Studies* 150, no. 2 (2010): 161–85.

Dummett, Michael. *Frege: Philosophy of Language*. Harper and Row, 1978.

———. *The Logical Basis of Metaphysics*. Harvard University Press, 1991.

———. "'Yes', 'No' and 'Can't Say'." *Mind* 111, no. 442 (2002): 289–95.

Galinsky, Adam, Cynthia S. Wang, Jennifer A. Whitson, Eric M. Anicich, Kurt Hugenberg, and Galen V. Bodenhausen. "The Reappropriation of Stigma-

tizing Labels: The Reciprocal Relationship Between Power and Self-Labeling." *Psychological Science* 24, no. 10 (2013): 2020–29.
Herbert, Cassie. "Precarious Projects: The Performative Structure of Reclamation." *Language Sciences* 52 (2015): 131–38.
Hess, Leopold. "Inferentialist Semantics for Lexicalized Social Meanings." *Synthese* 200, no. 5 (2022). https://doi.org/10.1007/s11229-022-03817-5.
Hom, Christopher, and Robert May. "Moral and Semantic Innocence." *Analytic Philosophy* 54, no. 3 (2013): 293–313.
Jeshion, Robin. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27 (2013): 307–35.
———. "Pride and Prejudiced: On the Reclamation of Slurs." *Grazer Philosophische Studien* 97, no. 1 (2020): 106–37.
———. "Slurs and Stereotypes." *Analytic Philosophy* 54, no. 3 (2013): 314–29.
Kukla, Quill, and Mark Lance. "Telling Gender: The Pragmatics and Ethics of Gender Ascriptions." *Ergo* 9, no. 42 (2022): 1130–59.
Leslie, Sarah-Jane. "Generics and the Structure of the Mind." *Philosophical Perspectives* 21, no. 1 (2007): 375–403.
———. "Generics: Cognition and Acquisition." *Philosophical Review* 117, no. 1 (2008): 1–47.
———. "The Original Sin of Cognition: Fear, Prejudice, and Generalization." *Journal of Philosophy* 114, no. 8 (2017): 393–421.
McCready, Elin. "Varieties of Conventional Implicature." *Semantics and Pragmatics* 3 (2010): 1–57.
Mühlebach, Deborah. "A Non-Ideal Approach to Slurs." *Synthese* 202, no. 3 (2023). https://doi.org/10.1007/s11229-023-04315-y.
Neufeld, Eleonore. "An Essentialist Theory of the Meaning of Slurs." *Philosophers' Imprint* 19, no. 35 (2019): 1–29.
Omaheimo, Raisa. *Ratkaisuja Läskeille*. Kustantamo S&S, 2022.
Peters, Uwe. "Hidden Figures: Epistemic Costs and Benefits of Detecting (Invisible) Diversity in Science." *European Journal of Philosophy of Science* 11, no. 1 (2021). https://link.springer.com/article/10.1007/s13194-021-00349-6.
Popa-Wyatt, Mihaela. "Reclamation: Taking Back Control of Words." *Grazer Philosophische Studien* 97, no. 1 (2020): 159–76.
Potts, Christopher. "The Expressive Dimension." *Theoretical Linguistics* 33, no. 2 (2007): 165–98.
Predelli, Stefano. "Unmentionables: Some Remarks on Taboo." *Organon F* 28, no. 3 (2021): 726–44.
Prior, A. N. "The Runabout Inference-Ticket." *Analysis* 21, no. 2 (1960): 38–39.
Ritchie, Katherine. "Social Identity, Indexicality, and the Appropriation of Slurs." *Croatian Journal of Philosophy* 17, no. 50 (2017): 155–80.

Rumfitt, Ian. "'Yes' and 'No'." *Mind* 109, no. 436 (2000): 781–824.
Tennant, Neil. *Taming of the True*. Oxford University Press, 2002.
Tirrell, Lynne. "Derogatory Terms: Racism, Sexism, and the Inferential Role Theory of Meaning." In *Language and Liberation*, edited by C. Hendrix and K. Oliver. State University of New York Press, 1999.
Valtonen, Pasi. "Gendered Slurs and the Subset Argument." *Journal of the American Philosophical Association* 8, no. 4 (2022): 762–79.
———. "Generic Inferential Rules for Slurs: Dummett and Williamson on Ethnic Pejoratives." *Synthese* 198, no. 7 (2021): 6533–51.
———. "Reclamation and Authorization: Cepollaro and Lopez de Sa on In-Group Restriction." *Philosophia* 52, no. 2 (2024): 463–73.
Whiting, Daniel. "Conservatives and Racists: Inferential Role Semantics and Pejoratives." *Philosophia* 36, no. 3 (2008): 375–88.
Williamson, Timothy. "Reference, Inference, and the Semantics of Pejoratives." In *The Philosophy of David Kaplan*, edited by Joseph Almog and Paolo Leonardi. Oxford University Press, 2009.

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · SEPTEMBER 2025
https://doi.org/10.26556/jesp.v30i5.4420

RESPONSIBILITY AND THE SOCIAL DIMENSION OF ADDICTION

Keyao Yang

If an addicted person commits wrongful acts in pursuit of drugs or alcohol, does her addiction mitigate or exempt her blameworthiness? This question has been a frequent test case for competing accounts of free and responsible agency.[1] However, as I show, the discussion so far focuses disproportionally on how addiction might undermine individual capacities. Although some recognize addiction's social dimension, in this article, I offer a more detailed defense of a view that takes this dimension and its impact on responsibility more seriously than previous accounts. I argue for a duress-like approach that incorporates the self-medication model of addiction, which is particularly well suited for evaluating the blameworthiness of addicted individuals who commit wrongdoings while living in adverse socioeconomic conditions.

My argument is structured as follows. In section 1, I describe the social dimension of addiction and present a real-world case (Kathy's case) to illustrate the demands this social dimension places on theories of moral responsibility, highlighting the theoretical gaps in current accounts. Sections 2 and 3 develop an account that fills these gaps. In section 2, I appeal to an account of responsibility that understands responsible action in terms of the fair opportunity to avoid wrongdoing. According to this view, there are two factors in determining the quality of an opportunity: the normative competence of an agent and her situational control.[2] Using this account as a map to locate and examine the main views concerning the responsibility of addicted agents, we can see that most common views focus on diminished normative competence. There is another approach that focuses on situational control and duress. However, the duress-oriented approach faces several challenges. In section 3, I bring Kathy's case into dialogue with the *self-medication hypothesis* of addiction, according

1 For some representative discussions on this topic, see Frankfurt, "Freedom of the Will and the Concept of a Person"; Watson, "Free Agency"; and Fischer and Ravizza, *Responsibility and Control*, 48.

2 See Brink and Nelkin, "Fairness and the Architecture of Responsibility"; Brink, *Fair Opportunity and Responsibility*; and Burdman, "A Pluralistic Account of Degrees of Control in Addiction."

to which addictive substance use is often initiated and sustained by a need to seek relief from painful living conditions. I argue that Kathy and arguably many others in similar conditions use addictive substances to alleviate the distress they endure from their social environments. Then I develop a duress-like framework of addiction to argue that Kathy's situation is duress-like. I further argue that this approach can be used to investigate the blameworthiness of other instances of addiction-related wrongdoing with pronounced socioeconomic dimensions. In section 4, I discuss the scope of the duress-like framework. Section 5 concludes the argument.

1. THE SOCIAL DIMENSION OF ADDICTION

The relation between addiction and adverse socioeconomic conditions is complex. On the one hand, not all addicted people live in poor socioeconomic conditions, and not all people who live in such conditions develop addictions.[3] This suggests that addiction and adverse socioeconomic conditions might not be inherently connected. On the other hand, empirical research from public health and social sciences consistently shows that a range of macrolevel social challenges are strong risk factors for the development and sustaining of addiction.[4] These challenges include poverty and unemployment; deficits in housing, education, and health care resources; inequality, marginalization, social injustice, political unrest; and violence in households and communities.[5] Populations that face multiple social challenges simultaneously are particularly vulnerable to addiction. Examples of these highly vulnerable groups include street-involved youth, sex workers, and displaced migrants and refugees, both in high-income countries with severe economic inequality and in low- and middle-income countries. The socioeconomic challenges affecting addicted populations certainly vary in severity. In addition to those who are most affected by socioeconomic hardships, there are populations who face fewer but still substantial challenges.[6]

Findings like these have led some researchers to conclude that social factors are a partial cause of addiction.[7] However, for many people, there is still insuf-

3 UNODC, *World Drug Report 2016*, 71.

4 UNODC, *World Drug Report 2016*, 72, 79.

5 UNODC, *World Drug Report 2016*, 69–71.

6 For empirical data that support these claims, as well as comprehensive summaries on the role of social challenges as risk factors for addiction on a global scale, see UNODC, *World Drug Report 2016*, 69–79; *World Drug Report 2018*, Booklet 4, 18–34, Booklet 5, 18–20; and *World Drug Report 2020*, Booklet 5.

7 Hart, *High Price* and "Viewing Addiction as a Brain Disease Promotes Social Injustice"; Alexander, *The Globalization of Addiction*.

ficient evidence to draw a definitive conclusion about the causal connection between addiction and socioeconomic conditions. To determine precisely whether and to what extent socioeconomic conditions contribute to addiction, much more research is needed. Currently, in many countries and regions with poor socioeconomic conditions, even basic demographic data on addicted populations are lacking, let alone research investigating the specific features and nature of addiction in these populations. Most studies still focus primarily on addicted populations in North America and Europe, which also receive the majority of research funding. Therefore, a conclusive understanding is unlikely to emerge in the near future.

Nevertheless, the available evidence is sufficient to show that within the global addicted population, a significant subpopulation experiences adverse socioeconomic conditions that contribute to both the onset and persistence of addiction. For compatibilist moral philosophers examining addiction and responsibility, although evidence of causal contribution does not automatically alter judgments of responsibility, it gives us a reason to consider the socioeconomic dimensions of addiction and to refine our theories accordingly. Specifically, we should develop frameworks that can accurately assess whether, to what extent, and in what ways these socioeconomic factors might mitigate or exempt a person's responsibility and blameworthiness for addiction-related wrongdoing.[8] While this subpopulation may not constitute the majority of addicted people (at least according to current available statistics), it is still a significant group with a high likelihood of becoming involved in various forms of wrongdoing. Therefore, evaluating moral responsibility and addiction in cases of addiction-related wrongdoing is particularly relevant to them. Moreover, many of these people already suffer from oppression and marginalization. Ensuring that our theories provide accurate assessments of their moral responsibility for addiction-related wrongdoing is the least we as moral philosophers of addiction can do for them.

However, current discussions of addiction and responsibility predominantly focus on the physiological or psychological characteristics of addiction. The most common reasons offered for mitigating an addicted person's blameworthiness involve addiction's impact on her normative capacity, including the way in which substance impairs her cognitive capacity to judge the morality of her actions and the way in which substance creates abnormally strong

8 To clarify, this article focuses on the moral responsibility (particularly the blameworthiness) of addicted individuals for wrongdoing committed to acquire and use substances. It does not cover their legal responsibility, wrongful actions under intoxication, or the legality of drug taking. These are separate questions.

desires that impacts the person's volitional capacity to control her drug-seeking behavior.[9]

There are some who discuss the impact of social conditions on addiction. Jeanette Kennett and colleagues, for example, offer a self-control theory of addiction.[10] On their view, an addicted person's psychological state and social environment collectively influence her self-control, thereby impacting her responsibility. Federico Burdman offers a reasons-responsiveness account to explain addicted people's degree of control and argues that addicted people's situational factors can undermine their degree of control.[11] Both Kennett and colleagues and Burdman acknowledge that social conditions can affect an addicted person's control and responsibility. However, they mainly focus on showing that social conditions are among the factors that affect the addicted person's responsibility. They do not offer a systematic and precise account of when, how, and to what extent these influences occur.

The scarcity of theoretical resources addressing the social dimension of addiction can make it difficult to analyze real-world cases of addiction and responsibility. Let us take a real-life case for example and consider what the current accounts can say about it: the story of Kathy, as told by Tanya Telfair Sharpe.[12] Kathy was a thirty-six-year-old Black woman living in an inner-city neighborhood in Atlanta, Georgia, receiving a monthly welfare check of $253. At the time she was interviewed, she was addicted to crack cocaine and was using it regularly. Her mother died from alcohol abuse when Kathy was twelve years old, and Kathy was the one to find her mother's body. Kathy soon dropped out of school to help care for her siblings. Since childhood, she had struggled with poverty. Economic decline in the inner city limited the jobs available to her. Her dropout status and poor reading skills further limited her employment opportunities, confining her to low-skill, low-paying jobs despite her continuous job search efforts. To secure funds for her everyday needs, Kathy worked as a prostitute. She married an abusive man, who introduced her to crack cocaine and repeatedly raped her. Although they eventually separated, his abuse continued. When crack cocaine was introduced into Kathy's neighborhood, drug

9 For discussion about addiction's impact on a person's cognitive capacity, see Morse, "Hooked on Hype"; Levy, "Addiction, Responsibility and Ego-depletion"; and Yaffe, "Lowering the Bar for Addicts." For discussion about addiction's impact on volitional capacity, see Wallace, "Addiction as Defect of the Will"; and Schroeder and Arpaly, "Addiction and Blameworthiness."

10 See Kennett, "Just Say No?"; Kennett et al., "Drug Addiction and Criminal Responsibility"; and Kennett et al., "Self-Control and Moral Security."

11 Burdman, "A Pluralistic Account of Degrees of Control in Addiction."

12 Sharpe, *Behind the Eight Ball*.

dealing and street-corner violence and crime increased drastically, which made her residential environment more dangerous than before. Kathy's work as a prostitute also exposed her to various safety hazards, including emotional and physical abuse.

Kathy's stressful and painful life is common among women in impoverished neighborhoods. Studies show that many women in such areas find only very limited employment opportunities beyond sex work despite their efforts to seek jobs and attend job-training programs.[13] These women frequently face violence and abuse, including police harassment, forced sex, robbery, rape, and murder.[14] Additionally, they struggle with homelessness, housing instability, domestic violence, and difficulties in accessing resources and services.[15]

At the time of the interview, Kathy had six children and reported severely neglecting them to procure and consume crack cocaine.[16] All but one had been removed from her custody. Why would she use drugs to the point of neglecting her children? Kathy's interviewers report that for women like Kathy, crack is often used to numb the feelings of loneliness and inadequacy felt by being poverty-stricken in a wealthy country.[17] Kathy explained her use as follows:

> The hardest part about [crack cocaine] is that you want to keep getting high, thinking that's going to ease the pain, and it helps a little bit. But once it's down, and speaking for myself, that's why I keep constantly wanting to get high, because the problems still be there—a lot of things that I didn't want to deal with that was happening in my life because of that. So I steadily kept smoking to ease the pain.[18]

13 Sterk, *Fast Lives*, 48–51.

14 See Maher, *Sexed Work*; Farley and Barkan, "Prostitution, Violence, and Posttraumatic Stress Disorder"; Sterk, *Fast Lives*, 72–73, 180–86; Valera et al., "Perceived Health Needs of Inner-City Street Prostitutes"; Raphael and Shapiro, "Sisters Speak Out"; Surratt et al., "Sex Work and Drug Use in a Subculture of Violence."

15 For research about female drug-using sex workers and housing instability, see Knight, *Addicted, Pregnant, Poor*, ch. 1. For domestic violence, see Sethi et al., "Experience of Domestic Violence by Women Attending an Inner City Accident and Emergency Department." For difficulties in accessing resources and services, see Kurtz et al., "Barriers to Health and Social Services for Street-Based Sex Workers."

16 In the interview, Kathy explained about her child neglect: "I got to the point where I didn't want to [care for my children] no more. Those things weren't important to me anymore. The only thing I cared about was getting that next high. . . . Cooking dinner for them was not important. Taking their baths was not important. When they got sick, somebody else had to take them [to the doctor], because it wasn't important. Crack was important" (Sharpe, *Behind the Eight Ball*, 172).

17 Sharpe, *Behind the Eight Ball*, 153.

18 Sharpe, *Behind the Eight Ball*, 151.

Her testimony suggests that her use of crack cocaine was partly for easing the overwhelming pain caused by the "deeper problems" in her life.[19] Some empirical studies of women who are in similar situations to Kathy's indirectly support this reading of the testimony. For example, studies show that women in impoverished neighborhoods who engage in prostitution use drugs and alcohol to cope with the negative emotions caused by prostitution.[20]

Many, especially those who view social factors as significant influencers of addiction, may intuitively feel that Kathy's blameworthiness for neglecting her children's care, such as failing to bathe or cook for them, is lessened by her substantial socioeconomic pressures. This intuition extends beyond mere acknowledgement of Kathy's impaired cognitive and volitional capacities to judge and control drug use or a simple sympathy for her difficult circumstances.[21] Rather, it arises from understanding that she used drugs for easing the distress caused by her overwhelming environmental difficulties, which she had limited power to change. However, due to a lack of adequate theoretical

19 Kathy expressed a belief that her deeper problems were consequences of her crack use, thereby assuming responsibility for causing them. However, many issues like poverty, an abusive relationship, occupational violence, and health concerns were not solely due to her drug use. It is likely that Kathy and her family overattributed responsibility to her to avoid feeling helpless. Moreover, Kathy's marginalized status might have subjected her to hermeneutic injustice, leaving her without the language to interpret her actions more charitably. Consequently, she viewed her drug use as a personal failure that was responsible for her problems. For further discussion on how marginalized people's actions might be uncharitably interpreted due to societal norms and power structures, see McKenna, "Power, Social Inequities, and the Conversational Theory of Moral Responsibility"; and Mackenzie, "Moral Responsibility and the Social Dynamics of Power and Oppression."

20 Women engaging in prostitution use drugs for various reasons: to detach from customers, perform their job, manage fear and feelings of worthlessness, and cope with degrading experiences. See Young et al., "Prostitution, Drug Use, and Coping with Psychological Distress"; Kramer, "Emotional Experiences of Performing Prostitution"; Bungay et al., "Women's Health and Use of Crack Cocaine in Context"; Sallmann, "Going Hand-in-Hand"; Bachman et al., "The Recursive Relationship Between Substance Abuse, Prostitution, and Incarceration"; and Daniulaityte and Carlson, "To Numb Out and Start to Feel Nothing." Additionally, they use drugs to alleviate physical pain when health care resources are unavailable or unaffordable, or when physicians hesitate to prescribe them pain medication. See Surratt et al., "Prescription Opioid Abuse Among Drug-Involved Street-Based Sex Workers"; and Bungay et al., "Women's Health and Use of Crack Cocaine in Context."

21 It certainly is true that our sympathy for a wrongdoer with a difficult life can conflict with our inclination to blame her. See Agule, "Being Sympathetic to Bad-History Wrongdoers." While this may explain our hesitation to blame, it does not explain why she is less blameworthy. Besides, sympathy alone does not capture all the subtleties of moral responsibility in Kathy's case. For sympathy to conflict with blame, wrongdoing does not necessarily need to be connected to a wrongdoer's difficult history. However, in Kathy's case, her difficult life was directly linked to her addiction-related wrongdoing.

resources on the social dimensions of addiction and responsibility, it is difficult to comprehensively explain why and to what extent socioeconomic pressures should mitigate Kathy's blameworthiness.

In this article, I hope to fill this theoretical lacuna. I articulate and defend a theoretical framework for discussing addiction's social dimension and moral responsibility. This framework rests on a duress-like account for excusing addiction-related infractions. It is particularly suitable for examining cases like Kathy's, where substance use is prompted and sustained by a felt need to cope with complex environmental pressures shaped by conditions of systemic poverty and violence. I demonstrate how it helps explain the intuition that Kathy's socioeconomic pressures mitigate her blameworthiness. Interestingly, even those who do not share this intuition can find my framework useful by providing a basis to explain why Kathy's socioeconomic pressures were not sufficient for mitigating her blameworthiness.

2. RESPONSIBILITY AND ADDICTION

Attending to a segment of the literature on moral responsibility can help highlight some of the tools on offer for supporting an excuse for addicted offenders like Kathy. Many theories of responsibility address conditions of excuse for moral wrongs. In this article, I focus on one in particular, namely David Brink and Dana Nelkin's "Fair Opportunity to Avoid Wrongdoing," as a framework for evaluating Kathy's case.[22] According to the fair opportunity account of responsibility, when an agent commits a wrong, whether she is responsible for it depends on whether she has a fair opportunity to respond to the practical reasons to resist the wrong. Whether she has such a fair opportunity is determined by two factors: normative competence and situational control. The latter factor, which is often neglected in discussions of moral responsibility, makes it a particularly useful tool for analyzing addiction's social dimension.

2.1. Normative Competence and Situational Control

Having normative competence is a necessary condition for a person to be held responsible. Normative competence consists of two capacities: cognitive and volitional. Cognitive capacity refers to the agent's ability to recognize practical reasons, which in the context of moral responsibility, are moral reasons and norms; that is, the agent can tell what counts as wrongdoing and what does not. Volitional capacity refers to the agent's ability to form the intention to

22 Brink and Nelkin, "Fairness and the Architecture of Responsibility"; and Brink, *Fair Opportunity and Responsibility*.

act in accordance with the practical reasons she recognizes. Suppose a person recognizes that she has a reason to do something because it is morally right. If she has sufficient volitional capacity, then she possesses the capacity to form the intention to do the morally right thing. In specific cases, possessing volitional capacity not only means that the person can produce motivation to act upon the practical reasons she recognizes but also means that the motivation is strong enough to defeat other motivations and temptations that compete with and distract her from the practical reasons.

Normative competence comes in degrees. People in different stages and moments of life have different degrees of normative competence. Moral responsibility comes in degrees too. One factor that affects a person's blameworthiness is her degree of difficulty to avoid the wrong.[23] Thus, a person's level of normative competence can affect her blameworthiness by affecting how difficult it is for her to avoid the wrong. If the person has low normative competence, she might have more difficulty telling right from wrong or resisting the wrong. If she makes a wrongful choice on an issue that needs high normative competence to make the right choice, her blameworthiness might be mitigated.

Having normative competence makes a person a responsible agent, which is necessary but not sufficient for her to be responsible for an action. When a person commits a wrongdoing, she must also possess situational control to be responsible for it. While normative competence is about an agent forming an intention, situational control is about the quality of the environment outside of the agent. When a person has full situational control, it means her external circumstances give her (or any person in her position with normative competence) a fair opportunity to execute the intention she forms. In other words, the external circumstances do not stand in her way to the extent that deprives her of a fair opportunity of carrying out her intended activities.

In most cases, when people commit wrongdoing, few external factors stand in their way of doing the right thing. Thus, they have full situational control and are responsible for their wrongdoing. However, sometimes people do not have full situational control. Making choices under the threat of another person or natural force are typical examples of lacking situational control. When a person commits a wrong under such a threat, she does not have a fair opportunity to choose the right action. Therefore, her situational control (or a significant part of it) is taken away, which excuses her from being held fully responsible for the wrongdoing.

Like normative competence, situational control comes in degrees, which can affect the degree of moral responsibility by affecting the degree of difficulty in avoiding the wrong. One view ties this difficulty to effort and sacrifice: the

23 Nelkin, "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness."

greater the effort and sacrifice a person has to make to avoid wrongdoing, the more difficult it is for her to avoid it, and the less blameworthy she is for committing the wrong.[24] For example, a person who ignores a stranger in need of an ambulance in order to rush to a grocery store sale retains full situational control and a fair opportunity to help. By contrast, if the person ignores the stranger because she is rushing to the hospital to see a dying loved one, then the sacrifice she faces is much greater than missing a store sale. It hence diminishes her situational control and reduces her responsibility for not helping.

In summary, a person's situational control, understood in terms of the sacrifice and effort she has to make to avoid committing the wrong, affects her blameworthiness. If the situation poses severe enough pressures, and the wrongdoer needs to make a big effort or sacrifice to avoid the wrong, then she might be partially or fully excused from blame.

2.2. Addiction and Diminished Normative Competence

Most attempts in the literature to excuse addiction-related offenses appeal to diminished normative competence: an addicted person's cognitive and/or volitional capacities are impaired in such a way that she cannot properly recognize and respond to reasons. For example, Neil Levy argues that addiction impairs the agent's ability to judge what is valuable to her.[25] Long-term drug use changes the person's brain, making her more likely to misjudge a drug's value and to be convinced by weak reasons to use drugs through processes such as ego depletion and belief oscillation. Gideon Yaffe argues that an addicted person's drug use indicates a momentary change in judgment about the drug.[26] When the person uses, she sees drugs as valuable; thus, addiction is connected to one's beliefs and judgments.

Other accounts connect addiction to abnormal desires. For instance, Timothy Schroeder and Nomy Arpaly argue that addictive drugs affect the brain's predictive system, and addiction is partly an obstinate habit or unconscious disposition, which is closer to volition (desire) than to cognition (judgment).[27] Richard Holton and Kent Berridge argue that addiction is an abnormal feeling of wanting something caused by incentive sensitization despite disliking it.[28] On their view, this feeling is cultivated through the drug's impact on the brain's

24 Nelkin, "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness."

25 Levy, "Addiction, Responsibility and Ego-Depletion," "Addiction as a Disorder of Belief," and "Addiction."

26 Yaffe, "Lowering the Bar for Addicts," "Are Addicts Akratic?" and "Compromised Addicts."

27 Schroeder and Arpaly, "Addiction and Blameworthiness."

28 Holton and Berridge, "Addiction Between Compulsion and Choice" and "Compulsion and Choice in Addiction."

desire system that tracks what is rewarding. Some philosophers also suggest that addiction might function as a strong and highly recurrent noise to the addicted person, which distracts the person and competes for her attention.[29] These theories suggest that an addicted individual's volitional capacity to avoid drug use is also impaired.

Within the structure of the fair opportunity theory of responsibility, we can see that these theories support the argument that an addicted person's blameworthiness should be mitigated due to impairments of normative competence. Whether mitigation of blameworthiness is partial or full depends on a philosopher's background view about how much impairment of normative competence is needed to provide a full excuse. For many philosophers, the empirical evidence shows that an addicted person's desire to use drugs is very strong, but it does not show that it is irresistible. Therefore, their blameworthiness can be mitigated but not fully excused.[30]

The literature so far makes a clear connection between addiction and the impairment of a person's normative competence and establishes the reason to partially excuse addicted wrongdoers for impaired normative competence. It allows us to argue that Kathy was at least partially excused when she was busy with getting crack cocaine and left her children aside. But I want to explore the ways to argue that, besides the impairments caused by the drug, Kathy was also (at least partially) excused owing to her difficult socioeconomic environment.

It is possible that Kathy's harsh social conditions caused or worsened mental illnesses like depression and post-traumatic stress disorder, impairing her reasons-responsiveness and potentially excusing her actions due to diminished normative competence. However, I want also to defend the intuition that, aside from causing her mental illness or diminishing her normative competence, Kathy's environment also directly contributed to excusing her wrongdoing. Recall that the Fair Opportunity to Avoid Wrongdoing view recognizes that a lack of situational control, alongside diminished normative competence, can also excuse behavior. In the next section, I explore whether a situational-control-based excuse can mitigate Kathy's blameworthiness.

2.3. Situational Control, Duress, and Addiction

When a person commits a wrong, if her situational control is compromised, she may be excused from full responsibility on the grounds of lacking a fair

29 See, e.g., Watson, "Disordered Appetites," 10–11; and Sripada, "Addiction and Fallibility."

30 See Morse, "Hooked on Hype" and "Addiction, Choice and Criminal Law"; Wallace, "Addiction as Defect of the Will"; Schroder and Arpaly, "Addiction and Blameworthiness," 235–36; and Pickard, "Psychopathology and the Ability to Do Otherwise" and "Responsibility Without Blame for Addiction."

opportunity to avoid the wrongdoing. *Duress* is a typical case of lacking situational control. While duress excuses are more familiar in legal contexts, we can apply the same reasonings to think about moral offenses. Duress has been explored as a possible excuse for addiction-related offenses. However, this approach faces several challenges.

In jurisprudence, duress refers to a situation in which a person is threatened such that unless she engages in wrongdoing, she will face death or grievous injury. For example, a mother who is forced at gunpoint by a terrorist to drive a getaway car to save her child is in a situation of duress. The mother is forced to choose between two options: helping a terrorist or risking her child's safety. Duress removes the mother's opportunity to execute her intention to do neither, leaving her with no choice but to commit a wrongdoing or suffer harm. Although the mother has sufficient normative competence, she is not fully responsible for helping the terrorist due to this external situation. Punishing her would be unfair because for any person of reasonable firmness, if they were in her situation, it would be too much to expect the person to make such a big sacrifice and resist committing wrong.[31]

To count a situation as one involving duress, several conditions must be met: (1) the consequence of not committing the wrongdoing should be imminent; (2) there is no reasonable escape from choosing among the limited options; (3) the agent reasonably believes that the hard choice is genuine; (4) the consequence of the choices under the duress is severe enough;[32] and (5) the agent is not responsible for putting herself under duress.[33] Failing to meet any of the conditions makes the situational excuse inapplicable to the wrongdoer.[34]

31 Dressler, "Duress," 284–87.

32 In the legal context, a severe enough consequence usually means death or grievous injury to oneself or a loved one.

33 Dressler, *Understanding Criminal Law*, 297–98; and Brink, *Fair Opportunity and Responsibility*, sec. 63.

34 These conditions are listed as criteria for duress not just for legal considerations but also because they align with common intuitions about what takes away an agent's fair opportunity to avoid wrongdoing and thus constitutes duress. Consider condition 5 as an example: suppose a person steals bread to feed her starving child. At first glance, we might view her actions as excusable under duress. However, if we later learn that her child is starving because she had previously chosen to invest aggressively, knowingly taking a financial gamble that jeopardized her child's basic security, her excuse becomes less compelling. In this case, she had a fair opportunity to avoid putting herself in this situation but voluntarily relinquished it. This illustrates why condition 5 is a reasonable criterion for duress. Similarly, the other conditions of duress align with our intuitions about what takes away a person's fair opportunity in comparable ways. While specific intuitions may vary for conditions like 4 (about, for instance, how severe is sufficiently severe) the overarching principles underlying all five conditions are broadly intuitive.

However, as several philosophers have argued, addiction faces difficulties in meeting some of these conditions. The first difficulty concerns the level of suffering required to qualify as being in duress. To count as duress, the person must face threats comparable to death, grievous injury, or the loss of something vitally important.[35] The most obvious suffering associated with abstaining from a substance is withdrawal, which may involve unsatisfied cravings, physical symptoms, dysphoria, and related effects. However, Stephen Morse argues that such suffering is often not sufficiently extreme.[36] We rarely treat the frustration of a desire as totally unbearable, and for at least some forms of addiction, physical withdrawal is not that unbearable. Several philosophers note that withdrawal from heroin, for example, is often likened to having the flu.[37] Other mental phenomena, such as intense cravings or the fear of dysphoria, may be significant and severe; but their severity is difficult to assess.[38] It remains an open question whether the consequences of refraining from drug use meet the threshold of severity required for duress. More empirical evidence is needed to support the duress model in light of this challenge.

The second challenge is that in some cases, an addicted person may bear responsibility for developing her dependence on the substances.[39] If the difficulty of abstaining results from her own earlier choice to begin using the drug, then condition 5 seems not satisfied, and the excuse of duress is less compelling.[40]

There are other disanalogies between addiction and typical duress cases. For instance, duress often involves an opposing individual who threatens an agent to act against her will, but addiction lacks such an outside agent. In addition, duress is typically a one-time event, while addiction is a long-term and repeated

35 In jurisprudence, only death and grievous bodily injury are considered severe threats. However, here we are discussing moral judgments rather than legal judgments. A person may be morally excusable even if not legally so. We can lower the bar to include significant but less severe threats.

36 Morse, "Hooked on Hype," 35–38.

37 Husak, "Addiction and Criminal Liability," 682; Morse, "Hooked on Hype," 36, and "Addiction and Criminal Responsibility," 183–88; and Pickard, "Psychopathology and the Ability to Do Otherwise."

38 Morse, "Hooked on Hype," 35–36.

39 Morse, "Hooked on Hype," 37, and "Addiction and Criminal Responsibility," 175–79; and Brink, *Fair Opportunity and Responsibility*, s. 114.

40 For some subgroups of the addicted population, such as those who acquire addictions in childhood, this challenge might not be a problem. For a discussion about the reasons to mitigate responsibility for developing addiction in some cases, see Kennett et al., "Drug Addiction and Criminal Responsibility."

problem that often leads to the commission of small wrongs.[41] Therefore, the standard legal excuses for duress do not readily apply to addiction.

Nevertheless, some philosophers take the possibility of a duress-based excuse for addiction seriously and explore conditions under which such an excuse or something analogous might apply. Gary Watson is one of the most prominent figures in this discussion. He argues that addiction is a form of acquired appetite, which can be understood by analogy to natural appetites such as hunger or thirst.[42] If a person commits a minor wrong to avoid the severe pain of hunger or thirst, we might judge that she is in a kind of duress: the pain is severe enough that it would be unreasonable to expect her to resist. Similarly, if an addicted person uses drug to avoid the suffering associated with deprivation of an acquired appetite, she too might be seen as acting under duress and therefore partially excused. This analogy helps illuminate why withdrawal might be seen as sufficiently severe. At the same time, as noted above, other philosophers draw on empirical evidence to emphasize variation among addictive substances, and not all forms of withdrawal appear severe enough to render resistance unreasonable. More discussion is needed to establish that addiction can satisfy condition 4.

Condition 5 is another major hurdle for applying the duress excuse to addiction. In response, Watson observes that failing to satisfy this condition does not automatically disqualify someone from invoking the duress excuse.[43] For example, people are typically seen as responsible for the emotional attachments that they form to their spouse and children. Yet if someone is coerced into wrongdoing under the threat of her child's life, the situation may still qualify as duress. More broadly, Watson argues that our judgments about whether someone is responsible for placing herself in a hard choice through her appetites or attachments are often affected by cultural, social, and political norms. These include norms governing the meaning and value of various appetites and attachments, the socially perceived difficulty of changing one's life to relinquish the appetite, and what is reasonable to demand of a person in such circumstances. (Social

41 Compared to the five central conditions of duress, these disanalogies are more peripheral and do not directly affect common intuitions about whether a wrongdoer is excusable.

42 See Watson, "Disordered Appetites," 11. In bringing out the analogy between natural appetite (such as hunger and thirst) and addicted appetite for substances, Watson mainly focuses on the commonalities between them. The commonalities include the subjective experience of intense desire and the distress of deprivation. However, not all features of natural appetites carry over. For example, the analogy does not include the objectively life-threatening consequences of food or water deprivation (such as malnutrition and dehydration) because withdrawal from many addictive substances is not fatal (although there are some exceptions).

43 Watson, "Excusing Addiction."

norms may also influence our assessments of condition 4 by shaping views about what kinds or degrees of harm count as sufficiently severe.)

Watson's discussion offers a fruitful starting point. It suggests that condition 5 may not straightforwardly disqualify addiction from counting as duress, thereby opening space for the possibility that addiction might sometimes qualify. However, his exploratory remarks do not yet provide a complete or practically applicable account of how and under what conditions addiction might meet the criteria for duress in real-world cases. Further work is needed to develop such an account. This may involve clarifying, examining, or revising relevant social norms, including addressing questions about their reasonableness and internal consistency. It might also require a more fine-grain taxonomy of addiction: if different subtypes of addiction correspond to distinct normative expectations, then some may plausibly qualify for a duress-based excuse.

Yaffe also examines addiction and responsibility through comparison with duress.[44] Rather than arguing that addiction constitutes literal duress, Yaffe argues for an analogous excuse that mirrors the underlying rationale of duress defense. In standard duress, a person is threatened with a severe consequence that will follow if she does not commit a wrongful act. The justification for the excuse is that the threatened harm is so serious that we cannot reasonably expect the person to resist wrongdoing and bear it. Yaffe argues that addiction involves a structurally similar situation. Normally, we expect people to act righteously, guided by their autonomy and understanding of moral rights and wrongs. However, due to the effects of addiction, particularly the way it compromises the dopamine system, addicted people face significant difficulties learning and responding to reasons to refrain from drug use.[45] For them, the only way to avoid drug use may be to relinquish their autonomy and significantly suspend their moral reasoning. This is not a case in which someone is threatened with a severe consequence for not using drugs. Rather, it is a case in which, in order to avoid drug use, the person must make a morally significant sacrifice, namely the sacrifice of autonomy. Like in duress, we cannot reasonably expect the addicted person to make such a sacrifice. Therefore, addicted people may not be fully responsible for failing to refrain from drug use.

Yaffe clarifies that this excuse applies primarily to the act of drug consumption in certain circumstances and its direct consequences and less so to wrongdoings committed as a means of acquiring drugs, such as theft to obtain money for purchase.[46] The reason is that the feature that undermines our expecta-

44 Yaffe, "Lowering the Bar for Addicts," "Are Addicts Akratic?" and "Compromised Addicts."

45 Yaffe, "Are Addicts Akratic?"

46 Yaffe, "Compromised Addicts."

tions of the addicted person—namely, that refraining from use would require a morally significant sacrifice of autonomy—is present at the point of drug consumption but not clearly present at the earlier stage of drug acquisition. When an addicted person commits a crime in order to obtain drugs, she has not yet encountered the burden of giving up autonomy in order to refrain from use. That burden arises only later when she must choose whether or not to consume the substance. Therefore, in cases involving preparatory wrongdoing such as theft, we can reasonably expect more from the agent, and the excuse is less appropriate.

What does the preceding discussion suggest about Kathy's case and her moral responsibility? First, it shows that mitigating Kathy's responsibility through a duress excuse would likely encounter the same challenges commonly faced by duress-based defenses of addiction, particularly in meeting conditions 4 and 5. Watson's analysis offers a potential avenue for addressing these conditions by examining the social norms surrounding addiction. Given this article's focus on Kathy's socioeconomic background, we might adopt Watson's approach and explore the norms governing how addiction is understood and morally evaluated in cases of socioeconomic hardship. This may be a productive line of inquiry, though further development is needed to determine whether and in what way Kathy's case might satisfy the criteria for a duress-based excuse. Yaffe's account offers insight into some aspects of Kathy's failure to fulfill her parental duties, particularly where those failures stem directly from drug use. However, it is less helpful in accounting for failures arising from drug acquisition and income-generating activities related to that end. Moreover, Yaffe's account is about addiction's impact on the brain and does not engage with the socioeconomic dimensions of addiction. Thus, if we are to explain the intuitive pull of the idea that Kathy's disadvantaged circumstances mitigate her responsibility for both drug use and drug-acquiring behaviors, we need to look for additional resources.

Besides Watson's and Yaffe's discussions surrounding addiction and duress, there are other accounts that discuss addicted people's situational control by considering their social environments. For example, Kennett and colleagues present a self-control theory of addiction, arguing that an addicted person's psychological state and social environment jointly affect her self-control, thereby impacting her responsibility.[47] Addiction's psychological and neurophysiological aspects, such as distorted beliefs and cravings, compromise the person's ability to control her decisions and desires, and hence, they impair her self-control. Adverse external conditions, including poverty, social exclusion,

47 Kennett, "Just Say No?"; Kennett et al., "Drug Addiction and Criminal Responsibility"; and Kennett et al., "Self-Control and Moral Security."

and trauma, can limit an individual's available life options and undermine her control. The interplay between psychological states and social circumstances can further diminish the addicted person's control over decisions and actions by degrading her sense of self-worth and self-efficacy, making her devaluate long-term goals in favor of immediate substance use.

Kennett and colleagues' account is valuable for recognizing and incorporating a broad range of factors that influence an addicted person's self-control, especially social factors, which have been underappreciated in earlier philosophical discussions on addiction and control. However, their account would benefit from finer distinctions. It is important not only to acknowledge the breadth of factors affecting self-control but also to differentiate their nature and moral significance. A key distinction lies between lacking self-control in a psychological sense and lacking self-control in a social sense. When a person lacks self-control in the psychological sense, it diminishes both her responsibility and her status as a moral agent, thereby justifying paternalistic intervention. By contrast, when a person's self-control is impaired in the social sense, it reduces her responsibility without compromising her moral agency or status. Recognizing these distinctions is crucial for appropriately addressing questions of blameworthiness and the ethics of intervention.

Burdman's account of addiction and control offers a more precise evaluation of self-control in addicted people.[48] Drawing on the reasons-responsiveness theory, Burdman argues that the loss of control over substance-related behaviors varies in degree. This degree of control is defined in terms of a person's degree of reasons-responsiveness, which is determined by psychological factors (such as cravings and anomalous drug-related beliefs) and situational factors (such as peer pressure, environmental cues, and difficulties in attaining community support and social identity). Burdman's approach provides richer theoretical resources for distinguishing the moral implications of psychologically and situationally impaired self-control. However, it could be further refined by specifying criteria for assessing which, how, and to what extent social factors influence control and responsibility in cases of addiction.

In the next section, I argue for an account of addiction and blameworthiness that focuses specifically on the social environments and situational control of addicted people. This approach provides more systematic and precise guidelines for assessing how social factors affect blameworthiness. Moreover, by separating normative competence and situational control, it enables us to evaluate addicted people's blameworthiness with a better appreciation of the moral consequences of our evaluation.

48 Burdman, "A Pluralistic Account of Degrees of Control in Addiction."

3. FOCUSING ON THE SOCIAL DIMENSION: SELF-MEDICATION AND DURESS-LIKE EXCUSES

The theories on addiction and responsibility reviewed in the last section do not appear to offer sufficient theoretical resources to fully account for the intuition that there is a direct link between Kathy's overwhelming environmental challenges, her drug use, and her blameworthiness. In this section, I present and defend an alternative framework.

Kathy's testimony suggests that her crack cocaine use was partially driven by a need to alleviate severe distress resulting from the underlying issues in her life. We can make sense of this claim through the *self-medication hypothesis* of addiction, which is often used by clinicians who employ the psychodynamic approach to treat addiction. According to this hypothesis, individuals with chronic, severe addiction consume substances to relieve distressing emotions, such as anxiety, sadness, and despair, analogous to using pain relief sprays for physical pain.[49] This hypothesis is compatible with seeing addictive substance use as highly compelled (in the sense that the desire is abnormally strong and very difficult to refuse) but resists seeing it solely as a compulsion. It therefore offers the possibility that addictive substance use can be both compelled to some extent and an intelligible choice made for a purpose to some extent. It offers a framework for addicted people to see potential for change in their behavior rather than viewing themselves as helpless victims of uncontrollable urges.

Hanna Pickard was the first philosopher to introduce the self-medication model into discussions of addicted people's responsibility and blameworthiness.[50] She uses the model to challenge the common assumption that addiction, as a compulsion, fully exonerates an addicted person from responsibility. Using the self-medication model, Pickard argues that substance use can be a choice made for purposes such as self-medicating distress or enhancing self-worth.[51] Therefore, addicted persons can still be responsible for their substance use and associated wrongdoings, although not blamed.[52] In her account, Pickard acknowledges that because addictive drug use is a choice, the context in which that choice is made (such as social disadvantage, insufficient social support, and limited material resources) can constrain the coping strategies and life options

49 Khantzian, "The Self-Medication Hypothesis of Addictive Disorders," "The Self-Medication Hypothesis of Substance Use Disorders," and "Understanding Addictive Vulnerability."

50 Pickard and Ward, "Responsibility Without Blame"; and Pickard, "Psychopathology and the Ability to Do Otherwise," "Responsibility Without Blame for Addiction," and "Addiction and the Self."

51 Pickard and Ward, "Responsibility Without Blame"; and Pickard, "Addiction and the Self."

52 Pickard, "Responsibility Without Blame for Addiction."

available to a person, thereby reducing (though not eliminating) her responsibility for choosing to use addictive substance. Sometimes, given the options available, an addicted person's choices and actions "may be justified by duress."[53]

Pickard's account offers valuable insights for understanding Kathy's case. However, while it acknowledges the impact of socioeconomic difficulties on responsibility, it does not provide a structured framework for systematically assessing which socioeconomic pressures are relevant for mitigating responsibility, nor to what extent they should do so. As a result, evaluators (such as clinical practitioners) must rely on their own discretion to determine which socioeconomic difficulties are relevant, whether they are severe enough to mitigate responsibility, and if so, to what degree. This lack of standardized criteria increases the risk of inaccurate judgments, potentially leading to the oversight or underestimation of certain socioeconomic factors' impact on responsibility. Furthermore, while Pickard recognizes that choices related to substance use may sometimes be justified by duress, her account does not explain how such justification can be established, especially given the challenges facing the duress excuse for addiction—namely, the difficulties in satisfying conditions 4 and 5, as discussed in section 2.3 above. Without addressing these challenges, the appeal to duress risks being seen as mere humanitarian sympathy for an addicted person's hardships rather than as a robust and objective basis for excuse with moral (or even legal) significance.

This is not to suggest that Pickard's account is incorrect. Rather, it highlights the need for further exploration of the relationship between addiction, socioeconomic factors, and duress. To address this gap and develop a more comprehensive framework for evaluating which socioeconomic factors are relevant to mitigating the responsibility of addicted individuals, I propose the following positive account, building on where Pickard's work leaves off. The self-medication model takes comorbid psychiatric illnesses to be the most common sources of distress that create the need to use addictive substances. However, empirical evidence also highlights social factors like systemic poverty and violence as significant sources of distress, creating the need to self-medicate among people in impoverished communities.[54] While both types of self-medication likely play a role in Kathy's and similar cases, the second type is my main focus here.

In the previous section, we noted the limitations of a duress excuse for addiction. However, if addictive substance use is a choice to self-medicate and

53 Pickard, "The Purpose in Chronic Addiction," 46.

54 See Young, Boyd, and Hubbell, "Prostitution, Drug Use, and Coping with Psychological Distress"; Bungay et al., "Women's Health and Use of Crack Cocaine in Context"; Sallmann, "Going Hand-in-Hand"; and Daniulaityte and Carlson, "To Numb Out and Start to Feel Nothing."

to cope with distress caused by environmental difficulties, it might constitute a choice under duress without facing those limitations. Consider Kathy's situation: in daily life, Kathy frequently faced a hard choice between using substances to self-medicate overwhelming distress and fulfilling moral obligations such as parental duties. She might have wanted to do both at the same time, but her situation removed her opportunity to do so. As a result, when she chose to self-medicate and failed to fulfill the obligations, she might not have been fully blameworthy. Blaming her would not be totally fair, because for any person of reasonable firmness, if they had been in her situation, it would have been too much to expect that person to resist self-medicating. Described in this way, Kathy's situation appears to highly resemble one of duress. There are certainly differences: typical legal duress usually involves coercion by another agent and occurs as a singular event.[55] Addiction does not involve another agent, and it tends to involve repeated episodes. Therefore, to be more precise, in the subsequent discussions, I call Kathy's situation *duress-like* rather than strict duress.[56]

The duress-like framework offers a new way of analyzing Kathy's case.[57] As noted earlier, the standard account of duress involves five conditions. Philosophers have said relatively little about conditions 1 and 3, which suggests that addicted people typically have little trouble satisfying them. Condition 2 may be more context sensitive: whether there is a reasonable means of escape often

55 One might be concerned that addiction involves no coercion by another agent and therefore resembles necessity more than duress. Given space constraints, I do not explore this question in depth here. Fortunately, duress and necessity share similar conditions, and both align with my main thesis by diminishing the agent's blameworthiness. Thus, my argument remains relevant regardless of which one is used. One reason to favor duress over necessity here is that duress typically provides an excuse, while necessity offers a justification that says the action is not wrong. Since it is intuitive that Kathy and others in similar situations wrong their children, duress seems the more appropriate comparison.

56 By 'duress-like', I mean that Kathy's situation does not match the technical or historical usage of 'duress' in legal contexts. Nonetheless, it shares the normative structure and underlying moral intuition of duress. I use the term 'duress-like' to signal these similarities while acknowledging the legal distinctions.

57 Some readers may question the persuasiveness of my argument, grounded in the self-medication model and choice theory of addiction, especially in light of competing theories such as Levy's belief oscillation theory ("Addiction as a Disorder of Belief" and "Addiction") and Holton and Berridge's incentive sensitization theory ("Addiction Between Compulsion and Choice" and "Compulsion and Choice in Addiction"). However, as extensive empirical research shows, addiction has a complex and multifaceted nature. No single theory claims to comprehensively explain all aspects of addiction; rather, multiple theories each illuminate some robust and undeniable aspects of the disorder. The self-medication model and choice theory are among them. Supported by empirical evidence, they underscore the significant roles of self-medication and choice in substance use. These aspects of addiction are compelling and relevant, even to proponents of alternative theories.

depends on the circumstances surrounding the case. By contrast, conditions 4 and 5 are widely recognized as persistent challenges in applying a duress excuse to addiction. While some promising suggestions have been offered, no systematic account has yet shown how addiction can satisfy both conditions. However, as I argue, Kathy's case (and others like it) shows that socioeconomic hardship can provide a basis for meeting conditions 2, 4, and 5.

First, the scarcity of resources in impoverished communities often constrains the viable alternatives to drug use that the residents in these communities have access to. These communities typically have very few treatment facilities, and residents may lack the means and time to attend treatment without significantly disrupting their daily lives.[58] Moreover, many facilities cannot accommodate residents' needs such as childcare while receiving treatment. Even when treatment is available, it is not always effective.[59] As a resident of an impoverished community, Kathy lacked sufficient viable alternatives to drug use. Therefore, her situation is likely to satisfy condition 2.

Second, Kathy reported resorting to crack cocaine as a coping mechanism for the stressors of living in her impoverished community. The stressors include the day-to-day struggle with extreme poverty, a bleak job market, as well as recurrent incidents of humiliation, abuse, and violence from her partner and customers. Plausibly, the distress, anxiety, and despair caused by enduring these harsh conditions without any assistance from the drug were severe and comparable to the loss of something of great significance in life. Therefore, Kathy's situation can potentially satisfy condition 4. Addicted people who do not face comparable socioeconomic difficulties, on the other hand, are not able to invoke this form of emotional distress to meet condition 4.

Third, if we understand Kathy's "threat" as her severe and distressing environmental pressures, it seems unjust to hold her solely responsible for it. Social problems such as the lack of access to health care, education, and housing, as well as the presence of street violence and domestic violence, are prevalent issues in impoverished areas. Kathy's personal ability to overcome these social problems is limited. One might argue that Kathy is still responsible for starting her drug use (she might have started it out of curiosity), for deciding to

58 Sterk's research indicates that many women in such communities urgently need affordable treatment programs, which sometimes have waitlists of more than six months (Sterk, *Fast Lives*). Such delays can change life circumstances, leaving an individual no longer ready for treatment when their turn arrives.

59 As shown in Sterk, women often view treatments as overly structured and restrictive, primarily aimed at detoxification, and inadequate for preparing them for real-life challenges (*Fast Lives*, 135–42). Aftercare and social support are scarce, leading many to relapse upon returning to their prior environment.

have children (which further compromised her financial security), and/or for choosing to remain in an abusive relationship and to work as a prostitute (which add to the violence that she endures.) However, given her lack of education, impoverished upbringing, and history of trauma and abuse, it may be unreasonable to hold her fully responsible for these choices.[60] Moreover, her living conditions at the time of reporting impeded mobility and limited her access to social support, which further reduced her ability to make different choices. Therefore, Kathy's situation is likely to satisfy condition 5.[61]

Another potentially relevant issue is that addiction is not a one-time hard choice (as in typical duress) but rather causes long-term, repeated hard choices, which might indirectly affect condition 2. Since Kathy's long-term addiction led to repeated hard choices, she had opportunities to recognize patterns of the hard choices and to learn from experience. Even if she could not prevent them immediately, over time, she had more opportunities to plan ahead, acquire resources, and develop strategies to minimize their occurrence. On this view, her duress excuse would be valid only for a limited period and would gradually weaken over time. I agree that addicted people's duress excuse does not last indefinitely. It can fade as the person has more opportunities making long-term plans to create alternative solutions. Nevertheless, as Watson notes, adjusting one's life circumstances takes time.[62] Thus, after developing an addiction, an addicted person may need an extended period to create alternative solutions.

60 In other studies of women who use drugs in impoverished neighborhoods, many share similar experiences of entering and remaining in prostitution due to economic necessity. One study shows that a significant number of women are victims of child abuse, leading them to run away and become homeless before adulthood (Silbert and Pines, "Sexual Child Abuse as an Antecedent to Prostitution"). They subsequently resort to sex work to earn money and survive, making them perfect prey for pimps and drug dealers (Yates et al., "A Risk Profile Comparison of Runaway and Non-Runaway Youth"; Sterk, *Fast Lives*, 31, 177; and Abramovich, "Childhood Sexual Abuse as a Risk Factor for Subsequent Involvement in Sex Work"). Such women often start using drugs in early to middle adolescence, before fully understanding and being responsible for their decisions. Regarding pregnancy and abortion, these women often fear negotiating safe sex with partners or customers for fear of violence (Sterk, *Fast Lives*, 102–3, 146, 149). Some women lack sexual and reproductive knowledge and have misconceptions about the impact of drugs on fertility (Sharpe, *Behind the Eight Ball*, 119–25). Irregular menstrual cycles caused by drug use further obscure pregnancy recognition (Sterk, *Fast Lives*, 105). Other factors preventing abortion include financial constraints, lack of access, religious beliefs, and feelings of guilt (Sterk, *Fast Lives*, 110; and Knight, *Addicted, Pregnant, Poor*, 78).

61 In some sense, the discussion here can be interpreted as a more fully worked-out instance within a broader Watsonian framework. It lays out the descriptive details of Kathy's case, which help bring into focus the kinds of social norms that might judge her more sympathetically.

62 Watson, "Excusing Addiction," 615.

For addicted people living in poverty in inner-city neighborhoods, this period may be even longer. Their resources and flexibility to modify their daily routines can be severely constrained.[63] Because they generally have fewer resources and less flexibility than addicted people in more stable circumstances, they may be able to invoke a duress excuse for significantly longer.[64]

Therefore, one might conclude, informed by her testimony and empirical research on the prevailing living conditions in her community, that Kathy's circumstances exhibit the necessary descriptive characteristics to satisfy the conditions of a duress excuse of addiction. These characteristics enable us to apply a duress-like excuse to argue that Kathy's hardships mitigate the blameworthiness for at least some of her addiction-related parental failures, particularly if such failures did not result in grave outcomes.

Even skeptics of Kathy's duress-like excuse can find this framework useful in helping them articulate why Kathy's socioeconomic condition might not excuse her wrongdoing. For example, Kathy's testimony about not wanting to "deal with" her problems is ambiguous. One possible interpretation is that she used drugs merely to avoid confronting her problems rather than to self-medicate the distress they caused. Under this interpretation, her situation might not constitute duress. The interview does not provide conclusive evidence for either interpretation. Besides, if it turns out that Kathy's neglect of her children's basic needs severely impacted their health, then the significance of her distress may not have been comparable to it. Additionally, if Kathy's poverty turns out not to have been that severe, it may have caused only moderate emotional pain—not enough to satisfy duress's severe distress condition 4. More information is also needed to verify whether Kathy started using crack cocaine only after becoming an adult with sufficient understanding of its effects (condition 5) and whether she had access to alternative supports like treatment (condition 2).[65]

63 Consider the well-documented phenomenon of poverty traps in economics: the demands of day-to-day survival consume time, money, and energy, making it difficult to plan for or invest in the future. Tasks that are relatively simple for most, such as securing stable housing, paying for laundry, affording next week's bus fare, or finding a safe place to store personal belongings, can be surprisingly challenging. This chronic scarcity of resources makes it even harder to break free from the cycles of poverty and addiction, potentially leaving those in poverty in addiction-related duress for months or even years.

64 Additionally, long-term addiction itself can potentially make managing drug use increasingly difficult. Neurological changes, entrenched habits, and psychological dependence may develop faster than a person's growing ability to make changes, further constraining their options over time. I am grateful to an anonymous reviewer and Reuven Brandt for pressing me to consider this issue.

65 One might question the reliability of Kathy's testimony, suspecting her of framing her substance use as a choice made under duress. However, the broader issue of trusting testimony

The final assessment of Kathy's blameworthiness is contingent on the specific details of her daily life. We should see the discussion over her blameworthiness as partially about the empirical facts of her circumstances. And the duress-like framework provides us the theoretical resources to identify what empirical facts about Kathy's socioeconomic condition are relevant to her blameworthiness for her child neglect and to investigate whether and to what extent they affect her blameworthiness.

More broadly, the duress-like framework can be used to evaluate similar addiction cases with a prominent socioeconomic dimension. We can examine whether a person has a duress-like excuse by investigating whether her purpose of substance use is for alleviating distress and whether her distress is partly caused by her socioeconomic environment. Just like in Kathy's case, whether and to what extent her offense is excused in this way depends on the nature and gravity of the offense and other facts about how she is positioned in her environment, including its harshness, the severity of the distress it causes, the availability of assistance or alternative resources to cope with her distress, and how she came to be in it.

It is important to note that the mere presence of socioeconomic factors—be it peer pressure, loss of community support, poverty, trauma, or others—is not sufficient to justify a situational excuse on this framework. What determines whether a person's socioeconomic factors constitute an excuse is these factors' combined effect and, most importantly, whether they create a duress-like situation for the person. The works of Kennett and colleagues, as well as Burdman, recognize that such factors can constitute a situational excuse. Nevertheless, my duress-like framework extends this recognition by providing detailed answers on when, how, and to what extent these factors can serve as a situational excuse. This development is essential for accurately evaluating real-world addiction cases influenced by socioeconomic factors.

The duress-like framework also helps identify what type of environmental factors are relevant to the mitigation of a person's blameworthiness. In short, any environmental hardship that contributes to the collective effect of a duress-like situation is relevant, including those seemingly minor or irrelevant ones. Everyday aspects of poverty and marginalization, such as sex workers' traumatic experiences with customers, the transportation difficulties to reach local laundry shops and grocery stores, or the lack of childcare resources in addiction treatment centers, can contribute to a duress-like situation when combined with other factors. Much like the straws that break a camel's back or the bars that form a bird's cage, these specific details of an addicted person's

is complex, and I do not address it in this article.

life and environment are important information for determining whether her blameworthiness is partially excused for a duress-like reason. Each case should be assessed individually, with specific case details determining the validity of the person's excuse.

4. Evaluating the Scope of the Duress-Like Framework

Some might raise the objection that my duress-like approach is too narrow in its scope, concentrating exclusively on addicted persons in poverty-stricken, violent communities who use substance for self-medication. However, given the widespread substance abuse in economically disadvantaged areas and the clinical support for the self-medication hypothesis, the duress-like approach remains valuable. It can effectively assess blameworthiness in a considerable number of real-world addiction-related offenses.

Besides, self-medication due to social marginalization and oppression is not the only way social factors might affect an addicted person's blameworthiness. My aim here is not to establish an exhaustive framework for excusing addiction influenced by social factors but to address theoretical gaps. The duress-like framework can be expanded and supplemented. Apart from self-medication, there might be other ways in which a socially disadvantaged addicted person finds herself in a duress-like situation. Apart from being in a duress-like situation, the socially disadvantaged addicted person might have other situational excuses to mitigate her blameworthiness. There is abundant room for future research.

Others may worry that my approach is overly broad. By allowing a duress-like excuse for socially disadvantaged addicted persons, we may inadvertently legitimize excuses for a wide array of wrongful acts. A related concern is that my framework focuses too much on social disadvantage that create duress-like situations, with addiction playing a minor role. Consequently, any wrongful behavior might be excused if linked to the "right" kind of social disadvantage.

My response to both objections centers on the same component: the self-medication model. Addressing the second objection, the self-medication model of addiction is central in forming the structure of a hard choice in duress (like Kathy's choice between self-medicating and caring for her children). Without addiction, we have no reason to say the wrongful behavior is about self-medication, and the hard choice structure collapses. Regarding the first objection: the self-medication model's application in addiction is supported by substantial empirical evidence and should not be arbitrarily extended beyond addiction-related behaviors. Concerns that this model could excuse a wide array of offenses overlook the rigorous standards required for its clinical application. While it is possible to worry about the model unpalatably excusing

villainous behaviors, such an extension would require establishing a self-medication model for those behaviors first, which is highly unlikely.

Nevertheless, I do think that my approach could be extended to other offenses related to substance use, including distress-motivated substance use triggered by treatment-resistant illnesses (such as treatment-resistant depression and severe chronic physical pain), as well as nonaddicted, distress-motivated substance use. Some might argue that if my approach can be applied to cases without socioeconomic challenges, or even to nonaddicted drug use, then addiction and socioeconomic challenges seem less central to the account. My response to this is as follows: while my account discusses factors highly relevant to the nature of addiction, it is not meant to define the nature of addiction or to demarcate addiction from nonaddiction. Nor does it aim to distinguish socioeconomic influences from other possible causes of self-medication, such as chronic pain or personal distress. Rather, it is grounded in the empirical reality that many real-world cases of addiction involve individuals in adverse socioeconomic conditions, which often contribute to the development and persistence of addiction. These cases place a demand on theories of moral responsibility to provide accurate evaluation of blameworthiness, and my account is developed to respond to that demand. It offers a focused framework for evaluating blameworthiness so that the blameworthiness of addicted people in these real-world cases can be assessed more accurately. If other nonsocial or nonaddiction cases can also be assessed more accurately in my account's framework, that should not be a reason against it.

5. CONCLUSION AND IMPLICATIONS

In this article, I have presented the case of Kathy, a person who was addicted to substances and committed a moral infraction, and I have discussed the ways in which her socioeconomic background might affect her blameworthiness. My broader goal has been to use Kathy's case as an example to construct a theoretical framework highlighting a duress-like reason that may partially excuse the blameworthiness for addiction-related wrongdoings, which connects to the socioeconomic dimensions of addiction.

Seen in the structure of the fair opportunity theory of responsibility, my work in this article can be described in the following way: cases of addiction-related wrongdoings and responsibility can be categorized along a spectrum. At one end, responsibility is mitigated purely by impaired normative competence; at the other, it is mitigated solely by a lack of situational control. The current philosophical literature predominantly explores the normative competence side. Though there are some discussions of the relevance of situational control, these are less common and remain incomplete.

The self-medication model and choice theory of addiction show that substance use by addicted individuals involves some degree of choice. This is important for the discussion about addicted wrongdoers' situational control because only when a person's action involves choice making does it make sense to talk about the situation in which she makes the choice. However, the choice theory stops short of examining the situation in which the addicted person makes the choice and how the situation might affect the person's responsibility. As a result, for real-world addiction and responsibility cases that are located near the other end of the spectrum and that have strong social dimensions (such as Kathy's), current resources fall short in assessing how social factors might influence blameworthiness. Since global data show a profound and widespread connection between substance abuse and socioeconomic factors across both developing countries and impoverished regions in developed countries, it is critical to address this gap. My duress-like framework bridges this gap by providing theoretical resources for a more effective evaluation of blameworthiness in cases of addiction that have a significant social dimension.[66]

University of California San Diego
keyaoyang917@gmail.com

REFERENCES

Abramovich, Evelyn. "Childhood Sexual Abuse as a Risk Factor for Subsequent Involvement in Sex Work: A Review of Empirical Findings." *Journal of Psychology and Human Sexuality* 17, nos. 1–2 (2005): 131–46.

Agule, Craig K. "Being Sympathetic to Bad-History Wrongdoers." *Pacific Philosophical Quarterly* 102, no. 1 (2021): 147–69.

Alexander, Bruce K. *The Globalization of Addiction: A Study in Poverty of the Spirit*. Oxford University Press, 2011.

Bachman, Ronet, Samantha Rodriguez, Erin M. Kerrison, and Chrysanthi Leon. "The Recursive Relationship Between Substance Abuse, Prostitution, and Incarceration: Voices from a Long-Term Cohort of Women." *Victims and Offenders* 14, no. 5 (2019): 587–605.

66 I am grateful to my dissertation committee, especially to my co-chairs Monique Wonderly and Dana Nelkin, for their invaluable guidance and support. I also thank two anonymous reviewers for their thoughtful comments and suggestions. An earlier version of this paper benefitted from questions and comments raised by audience members at the American Philosophy Association Central Division meeting in 2023 and was supported by the Institute for Practical Ethics, University of California, San Diego.

Brink, David O. *Fair Opportunity and Responsibility*. Clarendon Press, 2021.
Brink, David O., and Dana K. Nelkin. "Fairness and the Architecture of Responsibility." In *Oxford Studies in Agency and Responsibility*, vol. 1, edited by David Shoemaker. Oxford University Press, 2013.
Bungay, Vicky, Joy L. Johnson, Colleen Varcoe, and Susan Boyd. "Women's Health and Use of Crack Cocaine in Context: Structural and 'Everyday' Violence." *International Journal of Drug Policy* 21, no. 4 (2010): 321–29.
Burdman, Federico. "A Pluralistic Account of Degrees of Control in Addiction." *Philosophical Studies* 179, no. 1 (2022): 197–221.
Daniulaityte, Raminta, and Robert G. Carlson. "'To Numb Out and Start to Feel Nothing': Experiences of Stress Among Crack-Cocaine Using Women in a Midwestern City." *Journal of Drug Issues* 41, no. 1 (2011): 1–24.
Dressler, Joshua. "Duress." In *The Oxford Handbook of Philosophy of Criminal Law*, edited by John Deigh and David Dolinko. Oxford University Press, 2011.
———. *Understanding Criminal Law*. 6th ed. The Understanding Series. LexisNexis, 2012.
Farley, Melissa, and Howard Barkan. "Prostitution, Violence, and Posttraumatic Stress Disorder." *Women and Health* 27, no. 3 (1998): 37–49.
Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press, 2000.
Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (1971): 5–20.
Hart, Carl L. *High Price: A Neuroscientist's Journey of Self-Discovery that Challenges Everything You Know About Drugs and Society*. Harper Perennial, 2014.
———. "Viewing Addiction as a Brain Disease Promotes Social Injustice." *Nature Human Behaviour* 1, no. 3 (2017).
Heather, Nick, and Gabriel Segal, eds. *Addiction and Choice*. Oxford University Press, 2016.
Holton, Richard, and Kent Berridge. "Addiction Between Compulsion and Choice." In Levy, *Addiction and Self-Control*.
———. "Compulsion and Choice in Addiction." In Heather and Segal, *Addiction and Choice*.
Husak, Douglas N. "Addiction and Criminal Liability." *Law and Philosophy* 18 (1999): 655–84.
Kennett, Jeanette. "Just Say No?" In Levy, *Addiction and Self-Control*.
Kennett, Jeanette, Nicole A. Vincent, and Anke Snoek. "Drug Addiction and Criminal Responsibility." In *Handbook of Neuroethics*, edited by Jens Clausen and Neil Levy. Springer, 2015.
Kennett, Jeanette, and Jessica Wolfendale. "Self-Control and Moral Security." In

Oxford Studies in Agency and Responsibility, vol. 6, edited by David Shoemaker. Oxford University Press, 2019.

Khantzian, Edward J. "The Self-Medication Hypothesis of Addictive Disorders." *American Journal of Psychiatry* 143, no. 8 (1986): 1063–64.

———. "The Self-Medication Hypothesis of Substance Use Disorders: A Reconsideration and Recent Applications." *Harvard Review of Psychiatry* 4, no. 5 (1997): 231–44.

———. "Understanding Addictive Vulnerability: An Evolving Psychodynamic Perspective." *Neuropsychoanalysis* 5, no. 1 (2003): 5–21.

Knight, Kelly Ray. *Addicted, Pregnant, Poor*. Duke University Press, 2020.

Kramer, Lisa A. "Emotional Experiences of Performing Prostitution." *Journal of Trauma Practice* 2, nos. 3–4 (2004): 186–97.

Kurtz, Steven P., Hilary L. Surratt, Marion C. Kiley, and James A. Inciardi. "Barriers to Health and Social Services for Street-Based Sex Workers." *Journal of Health Care for the Poor and Underserved* 16, no. 2 (2005): 345–61.

Levy, Neil, ed. *Addiction and Self-Control*. Oxford University Press, 2013

———. "Addiction as a Disorder of Belief." *Biology and Philosophy* 29, no. 3 (2014): 337–55.

———. "Addiction, Responsibility and Ego-Depletion." In *Addiction and Responsibility*, edited by Jeffrey Poland and George Graham. Massachusetts Institute of Technology Press, 2011.

———. "Addiction: The Belief Oscillation Hypothesis." In *The Routledge Handbook of Philosophy and Science of Addiction*, edited by Hanna Pickard and Serge H. Ahmed. Routledge, 2019.

Mackenzie, Catriona. "Moral Responsibility and the Social Dynamics of Power and Oppression." In *Moral Responsibility and the Social Dynamics of Power and Oppression*, edited by Katrina Hutchison, Catriona Mackenzie, and Marina Oshana. Oxford University Press, 2018.

Maher, Lisa. *Sexed Work: Gender, Race, and Resistance in a Brooklyn Drug Market*. Oxford University Press, 2002.

McKenna, Michael. "Power, Social Inequities, and the Conversational Theory of Moral Responsibility." In *Moral Responsibility and the Social Dynamics of Power and Oppression*, edited by Katrina Hutchison, Catriona Mackenzie, and Marina Oshana. Oxford University Press, 2018.

Morse, Stephen J. "Addiction, Choice, and Criminal Law." In Heather and Segal, *Addiction and Choice*.

———. "Addiction and Criminal Responsibility." In *Addiction and Responsibility*, edited by Jeffrey Poland and George Graham. Massachusetts Institute of Technology Press, 2011.

———. "Hooked on Hype: Addiction and Responsibility." *Law and Philoso-*

phy 19, no. 1 (2000): 3–49.
Nelkin, Dana K. "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness." *Noûs* 50, no. 2 (2016): 356–78.
Pickard, Hanna. "Addiction and the Self." *Noûs* 55, no. 4 (2021): 737–61.
———. "Psychopathology and the Ability to Do Otherwise." *Philosophy and Phenomenological Research* 90, no. 1 (2015): 135–63.
———. "The Purpose in Chronic Addiction." *AJOB Neuroscience* 3, no. 2 (2012): 40–49.
———. "Responsibility Without Blame for Addiction." *Neuroethics* 10, no. 1 (2017): 169–80.
Pickard, Hanna, and Lisa Ward. "Responsibility Without Blame: Philosophical Reflections on Clinical Practice." In *The Oxford Handbook of Philosophy and Psychiatry*, edited by K. W. M. Fulford, Martin Davies, Richard Gipps, et al. Oxford University Press, 2013.
Raphael, Jody, and Deborah L. Shapiro. *Sisters Speak Out: The Lives and Needs of Prostituted Women in Chicago*. Center for Impact Research, 2002.
Sallmann, Jolanda. "'Going Hand-in-Hand': Connections Between Women's Prostitution and Substance Use." *Journal of Social Work Practice in the Addictions* 10, no. 2 (2010): 115–38.
Schroeder, Timothy, and Nomy Arpaly. "Addiction and Blameworthiness." In Levy, *Addiction and Self-Control*.
Sethi, D., S. Watts, A. Zwi, J. Watson, and C. McCarthy. "Experience of Domestic Violence by Women Attending an Inner City Accident and Emergency Department." *Emergency Medicine Journal* 21, no. 2 (2004): 180–84.
Sharpe, Tanya Telfair. *Behind the Eight Ball: Sex for Crack Cocaine Exchange and Poor Black Women*. Haworth Press, 2005.
Silbert, Mimi H., and Ayala M. Pines. "Sexual Child Abuse as an Antecedent to Prostitution." *Child Abuse and Neglect* 5, no. 4 (1981): 407–11.
Sripada, Chandra. "Addiction and Fallibility." *Journal of Philosophy* 115, no. 11 (2018): 569–87.
Sterk, Claire E. *Fast Lives: Women Who Use Crack Cocaine*. Temple University Press, 1999.
Surratt, Hilary L., James A. Inciardi, and Steven P. Kurtz. "Prescription Opioid Abuse Among Drug-Involved Street-Based Sex Workers." *Journal of Opioid Management* 2, no. 5 (2006): 283–89.
Surratt, Hilary L., James A. Inciardi, Steven P. Kurtz, and Marion C. Kiley. "Sex Work and Drug Use in a Subculture of Violence." *Crime and Delinquency* 50, no. 1 (2004): 43–59.
United Nations Office on Drugs and Crime (UNODC). *World Drug Report 2016*. United Nations, 2016. https://www.unodc.org/wdr2016/.

———. *World Drug Report 2018*. United Nations, 2018. https://www.unodc.org/wdr2018/.

———. *World Drug Report 2020*. United Nations, 2020. https://wdr.unodc.org/wdr2020/.

Valera, Roberto J., Robin G. Sawyer, and Glenn R. Schiraldi. "Perceived Health Needs of Inner-City Street Prostitutes: A Preliminary Study." *American Journal of Health Behavior* 25, no. 1 (2001): 50–59.

Wallace, R. Jay. "Addiction as Defect of the Will: Some Philosophical Reflections." *Law and Philosophy* 18, no. 6 (1999): 621–54.

Watson, Gary. "Disordered Appetites: Addiction, Compulsion and Dependence." In *Addiction: Entries and Exits*, edited by Jon Elster. Russell Sage Publications, 1999.

———. "Excusing Addiction." *Law and Philosophy* 18, no. 6 (1999): 589–619.

———. "Free Agency." *Journal of Philosophy* 72, no. 8 (1975): 205–20.

Yaffe, Gideon. "Are Addicts Akratic?" In Levy, *Addiction and Self-Control*.

———. "Compromised Addicts." In *Oxford Studies in Agency and Responsibility*, vol. 5, edited by D. Justin Coates and Neal A. Tognazzini. Oxford University Press, 2019.

———. "Lowering the Bar for Addicts." In *Addiction and Responsibility*, edited by Jeffrey Poland and George Graham. Massachusetts Institute of Technology Press, 2011.

Yates, Gary L, Richard MacKenzie, Julia Pennbridge, and Eric Cohen. "A Risk Profile Comparison of Runaway and Non-Runaway Youth." *American Journal of Public Health* 78, no. 7 (1988): 820–21.

Young, Amy M., Carol Boyd, and Amy Hubbell. "Prostitution, Drug Use, and Coping with Psychological Distress." *Journal of Drug Issues* 30, no. 4 (2000): 789–800.

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · AUGUST 2025
https://doi.org/10.26556/jesp.v30i5.4144

WAR, LEGITIMACY, AND DEMOCRACY

COMMENTS ON RENZO

Thomas Christiano

Massimo Renzo's view on the duty to fight in war is an eminently reasonable position.[1] It is one with which I find myself in substantial agreement. The mix of revisionism and orthodoxy is a good one. The basic thesis is that citizens of a legitimate state have a presumptive duty to fight when ordered by the state to fight in a war. The duty is presumptive in the sense that it can be defeated by a justified belief on the part of the citizen that the state is about to engage in a seriously unjust war. There is also a duty for each citizen to figure out whether the proposed war is just or unjust. But if the citizen does not arrive at a justified belief that the war is unjust, then she has a duty to obey the state's command to fight. The presumption is in favor of duty to obey the state's command as long as the state is legitimate. This is meant to contrast with the revisionist view that the citizen has a duty not to do as the state commands if the citizen does not justifiably think the war is just. The revisionist presumption is against the duty to obey.

Renzo's view is said to vindicate a weak version of the moral equality of the combatants. What this means is that some combatants on both sides of the war can be permitted to fight. But this is meant to be a weaker version than the strong classical (or "traditional") version of the argument, which asserts that all combatants are permitted to fight.[2] The difference is that only combatants from legitimate and therefore minimally just states are permitted to fight.

1. MORAL EQUALITY

I do not see how the moral equality idea is supposed to work: the outcome seemingly is that in order to be permitted to fight, one must be unjustified in thinking that the war is unjust. In this case, moral equality holds only between those who fight for a legitimate state and those who are not justified in thinking

1 See Renzo, "Political Authority and Unjust War."

2 See Walzer, *Just and Unjust Wars*.

that the war is unjust. Those who fight yet justifiably think that the war is unjust are not moral equals: they are criminals.

If a legitimate state seems to be attacking a nonlegitimate state, does this imply that the members of the nonlegitimate state are not allowed to fight back under the authority of the illegitimate state? Are they to be thought of as criminals if they fight under the authority of the mostly illegitimate state?

Perhaps the members of an illegitimate state can treat their state's commands as if they are authoritative. Renzo seems to reject this idea. He seems to say that a state is either legitimate or illegitimate.[3] Legitimacy must depend on possessing a complex set of properties simultaneously. Hence, it appears that the legitimacy of the state cannot be piecemeal; it must be holistic. This approach is not adhered to on all dimensions, as far as I can tell, because Renzo does say that people with differing amounts of knowledge or differing access to knowledge might have different obligations.[4] So there is a certain amount of piecemeal legitimacy regarding who is under a duty. But Renzo seems to think that if a state is legitimate in one area, it must be in most or all other areas. But it is not clear that this is consistent with a thoroughgoing instrumentalism. Why not think that a state's commands enable you to act better in accordance with duty in some areas but not others? Suppose there is a state that is mostly illegitimate in the above sense, but that state gives good commands on reasonably just grounds in particular areas. Why not think, on the instrumentalist view, that these commands are authoritative? This might then be applied to war-making authority. Suppose a state is mostly illegitimate because unjust, but it is right for it to resist attack by another normally legitimate state. It is right to resist the overbearing behavior of the legitimate state. It seems not unreasonable to think that the commands of a mostly illegitimate state in wartime can have legitimacy and can bind its citizens in this area.

The notion of moral equality inevitably becomes very messy here. Might we take a political approach to this question, just as Renzo takes a kind of political approach to the question of the justification of an individual's going to war? In light of the messiness of actual assessments of guilt, might we try to construct legal and political institutions that determine this and replace messy moral assessments with legal assessment? Revisionists are not generally unfriendly to legal institutions with more tolerant attitudes toward ordinary soldiers, on the grounds that such institutions have better consequences. But the question becomes more pressing for Renzo since he introduces political institutions in a way that affects the justification of individual behavior.

3 Renzo, "Political Authority and Unjust War," 346.

4 Renzo, "Political Authority and Unjust War," 351n39.

2. INTERNATIONAL LEGITIMACY

The worry, expressed by Jeff McMahan, is that it is unclear how the legitimacy of a state can be relevant to the question whether it is morally justified for a member of that state to go to war against some other society.[5] The problem is that it appears that the relations among members of the state do not seem to justify killing innocent persons of another society, and it seems unnecessary to the justification of defending a society against an unjust attack. Since the legitimacy of the state in question does not affect the justice or injustice of attack, it is hard to see why legitimacy ought to play any role in justifying the decision of a member to go to war.

Renzo argues that an instrumentalist account of legitimacy is uniquely able to answer this question. His instrumentalist account asserts that a political authority has legitimate power over a person to the extent that the authority enables the person to discharge some of her most important duties of justice. It does this by giving content-independent and presumptive reasons for action to the members. Among the most important duties of justice a person has is the duty not to kill innocent persons. The main point is that part of the ground of the legitimacy of a state, on this conception, is that the state does not or is highly unlikely to engage in unjustified wars. And so it is highly unlikely to require individuals to fight in unjust wars. To be sure, this does not imply that the state never makes mistakes or that it never orders persons to fight in unjust wars, but it does imply a significant reduction in such unjust adventures.

The instrumentalist approach to legitimacy looks like a plausible way to respond to the McMahan worry. Some intrinsic approaches focus exclusively on relations among members of states. Associative conceptions of political obligation and authority focus on the qualities of relations among citizens and ground duties to obey on the qualities of those relations. Consent theory, fairness theories, and democratic theories all focus on the relations among members of states. The way they are usually developed ignores the relations between members and nonmembers. This is clearly a kind of incompleteness in the way these theories are usually elaborated.

Let us consider the instrumentalist approach. The thesis is that a state must have authority in order to avoid the perils of the state of nature, and a state must have authority to decide to go to war against other states and impose duties on citizens also to avoid the perils of the state of nature. Renzo argues that when legitimate states decide to go to war, it is in everyone's interests, even those who are not part of these states, that the war-making states are legitimate states. And

5 McMahan, *Killing in War*.

this is because legitimate states usually do not order people to go to war when there is little acceptable reason for it; consequently, people's general interests against unjustified warfare are better advanced and the duties of persons to avoid unjustified war are better fulfilled when legitimate states have the powers to declare or not declare war and to impose duties on their citizens to participate or not.[6] And this authority implies that it is better for a citizen to obey a state that acts in good faith even if mistakenly as long as the citizen does not justifiably think the war is unjust.

The basis for this is that a world in which legitimate states do not have the authority to go to war and impose duties to fight would not be a world without conflict. What would happen is that many other groups would fight with each other without any concern for justice in warfare. Hence, justice would be less well served, and everyone's legitimate interests, including even those who would otherwise have been attacked by a legitimate state, would be set back. But I am not sure what this counterfactual is, and I am not sure whether the counterfactual holds. Does the counterfactual merely refer to many nonlegitimate states fighting each other? How is this a counterfactual, since of course this can occur whether there are legitimate states with war-making authority or not?

A more reasonable counterfactual is that a world in which states do not have the power to require people to go to war is one in which many independent groups would end up fighting with each other: there would be small-scale violence occurring in many places. But I am not sure how this counterfactual makes sense either. Presumably, states would suppress these conflicts just as they try to suppress conflicts between gangs and between individuals. There may be some tendency to want to fight among persons, but a state's police power might well be sufficient to dampen this. This would not require war-making powers on the part of states, just ordinary police powers. Furthermore, if a state could not suppress all warlike actions within its boundaries, why should we assume that these actions would necessarily be worse than those of states? And is there a general answer to this question? Maybe it would be better to have legitimate states without the power to wage war in some important circumstances. So I am not sure this counterfactual supports Renzo's instrumentalist argument.

But there is also a worry that these counterfactuals are not the right ones. Why not consider a counterfactual world in which states have the power to impose duties of soldiering on their citizens—but only in justified wars? There is still authority here: states must decide whether to engage in justified war or not, and they must decide how to do this. But a state has no authority to

6 Renzo, "Political Authority and Unjust War,"352.

engage in unjustified war, and it cannot impose duties on citizens to participate. Instead of looking at a world in which the state cannot go into war at all, we look at one in which the state cannot declare an unjustified war. There are two points to be made here: I do not know why this is not the right counterfactual test, and I do not know what this counterfactual world would likely look like. If it is the right counterfactual test, then we have a view that is closer to McMahan's. Good-faith mistakes are not compatible with the exercise of legitimate authority, and soldiers have no duty to fight in wars that are the result of good-faith mistakes.

The problem of disagreement poses a challenge to this approach. What if too many people think too often that the state is acting in an unjustified way? Would this disable the state's legitimate war-making activities? Or is this unlikely to happen so that the state's war-making capacity would not be unduly jeopardized? In the end, the instrumentalist approach must answer this question, I think.

3. DEMOCRACY WORRIES

Renzo's discussion is a plausible one and balances the considerations backing the orthodox view and those behind the revisionist view. But there is still a puzzle that he does not entertain. This is the result of his official account of legitimacy as being grounded in the instrumental value of the state helping people satisfy their duties of justice.[7] The idea seems to be that the state is legitimate because it helps people fulfill duties of justice that are not inherently connected with the operation of the political system.

Though this is certainly a possible way to think about legitimacy when we think a state is legitimate because it enables people to fulfill their duties of justice, it is not the only way. For the duties of justice that a state enables citizens to fulfill may include democratic duties, and the state may enable people to fulfill these duties by being democratically organized and run. One way of giving an account of this duty is to say that one must treat one's fellow citizens as equals in the context of disagreement and fallibility, and the way to treat them as equals is to respect the outcome of democratic decision-making even when it goes against one's own judgment.[8] Here there is a duty of justice to go along with the democratic decision even when one disagrees with it. And this duty can be among the duties the state helps one fulfill when it is reasonably just. In this case, the democratic state's activity is not instrumentally connected with the

7 Renzo, "Political Authority and Unjust War,"352.

8 See Christiano, *The Constitution of Equality*.

fulfillment of this duty; rather, it is the intrinsic features of the state that generate the duty, which then must be fulfilled. So though this account satisfies the formal account of legitimacy that Renzo announces, it is not instrumentalist. Granted, not everyone accepts the idea that democracy is a requirement of justice, but it is widely accepted as a just solution to the facts of disagreement that are pervasive in societies. That fact alone gives us reason to consider the implications for the context of the duty to comply with democratic decisions to go to war.

The trouble is that this account poses a problem for Renzo's view of when the presumptive reason for obedience is defeated. Recall that the presumptive reason for defeat is that one has a justification for believing the war is unjust and is confident in one's judgment. But it is actually a fairly standard case in democracies that people on different sides of a question have justification for their views and confidence in them. Some proportion of the population disagrees with the majority decision, thinks of itself as justified in this disagreement, and is confident in its belief. They then have reason not to go along. If, however, we think that there is a distinctive democratic set of duties that includes complying with a decision that has been made in an egalitarian way for the reason that there is good-faith disagreement, then it looks like one may have a duty to go along with the decision and comply.

So let us just spell out the case. What if there is disagreement about the question whether a war is justified or not? What if some think it is justified, and others think it is not? Let us suppose that after lively debate, those who think the war is justified win a majority for their point of view. And suppose that they choose war. Do the others have a duty to go along? Suppose that they disagree because of a difference in underlying values. One group thinks that going to war to preempt a probable future attack is justified, while the other thinks that only a near certain future attack can justify preemptive war. After lengthy and lively debate, the majority sides with the probable-future-attack group. The majority and the minority both think they are justified. They may be confident in their views.

Is there no reason here for the minority, which thinks its own view is justified, to go along with the judgment of the majority to participate in the war? Their views have been carefully considered and argued against in good faith (let us suppose). Are they still permitted, indeed required, to disobey? It is not clear to me that they are. But Renzo must think that they are because, by hypothesis, they have justification for their beliefs. This justification must be either belief relative or evidence relative (under the supposition that people have access to different evidence) and not fact relative. But I am not convinced that there is not some kind of duty still to go to war here.

The problem is that a decision to go to war is a collective decision. A democratic conception of authority implies that it must be made in an egalitarian way and that when it is made in an egalitarian way, it must be obeyed on pain of failing to treat one's fellow citizens as equals. We here run directly into the worry that McMahan raises. The question is: How can my relations with my fellow citizens have a bearing on whether it is right to go to war against someone who is not a fellow citizen? The worry is acute here because those one is warring against have not participated in the process of democratic decision, and they are not being treated as equals when we go along with the majority. And in addition, if those the society is planning on warring against are genuinely not liable to be killed, then the killing is wrong. Or so someone might be justified in thinking.

It seems to me that the only way to think about this situation is as a kind of ethical dilemma. This is in contrast to Renzo, who more neatly characterizes the duty to obey as being undercut by any justified and confident belief in the wrongness of the war. On a democratic view, the duty to one's fellow citizens conflicts with the duty to those who are the victims of an unjust war. The duty to one's fellow citizens provides a reason to exclude consideration of the wrongness of the war even if one has lost out in the democratic contest, precisely because it was a contest carried out in a just democratic manner. But this is only in the context of fellow citizens. Once we take into account others who are not part of the democratic society but to whom one owes duties, they too become sources of independent reasons that cannot be excluded by democratic considerations.

There may be resources in many of these theories to help ground an extended notion of legitimate state action in the international realm. To go to war for the sake of conquest looks like a violation of a kind of democratic norm because the conquered, by hypothesis, do not have a say in the process and will likely be killed if they resist. They certainly do not seem to be treated as equals in any democratic way. (Although what if the proposed conquest is of a small community that would be a kind of minority?) And while consent theory may assert that people have consented to the authority that is ordering the attack, it can also say that an unprovoked attack on another society is itself a violation of consent theory. These theories could suggest that when an authority sends persons into an unjust war, it violates the very norm that grounds the legitimacy of the authority and thus cannot be acting legitimately. A consent theorist, for example, might argue that one cannot or may not consent to an action that violates the fundamental rights of another person. And I have argued elsewhere that a democratic approach to international politics, under the conditions in which states are the main actors, implies that the requirement of state consent

on reasonable terms is the appropriate constraint on how states relate to each other.[9] Obviously, invasion for the sake of conquest or other merely self-interested concerns does not satisfy the state-consent norm. War for the sake of self-defense, for the sake of protection of other states, and maybe even for the sake of humanitarian intervention in extreme cases may be defensible on these accounts.

Democratic societies seem to observe these constraints well in relation to each other; they never go to war with each other.[10] But there is substantial disagreement within democratic societies as to whether and when it is permissible to go to war against nondemocratic societies. The above limitations do not solve this problem or guarantee a clear and public answer. In those cases, a moral dilemma arises.

University of Arizona
thomasc@arizona.edu

REFERENCES

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

———. "The Legitimacy of International Institutions." In *Routledge Companion to the Philosophy of Law*, edited by Andrei Marmor. Routledge, 2012.

Levy, Jack, and William Thompson. *Causes of War*. Wiley Blackwell, 2010.

McMahan, Jeff. *Killing in War*. Oxford University Press, 2011.

Renzo, Massimo. "Political Authority and Unjust War." *Philosophy and Phenomenological Research* 99, no. 2 (2019): 336–57.

Walzer, Michael. *Just and Unjust Wars*. Basic Books, 1977.

9 See Christiano, "The Legitimacy of International Institutions."

10 See Levy and Thompson, *Causes of War*.

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · AUGUST 2025
https://doi.org/10.26556/jesp.v30i5.4142

SOLDIERS AND MORAL TRAGEDY

COMMENTS ON RENZO

Christopher Kutz

IT HAS BEEN a great pleasure to enter into dialogue with Massimo Renzo regarding his prize-winning article on the individual moral culpability (or not) of soldiers called to do violence by their states.[1] Whether individual soldiers enjoy an individual moral excuse for that violence—outside the morally clear context of immediate defensive war and so long as it remains within the bounds of the law of war—has been a central question in debates about the morality of war since America's Vietnam War, resurging with the dubious if not criminal coalition war in Iraq. As Renzo says, much of the current popular and philosophical discussion takes one of two polar positions: the conventional position, reflected in the actual law of armed conflict (LOAC), puts the moral onus of unjust wars solely on national leaders, not soldiers; on the other hand, the radically democratic position assesses the liability of each soldier one by one, based on the information available to and pressures on them. Such positions are typically driven by a range of complex moral intuitions and pragmatic concerns, including concerns for the integrity and character of soldiers, a reciprocal interest in sparing the lives of captured soldiers who might otherwise be executed for illegal war, and a desire to unify the domestic morality of violence (with its strict limits on lethal self-defense) with the capacious permissions convention grants to soldiers in war.

Renzo's promising strategy is to bypass the immediate moral inquiry into killing and self-defense (both individual and collective)—or at least to delay that inquiry. Instead, he turns to the logic of law and legal normativity, taking inspiration from the work of the late Joseph Raz. The result is a genuinely new, subtle, and persuasive account that shows why we have to treat as distinctive our moral judgments of individuals acting through the collective technology

1 Renzo, "Political Authority and Unjust Wars." That article was the subject of an American Philosophical Association (Pacific Division) symposium discussion with me, David Estlund, and Thomas Christiano on April 5, 2021. This article reflects the remarks I made on that occasion, when I also benefitted from the thoughts of my co-panelists. See also the articles in this issue by Renzo, Estlund, and Christiano.

of law. In what follows, I probe Renzo's position, not so much to challenge it as to note its nuances and the questions it does not (yet) answer. I impose a narrative frame because discussions of legal authority, especially in a Razian vein, become abstract very fast, even though their subject is the most concrete possible embodiment of state power—in this case, a state that commands its young people to kill and die for their country.

So let us consider as moral agent a volunteer soldier of a reasonably just state, sent to fight and kill in a country not his own, whose leaders insist that their war is lawful, and who has conducted himself consistently with the *jus in bello* rules of legal combat, targeting only combatants and inflicting only proportionate harm on noncombatants. As a result of the deployment, the solider—let us call him *Jose*—has killed other combatants, at least some of whom posed a direct threat to Jose himself, as well as killed or displaced a larger number of civilians, rendering their lives substantially worse.

At the time of enlistment and then deployment, Jose was aware of a debate in the public sphere about the legality and morality of the war, perhaps even aware that a number of other just and democratic states regarded the war as illegal. But the soldier accepted the view of the civilian government, transmitted through the military authorities, that the war was legally permissible because it was being waged pursuant to some international legal authorities—and also that the war was morally righteous. Years after the deployment, with the war no longer a live issue, most jurists have now come to believe that many of the factual predicates for the war were false and/or invented and that the war was in fact illegal and manifestly unjust. Indeed, Jose himself accepts this verdict.

Theorists have asked a number of questions about situations like this, questions pertaining both to the time before Jose engaged in any acts of violence and to the time after, in retrospect. Some of these questions are:

1. Was Jose rationally justified in acting in obedience to the orders of the military, specifically in following an order that was (in the circumstances but not as known to Jose) morally impermissible?
2. Was Jose morally *obligated* to follow this order?
3. Is Jose morally culpable for engaging in acts of violence, assuming that the war was in fact illegal and unjust?
4. Would it be morally legitimate for some actual or possible juridical body, domestic or international, to try and punish Jose for what he did?
5. Are the answers to the questions above determined just by first-order moral facts (e.g., who posed an illegitimate threat to whom?) or are they at least partly determined by the further political fact that Jose

was a member both of a free and just state (when he volunteered for service) and of its military (when he followed orders to kill)?

1. THE POWERFUL TRUTH IN RENZO'S ARGUMENT

According to Renzo, philosophers' views on these questions fall into two groups: traditional and revisionist. Traditionalists, represented for Renzo principally by Michael Walzer, offer a philosophical view that coincides with the legal status quo as manifested in the law of armed conflict—a status quo that exonerates any uniformed soldier of a sovereign state who fights consistent with the *ius in bello*, regardless of the justice of the war.[2] Thus, to the questions of moral permission and obligation (1 and 2), traditionalists answer yes and yes; and to the questions about Jose's culpability, liability, and the relevance of his political membership (3, 4, and 5), they answer no, no, and yes. Traditionalists often add the judgment that different, nonpolitical answers would decrease compliance with the *ius in bello*, for instance by removing the incentive offered by prisoner-of-war status, on the grounds that if soldiers were subject to punishment if captured merely for fighting, then they would lose any reason to fight within the rules. For traditionalists, Jose's political membership is fundamental and creates, as it were, a moral division of labor: while leaders of his state may be judged culpable and punished, Jose himself is responsible only for his personal conduct on the battlefield and has a claim to be freed if captured at the end of hostilities.

Revisionists, represented for Renzo principally by Jeff McMahan and Cécile Fabre, press a different view.[3] They begin with an opposed answer to question 5 and hence potentially to question 1: if the war was in fact unjustified, Jose's being a member of a political group alone cannot turn a homicidal act into one that is permissible, much less a proper object of duty. If the war was unjust, then Jose's conduct in its pursuit cannot be justified, no matter how many people are accompanying him in the pursuit. Moreover, revisionists give possible answers of no and yes to questions 2 and 3, respectively: if the war was illegal, and Jose should have been aware of that fact, then it is morally irrelevant that his political authorities took a contrary view; Jose could not be obligated to obey, is morally culpable, and could, in an ideal world, be subject to punishment. One cannot be rationally justified, much less duty bound, in killing the innocent. Revisionists grant, however, that the answer to 4 is more complicated: even if Jose is morally culpable for killing, his liability to be punished is a complicated question,

2 See Walzer, *Just and Unjust Wars*.

3 McMahan, *Killing in War*; and Fabre, *Cosmopolitan War*.

involving Jose's particular epistemic position with regard to the legality of the war, as well as consequentialist considerations of the sort I mentioned above.[4]

An especially fine feature of Renzo's article is how he threads the needle between the traditionalist insight that political membership matters to responsibility and the revisionist insight that killing and dying under orders needs a much less causal justification than traditionalists typically offer. His argument works by coupling a Razian account of the nature of rational authority with a Kantian account of the justification of state authority. The basic element of the view is grounded in the individual rationality of relying on forms of epistemic or practical technology—investment schemes, autopilots, and so forth. On Renzo's view, Jose—and all citizens of legitimate states—are properly under a duty to comply with state authority because state authority is the only feasible means of avoiding the turmoil and destruction of the state of nature in which each follows her own counsel. The authority of legitimate states serves to connect citizens to first-order values and reasons better than they could do by deliberating themselves in first-order terms. We therefore have a duty as social beings to follow the dictates that are constitutive of the political institutions that make that sociality not only possible but valuable.

But Renzo's view is not purely Razian. Its key difference is that the duties we have, *qua* members of generally just, functionally successful polities, is that the duties they impose are only *pro tanto* and are subject to being overridden by strong counterclaims. Raz, by contrast, regards the category of authoritative reasons as categorically exclusionary.[5] Though Renzo is clear that his is a practical, not epistemic conception of political authority, I think of this as a "Wikipedia model" of authority—general reliance is rational and appropriate, but when stakes are high, we should verify separately and be open to deviation (for instance, before publishing or betting big in a bar argument).

The Wikipedia/*pro tanto* model makes better sense of the specific context of political duties because it acknowledges two plausible truths about citizenship in legitimate states. The first is that the benefits of state authority, including over soldiers, are substantial for both their own citizens and outsiders. Concretely, one might think that the sanctity of borders, especially of smaller states, is valuable in minimizing interstate violence and ensuring investment and production.

4 Although Renzo does not say so directly, there is something half-hearted about the standard revisionist position, which insists on moral responsibility but denies any express political expression of that responsibility because it would be generally unfair to punish people acting on even misguided orders in the fog of war. My own view is that this aspect of the revisionist view is coherent, and indeed, we need to make room for conceptions of responsibility that do not entail permissions to punish.

5 Raz, "Postscript."

Such borders are practically guaranteed by systems of alliance. If the soldiers who are asked to defend allied borders were invited to consider on their own the justice of interstate conflicts, the security guarantees that alliances offer would be weaker, and the status of smaller states much more vulnerable. Global turmoil would follow—as indeed it has in Ukraine, which was unable to rely on alliances to deter Russia's attacks. It is only because a small state can reliably count on both its own soldiers and those of its allies following orders that it can hope to deter larger, hungry neighbors. And this arguably preserves lives and productivity all around.

The second truth is that even generally legitimate states can act wrongfully in pursuit of their own interests, in straightforward contravention of international law. Exaggerated (and politically rewarded) conceptions of national self-interest, jingoistic politics, and groupthink are all parts of legitimate state politics and practice. And when states act wrongfully, it cannot also be the case that those they command do wrong to question their orders. Given the stakes, Jose cannot be justified in following any order to deploy, any more than he would be justified in following any tactical order. Even in a just war, Jose is morally obligated to refuse an order, say, to shell a position where he knows civilians are sheltering. Superior orders are a limited defense and do not apply when a given order is manifestly illegal, by intent or by effect. Taking both truths together, Renzo's view rightly permits Jose to regard himself as primarily morally responsible for following orders and thus not subject to the kind of moral responsibility that attends purely individually; but it also crucially makes room for complicity and personal responsibility in those circumstances when the presumption of legal orders is overcome, and Jose obeys rather than resists.

These seem to me not only to be inherently sensible philosophical considerations but also to account for phenomenology thoughtful soldiers in the midst of war—namely, though they are duty bound not to second-guess orders as a matter of course, they perform what Renzo, quoting Fred Schauer, calls a "peek" at the underlying reasoning.[6] Peeking done, they feel they can reconcile their identities as free and rational citizens with their roles as soldiers, at least until their threshold of moral justification is breached.

2. BUT ARE THEY REALLY JUSTIFIED?

Renzo's view, I have suggested, accounts for the feeling of justification that soldiers in a legitimate state have (at least at the outset of their service) and makes

6 Renzo, "Political Authority and Unjust Wars," 348, quoting Frederick Schauer, "Rules and the Rule of Law."

for a satisfying reconciliation of moral autonomy with political membership (contrary to the philosophical anarchists who say this cannot be done). But is it more than a feeling? Are the soldiers individually genuinely doing what "they have moral reason to do, all things considered" when they kill in an unjustified war, so long as their "peek" does not reveal a serious injustice?[7] In what follows, I suggest that this model of justification is unstable, and because it is unstable, it leaves open powerful dimensions of blame and regret. In that sense, Renzo's argument may veer from one side to the other in the traditionalist-revisionist debate rather than forming a sturdy middle way.

In Raz's standard service conception of authority, justification is one-dimensional: if a political authority is justified, it is because those who follow its dictates generally do better by following them (by reference to underlying reasons) than those who try to deliberate for themselves.[8] This can generate a kind of paradox or local-general inconsistency when the authority is occasionally wrong: in this instance, I may actually make the situation worse even though following an authority's dictates overall makes the situation better. The paradox is only apparent, though, because the reasons that support my following the (incidentally) wrong dictate track overall reasons and are not cancelled out by occasional error. So my justification is solid, even if I have reason to wince about the local costs.

The problem for war, as Renzo notes, is that the "costs" of error are too high to simply be assimilated into the calculus of net political benefit. They are the deaths of innocents, as well as moral harm to the soldiers themselves. This is why Renzo requires a "peek" at war's justification, while trying to preserve the overall justification for following political authority. The result is a more complex form of hybrid justification than in Raz's is simple model—hybrid because it combines a subject-relative judgment about whether Jose came to a reasonable conclusion on the basis of his "peek" with an objective judgment about the justification of the war as a whole (and, beyond that, of an authoritative political structure). This hybrid justification is thus both the promise and peril of Renzo's model.

Here is a familiar example of a hybrid justification: in the United States, when prosecutors ask whether a police officer was reasonable in using deadly force, they begin with as many facts of the situation as they can assemble and then ask whether the ways in which the officer's perceptions or beliefs departed from how an objective officer would have viewed the same situation.[9] The

7 Renzo, "Political Authority and Unjust Wars," 355.

8 Raz summarizes his view in "The Problem of Authority."

9 The standard I describe here is that laid down by the US Supreme Court in *Graham v. Connor* 389 U.S. 486 (1989).

theory is that a fully relativized view—Did this particular officer perceive a need to use force?—vindicates all uses of force and results in no standard at all. But a purely objective standard—Would a reasonable officer knowing everything about the situation have used force?—makes no room for the particularities of the situation and so would be unfair to the actual officer. The justification must therefore be both relative and objective.

Renzo's standard similarly has two components, relative and objective, that combine for him into a judgment of the soldier's justification. Here is how I understand them as combining, first at the extremes and then in the problematic (and more realistic) middle case.

1. Relative and objective justifications line up in support: any person in Jose's epistemic position would have concluded that following orders to fight was justified in light of the character of justifications surrounding him. From the perspective of history, the war seemed clearly just.

Traditionalists and revisionists would both endorse exoneration, as would Renzo: this is a just war, and Jose has performed his moral duty to obey, to inquire, and to fight. (Think: a war of clear national, territorial, self-defense fought by a legitimate state.)

2. Both relative and objective justifications line up negatively: no person in Jose's epistemic position could have reasonably concluded that fighting was justified, and that view is objectively correct. Jose himself did not regard the fighting as justified but did so anyway. This category is again unproblematic: Jose's actions were unjustified.

Here traditionalists and revisionists divide on the question of moral culpability, though most revisionists might argue against punishing Jose on grounds of fairness. Renzo sides with the revisionist here: the moral error was apparent, and no one can take moral refuge in a defense of superior orders. (Think: a soldier who doubts his state's propaganda endorsing invading a neighbor but does so for the adventure.)

3. The relative and objective justifications divide: the war was locally and internationally controversial; the government and its allied news agencies provided a great deal of prowar propaganda; some individuals concluded that it was wrong to fight, others that it was a patriotic duty; individual views were mostly to follow political identification with or against the party in power. The balance of internal expert juridical and moral opinion was similar in being mainly a product of

> polarized identification, but most objective outsiders regarded the war as illegal and wrongful. Jose decided that although his duty to double-check the state's justification was triggered by the fact of the debate, there was enough apparent epistemic support around him to justify him, and so he fought and killed.

Traditionalists and revisionists split here. Traditionalists say that Jose ought not to be faulted for having followed orders, period, regardless of any later verdict of history. Revisionists suggest that Jose had a duty not to fight in these circumstances and can be faulted and perhaps punished, given good objective reason to doubt the justice of the war.

It would be ideal if Renzo's position could lead us out of the impasse. What he says is that whether traditionalists or revisionists have the correct answer depends on the specifics of Jose's epistemic position—whether Jose was reasonably in a position to see that the case for disobedience was met.[10] Otherwise, even in an objectively unjust war, Jose is justified in fighting, as traditionalists say (although not, Renzo, emphasizes, for quite the same reason).

To generalize: the problem is that it is very hard to imagine circumstances in a generally just state that provides its rationalizations to its soldiers, in which they could come to see the presumption of obedience as overcome. For every criticism voiced by war critics, the state has a response. And this means, in turn, that soldiers virtually always regard themselves as justified, simply by pointing to the systemic benefits of political order, even postwar when the mistakes of the war get out (which might include packets of lies wrapped up in others' good faith acceptance of those lies). But if this is so, then Renzo's account puts Jose in an extraordinary situation: Jose must accept that he was justified in fighting and killing in an unjustified war, even as he comes to realize that information about that injustice was available to him.

Is this the same merely apparent paradox of Razian authority, which sets short-term error against long-term benefit? It seems to cut deeper in Renzo's view, and Jose in retrospect experiences Renzo's position as paradoxical, as a matter of being both justified and unjustified: Jose justly fought an unjust war. He killed wrongly, and it was possible for him to have recognized the wrongness of his act at the time. It was his own failure to make more from the "peek" that led to this failure. This is quite different from following traffic directions or a construction code that in the circumstances are less than optimal.

The traditionalist position can avoid the paradox because Jose is justified, period—there is no demand for individual justification. But Renzo cannot avoid it because of this individual, relative dimension. Once Jose had reason

10 Renzo, "Political Authority and Unjust Wars," 355.

to question his duty to fight, he could either have regarded himself as justified, because in fact he regarded the reasons apparent to him as sufficient, or have regarded himself as unjustified. But the epistemic middle ground, where he regarded himself as still justified even though, in his own view, he did not see sufficient reason to conclude that the war was either just or unjust, is at the verge of incoherence. Compare a case with much lower stakes, where I must make a guess about some future outcome, knowing that I do not know enough to assess the probabilities thoroughly. I might decide that I have to make the guess anyway, but I do not regard my guess, in respect to its content, as itself justified. Jose's case is more serious: he is supposed to regard himself as justified in killing, even though he also (in retrospect) regards those he killed as innocent victims of a wrongful war. The idea of justification here strains at the edge of coherence.

The policing example shows this. If the hybrid standard is understood as justifying the objectively unjustified use of deadly force, then it is incoherent. The best that can be said for the *Graham v. Connor* doctrine is that it might protect police from unfair liability and thus might be necessary in order for them to have the right incentives to intervene in ongoing criminal activity. But that is just a matter of consequentialist justification and of little solace to either the police officer or the victim killed wrongfully—and it appears to be tied to the distinctly American problem of excessive police violence, especially against members of disadvantaged groups. Once we introduce the individual perspective, we can offer excuses from punishment, but we cannot create a justification unsupported by the facts.

In the case of policing, we can (arguably) afford to ignore the incoherence of the idea of individual justification because police departments and their political overseers have ways of turning up or down the level and intensity of police intervention, as well as a system of courts and disciplinary boards that can help individual officers determine the proper scope of their right to use force. War is not like that, at least for the *ad bellum* issues here. These are usually profoundly difficult, both in their international dimensions and in their domestic ones, and they are very rarely resolved. Witness, for example, the quixotic 2016 lawsuit by Army Captain Navy Smith contesting his orders to attack ISIS in Iraq, which were purportedly authorized by the 2001 Authorization for Use of Military Force, which was aimed at Al Qaeda in Afghanistan.[11]

11 The basis of the lawsuit is described in Ackerman, "How to Stop Trump Blowing It Up." The suit was dismissed in district court on the philosophically curious but legally predictable grounds that a soldier contesting unconstitutional orders lacked standing or distinctive injury as a basis for the suit and that only political institutions, not courts, can answer such questions. *Smith v. Obama*, CV 16-843 (CKK), 2016 WL 6839357.

Further, since legal positions do not determine moral positions, the moral questions are broader yet—witness the debates about humanitarian intervention as "illegal but legitimate," in the phrase of the UN Independent International Commission on Kosovo.[12] If Renzo's vote is to have bite, then there must be a basis amidst moral debate for Jose to regard himself as unjustified. But since all wars beyond direct homeland defense are deeply contested, it is hard to imagine situations that do not simply collapse into the traditionalist/Tennysonian position on justification—that "theirs is not to reason why."

3. BLAME MAKES IT WORSE

I do not mean to suggest that soldiers should never regard themselves as having made difficult but reasonable choices to fight, even in the wake of history's negative verdict on their cause. I do, however, mean to contest the promise of righteousness in these cases that Renzo offers. In the last section, I discussed the problematic notion of justification as applied to individuals and suggested that the semi-Razian approach to authoritative justification makes the concept unstable because the situation of a soldier in the midst of fraught debate cannot be treated as simply justified or unjustified, lest we lose sight of the fundamental tension between individual and objective perspectives.

Perhaps you are comfortable with the idea of an individual being justified in killing relative to his perspective, though unjustified objectively. If so, the character of blame may make my concern more vivid. We need to consider, I suggest, that many of these situations have the character of an Antigone-like tragic situation: Jose faces both a duty to obey and a duty to resist, neither overriding the other. This tragic dimension of individual political membership (perhaps like the so-called "dirty-hands" duty of leadership) may simply be an irreducible fact of life even in a generally legitimate state so long as that state exists in a confusing and unjust world. We disguise it by pretending that the duties can be ranked and thus reconciled, as Renzo does. But they cannot be.

There is a hint that Renzo is aware of this possibility, reflected in the often passive and acontextual language in which he frames his position:

> My argument is that when a legitimate state wages an unjust war, we are under a duty to comply with its order to fight *unless* the presumption in favor of obedience *is rebutted*.[13]

12 IICK, *The Kosovo Report*. The commission concluded that the NATO intervention in Kosovo was illegal because a Security Council resolution had been vetoed by Russia but legitimate because it was sustained by a regional coalition and in defense of human rights.

13 Renzo, "Political Authority and Unjust Wars," 351 (emphasis added).

And:

> They are morally innocent not simply in the sense that *they are blameless* but in the sense that *they are justified* in so fighting.[14]

Such impersonal language rings natural in a philosophy article, but it is misleading in an account of actual moral reflection and judgment. The best way to see this point is to return to our narrative frame, abjuring the perspective of the philosophical writer. That is, instead of asking questions of blame and permissibility from the point of view of a Walzer or Fabre, for example, let us ask the question from, say, the perspective of a villager whose corrupt and undemocratic government has agreed to the presence of Jose and his fellow soldiers or who simply has suffered in a war that, from her perspective, has always been an obvious assault on her land. Speaking of Jose's justification or blameworthiness from this perspective seems different because it reflects the specific, individual claims of those who suffer from the unjust violence, not just the abstract judgments of third-party observers.

There is an obvious analogy to cases of downside risk or to the moral cases Bernard Williams famously labelled *agent-regret*.[15] When we take a gamble that is rationally justified in advance by the information available, and that gamble turns sour, we may find some solace in its intrinsic reasonableness. But we regret having been stupid enough to take the gamble in the first place, especially when the loss is severe. *Ex ante* rationality cannot cancel out *ex post* loss, at least for human beings with limited resources and finite time horizons. And when there is a moral cost to our acts—when we cause an accident even though we could not anticipate a darting child or a slippery road—the recriminations are even more severe. The philosophical story we might tell about culpability always stays at a distance from the moral reactions that are a necessary part of human agency. As Williams remarks, while punishment or harsh judgment—certainly by bystanders—will be inappropriate, there is nothing inappropriate from the first-person perspective about offering apologies and regrets. Nor should we insist that victims accept the impersonal perspective either.[16]

Let us imagine Jose's encounter with a villager widowed by his actions or whose house was destroyed. Call her Hana. By hypothesis, Jose can say, "You realize, my state was generally legitimate, my president ordered me to go, and when I read the papers, I saw that we were invited by your government. I realize now that the conflict was founded on lies, that the inviting government was

14 Renzo, "Political Authority and Unjust Wars," 354 (emphasis added).

15 Williams, *Moral Luck*, 20–40.

16 I discuss the relevance of nonblameworthy moral regret in *Complicity*, ch. 1.

corrupt, but I did not know that at the time." But what comes next? Can he say to Hana, "So while I sympathize with your loss, I must insist that I was rationally justified at the time, and so I bear no moral culpability here, much less liability to punishment. You will need to direct your objections to my president"?

I think the answer is clearly no—and not because I have manipulated rhetoric to make such a response seem heartless.[17] The problem, rather, is that such a response would recognize only one side of Jose's agency—namely, his role responsibility as a soldier. Being a soldier does not absolve him of his further humanity. Such a response would fail to acknowledge the agency behind Hana's loss by treating it as a collateral harm to the task of having legitimate but fallible political institutions. And it would fail to acknowledge the first-personal responsibility that Jose must feel at the fact that his actions were a component of the causal storm that overtook Hana's life. Hana might well feel sympathy at the same time for Jose's position, and she may even choose to forgive and to urge him to forgive himself. But the very fact that forgiveness is relevant here underlines the fact that there is something to forgive, a moral residue of the authorized act. Even the *pro tanto* version of the Razian view makes no room for this because it looks at justification only as an exercise in the space of reasons, not as an interpersonal relation. It makes no space for tragedy.

4. Where the *Pro Tanto* Model Works: *In Bello* Justification

I have suggested that the Renzo model of justification is unlikely to work for plausible cases of *ad bellum* debate. But this is to sell it short in one respect. As my analogy to police use of force suggests, the model offers promise with questions about justification that are more easily resolvable both *ex ante* and immediately *ex post*, when individuals can correct their future actions be reflecting on good-faith errors. This is the situation of *in bello* conflict, where indeed, something like the model is already structurally incorporated through the "obedience to orders" defense.

On Renzo's view, recall, Jose has a duty to obey all orders that are not manifestly illegal—that is, orders that would be regarded as illegal by a person of ordinary sense and understanding. Compare, on the one hand, a (mistaken) order to shell a building that is known to have served before as a hospital but is now said to be used also as a weapons cache with, on the other hand, an order to shell a facility that is known to be used now as a hospital. In the latter case, Jose would commit a prosecutable war crime if he does not reject the order, as

17 We can be rightly suspicious of this technique in philosophical argument. But I think here its role is defensible.

he would be expected to do, given that anyone in his position would have been taught rules of engagement and principles of LOAC relevant to the context in which he operates. That standard is fair because he can be expected to have mastered the relatively clear standards of *in bello* conflict and to apply them against the presumption of obedience.

What about the former case, of an order to attack a purported dual-use facility? Jose would be permitted and indeed duty-bound to follow the command provided that he does not see anything to contradict the intelligence that weapons are present and that, at least for US armed services, legal officers have agreed that the military advantage was sufficient to offset the collateral deaths of hospital patients. Assuming these conditions are met, Jose would be exonerated even if it turns out that the target is in fact only a hospital, and he could be court-martialed for refusing the order unless it is on the basis of a good-faith belief that the intelligence is false. Jose would of course feel moral regret for the deaths he causes—indeed, he would likely feel regret whether or not the order was factually mistaken—but he would not likely feel the sort of guilt that I suggested would be present in the *ad bellum* case. The Renzo model here works well both in theory and in practice.

A lesson one might extract, as some revisionists have, is that we should try to make *ad bellum* rules as clear as *in bello* rules.[18] And certainly some international lawyers think we could prohibit almost any use of force that is not an immediate defense of one's own or one's allies' territory of military facilities.[19] Renzo's model of justification would work better with bright-line rules that can be internalized as easily as rules of engagement and the Geneva Conventions, and for which a jurisprudence exists. But for a variety of reasons, this seems to me highly unlikely. First, strong states—and therefore international law—will continue to claim discretion over uses of force far outside their territories in retaliation, prevention, and humanitarian intervention. Preserving the ambiguity of international law and principle over the use of force is key to their power. Second, the veto privileges of the permanent members of the Security Council mean that its dictates are unlikely to mirror general consensus, even in the clearest of cases, such as the Russian war on Ukraine. Third, the inherent vagueness of the concept of self-defense as applied to national interests means that justifications can almost always be sought for putatively defensive wars.

18 See, e.g., Rodin, "The Moral Inequality of Soldiers." Rodin proposes an international tribunal to resolve *ad bellum* issues against individual soldiers, which would require similar clarity.

19 This is the hope expressed in Hathaway and Shapiro, *The Internationalists*. They explore and defend the project of the Kellogg-Briand Pact to make war illegal except under extraordinary circumstances.

These ambiguities mean that no individual solider of a generally just state is ever likely to have clear enough rules for the standard to have any bite.

5. A HALF-HEARTED ENDORSEMENT OF THE TRADITIONALIST VIEW

I worried above that the difficulty of overcoming the presumption of obedience means that Renzo's view collapses into traditionalism in practice. But would that be a bad thing? Renzo says he prefers his view because it does not require belief in a distinct and less demanding political morality of violence, as he thinks traditionalism requires. But the view that political violence is justified differently from individual, interpersonal violence need not be as mysterious as that, even if Walzer's presentation in *Just and Unjust Wars* might lead one to think otherwise.

Certainly, one way of endorsing traditionalism is very close to Renzo's own view, and that is to see the "war convention" (as Walzer calls the impunibility of line soldiers who fight lawfully) as simply justified in consequentialist terms. I mentioned above the advantage it is thought to provide in giving individual soldiers a reason to adhere to *in bello* rules, so that they can be protected under POW status. Other reasons, specific to war, arguably include a greater likelihood of a peace treaty, because the issue of individual *postbellum* justice is off the table and the increased deterrence of international uses of force provided by a system that can more reliably call up soldiers to defend against them, since those soldiers will not fear post-conflict reprisal. These may all be specious, but they are not very far from Renzo's own view, which insists on the Kantian value of good international order, even if that Kantian value is not itself exhausted by the gains it provides to aggregate welfare. So Renzo's desire to distinguish his account from consequentialist traditionalism probably needs more justification than he provides.

However, Renzo is of course correct that Walzer, in particular, grounds his view not in welfare but in the distinctive value of loyalty as constitutive of a political community. We can call this the Romantic view, rooted as it is in a nineteenth-century conception of the importance of nationalism, as expressed by Johann Gottfried von Herder and others.[20] The view makes supremely important collective self-determination and expression, and it demands the subsumption of individual moral personality. A view that treats politics as purely instrumental, as a Razian liberal view does, cannot make much sense of these values except to see them as individual interests protected by an international system that generally protects self-government. This is to say that

20 See, e.g., Berlin, "Alleged Relativism in Eighteenth-Century Thought."

it cannot make sense of a people's moral interest in acting against perceived threats, except insofar as the system of liberal states is threatened. Walzerian traditionalists, by contrast, want to grant absolution even to law-abiding Wehrmacht or Russian soldiers, not because their obedience furthers liberal values but because they are properly subsuming their identities to their illiberal states.

Most people—certainly most liberals—no longer regard state membership as valuable as such. Membership in corrupt, pathological states has no redeeming value. But we should recognize two strands of Romantic, collectivist thought that do play anchoring roles in political morality for at least moderately decent if not liberal states. The first is how our actual moral and legal systems tolerate violence visited on persons as justifiable in a substantially different manner than we do from a purely individual, moral framework. Revisionists challenge this at one level, even though they (mostly) accept a system of individual excuses from punishment for wrongful killing that they reject in ordinary, interpersonal cases. In this, they are conceding to how we view almost all goal-directed political history, including revolutions, civil wars, and territorial wars, where many of the interests at stake, if they have an interpersonal analogue, come up short against the toll of death and destruction. A traditionalist who incorporates this different way of viewing and weighing violence need not be blood thirsty—only realistic in recognizing the lived moral and political values of most people over most time.

A second reason for recognizing the force of a distinct, political perspective on soliders' duties is to see that they rarely reflect in purely individual terms (What are my duties?) but instead reflect in terms of "we" (What should I do as part of my membership in this army and this political community?). Raz himself endorses a highly individualistic liberalism, and so in his model, an agent asks simply, "What should I do, given norm *X*?" But even with a correction, we do not end up with a collective, political view of the rights and permissions of warfare. This is because the collective aspect of the Romantic view underlies the permission structure of war. Just as the Razian view does not make room for the we, the Romantic view has trouble making room for the reflective, individual citizen half of a citizen-soldier.

I mentioned in my brief discussion of blame that many situations of war have an Antigone-like, tragic structure. Jose is caught between the demands of personal and political morality, between liberalism and Romanticism. This is why loyalty remains both prized and suspected in modern societies. Renzo's view attempts to make sense of that dual capacity, but unless it can take on the kind of inherently collective values at the root of national belonging, it provides a traditionalist resolution of the problem of obedience, without the moral ontology to support it. This brings us back around in a circle. We should

celebrate Renzo's attempt to capture the insights of both traditional and revisionist views. But the middle ground he sketches seems neither individualistic enough to capture the moral horror of fighting a wrongful war nor collective enough to support the moral comfort it offers its soldiers. We may need to settle for a less than satisfactory resolution.

University of California, Berkeley
ckutz@berkeley.edu

REFERENCES

Ackerman, Bruce. "How to Stop Trump Blowing It Up." *New York Review of Books*, November 28, 2017.

Berlin, Isaiah. "Alleged Relativism in Eighteenth-Century Thought." In *The Crooked Timber of Humanity*, edited by Henry Hardy. 2nd ed. Princeton University Press, 2013.

Fabre, Cécile. *Cosmopolitan War*. Oxford University Press, 2012.

Hathaway, Oona A., and Scott J. Shapiro. *The Internationalists: How a Radical Plan to Outlaw War Remade the World*. Simon & Schuster, 2017.

Independent International Commission on Kosovo (IICK). *The Kosovo Report: Conflict, International Response, and Lessons Learned*. Oxford University Press, 2000.

Kutz, Christopher. *Complicity: Ethics and Law for a Collective Age*. Cambridge University Press, 2001.

McMahan, Jeff. *Killing in War*. Oxford University Press, 2009.

Raz, Joseph. "Postscript: Rethinking Exclusionary Reasons." In *Practical Reason and Norms*. Rev. ed. Oxford University Press, 1999.

Raz, Joseph. "The Problem of Authority: Revisiting the Service Conception." *Minnesota Law Review* 90 (2006): 1003–44.

Renzo, Massimo. "Political Authority and Unjust Wars." *Philosophy and Phenomenological Research* 99, no. 2 (2019): 336–57.

Rodin, David. "The Moral Inequality of Soldiers: Why *Jus in Bello* Asymmetry Is Half Right." In *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, edited by David Rodin and Henry Shue. Oxford University Press, 2010.

Schauer, Frederick. "Rules and the Rule of Law." *Harvard Journal of Law and Public Policy* 14 (1991): 645–94.

Walzer, Michael. *Just and Unjust Wars*. 4th ed. Basic Books, 2006.

Williams, Bernard. *Moral Luck*. Cambridge University Press, 1981.

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · SEPTEMBER 2025
https://doi.org/10.26556/jesp.v30i5.4143

THE WEIGHT OF AUTHORITY IN WAR

COMMENTS ON RENZO

David Estlund

EVEN THOUGH many wars are unjust, they can sometimes be projects whose terrible success all should earnestly hope for. There we should hope that those legitimately in command have what they need at their disposal, including the obedience even of soldiers who disagree about the war's justice. When the war is wrong, the deaths are wrongful, so does this not show that obedient soldiers kill wrongly? Disputing this, one view—in some ways traditional—holds that it is not the soldiers' business "to reason why," and they ought morally to carry out the ordered fighting and killing. Each death is still a wrong, but not a wrong by the obedient soldier. Critics of that traditional view, often known as *revisionists*, argue that an obedient soldier in an unjust war commits the wrong of murder, even if the relevant superiors are in the wrong as well. (I use the term 'revisionist' narrowly here in this way.)

In his 2019 article "Political Authority and Unjust Wars," Massimo Renzo defends a nuanced version of the traditionalist view that it is not (always) for the soldier "to reason why."[1] He argues that legal directives from a legitimate state can convert lethal acts by a commanded soldier that would otherwise be morally impermissible into permissible ones. The picture is less clear than that makes it seem though, since Renzo also holds that the soldier is still "wrong" to kill in that case (342). I begin with a critique of Renzo's implication that an act can be both wrong and permissible, morally speaking. While his formulation is problematic, Renzo's main points can be captured without it. A second interpretive problem comes up, but here too it can be overcome. Renzo appears to misstate his position in several places, as if it were simpler than it is. Clearing that up allows us to see the subtle nature of his actual position by starting with that simpler view and criticizing some arguments offered in its favor—siding, in that way, with Renzo's real view. After, I hope, resolving those interpretive challenges, I raise and then address an important limitation of his argument that derives from the nature of authority: even if he is right about

1 Hereafter, Renzo's article is cited parenthetically.

political authority that it presumptively brackets certain considerations from the soldier's proper deliberation, surely some bracketed conditions can still be compelling enough that the authority lapses. From there, however, I side with Renzo (or at least what I argue is an implication of his position) in holding that the taking of innocent human life, as weighty as that is, is not enough—or more specifically, not obviously enough—to nullify political authority even in an unjust war.

1. FORBIDDEN AND PERMITTED

Here are two relatively simple positions in a dispute about following orders in an unjust war. (It is a helpful simplification to suppose that one side is just, the other unjust.)

> *Permitted Obedience*: In certain cases, a military command can render permissible (i.e., not wrong) a soldier's act of killing that otherwise would be a wrong by them, even if the military command itself is wrong and the targeted soldier ought morally not to be killed.[2]
>
> *Forbidden Obedience*: Where a commanded war and its killings would be serious injustices by the commanding state, whatever general moral reasons a citizen has to obey legal state orders, those are outweighed or overridden, and a soldier must, all things considered, not obey.

An important formal point that Permitted Obedience exploits is this:

> *Permitted Obedience's Formal Point Against Forbidden Obedience*: That the state would be wrong to kill and yet orders its soldiers to kill, rendering the killing a wrong to the victim, *does not imply that it is a wrong by the soldier*.

This is formal in the sense that it does not say anything about what is right or wrong but makes a logical observation. It is "against" Forbidden Obedience only in the sense that it denies one argumentative route to Forbidden Obedience, though there may be others.[3]

There is a second formal point, this time available to Forbidden Obedience and raising a question about Permitted Obedience:

2 I will (with notice) start calling this position *Authority* later in the article, for reasons to be explained then.

3 I can steer clear of the question whether collectives can be blameworthy, so I always mean to refer to individual agents who are involved in the decision-making in whatever way the context would require for the point at hand.

> *Forbidden Obedience's Formal Point Against Permitted Obedience:* That there is a duty to obey that is enough sometimes to render even killing permissible and sometimes to require and permit obeying wrongful orders *does not imply* that the duty of obedience is *not outweighed in the case of the unjust side in a war*.[4]

A duty to obey the state *could*, as this says, be outweighed in such cases, though it might not be. Renzo defends this formal point with a close look at the difference between two views of the reasons that arise from practical authority, arguing that they are not "exclusionary" but "presumptive." The details of that argument do not matter for my purposes, and we concentrate here on its normative upshot and some challenges to it. Broadly speaking, I defend Renzo's support for Permitted Obedience against some instructive challenges that are in a revisionist spirit. But I start with some preliminaries.

2. PERMISSIBLE WRONGS?

I normally use the terms 'morally wrong' and 'morally impermissible' interchangeably, as I believe most do.[5] Renzo, however, does not. He writes that his article's conclusion

> grants combatants on both sides a *permission* to fight (under the conditions specified . . .), while acknowledging at the same time that combatants fighting on the just side are *wronged by* those fighting on the unjust side (since the former have done nothing to lose their right not to be killed). (356)

That helps to explain why Renzo does not seem to think that it is tendentious or question begging—even while the question at hand is whether the soldier's act of killing is a moral violation—to employ, as if they were neutral in the dispute, such wrongness-implying descriptions of what the soldier does as "to commit something wrong," "acting unjustly" (343), "doing something unjust" (345), being an "accomplice" to serious injustice (346), and "perpetrating an injustice" (350). I would expect him to argue that these are not question begging at all, since a given act at a given time can, he thinks, be both morally wrong and morally permissible. The idea seems to be that there is a standing duty not to kill soldiers on the just side, and even when one is, all things considered, permitted

4 Like Renzo, I use the terms 'duty' and 'obligation' interchangeably.

5 Sophia Moreau, however, organizes things roughly as Renzo does (*Faces of Inequality*). In "Moreau on Discrimination and Wrong," I argue, along lines complementary to what I say here—that this leads her view into trouble.

to do so, one violates that duty (a *pro tanto* duty, he says, meaning that it "maintains its original force" even when it is "overridden") and thereby "wrongs" the soldier on the just side (341). The killing soldier kills wrongly—but permissibly.

I think this is unnecessarily paradoxical. Fortunately, Renzo's main normative points can be captured under a modest reformulation that drops the paradox. We can cast overridden duties as not duties at all in the instant case. Based on *pro tanto* reasons, they were only *prima facie* duties, which are not a species of duty but only what might seem at first to be duties. There is a *prima facie* obligation not to contribute to unjustified killings and a *prima facie* obligation to obey commands under certain specified conditions.[6] More fully (in my words, not Renzo's):

> *Renzo's View Adjusted*: If a soldier's state is reasonably just and legitimate, then even as its war effort is unjust, there is a moral presumption that the decision to go to war is the state's to make, which takes the weighty reasons not to kill others out of the soldier's consideration with respect to the decision whether to obey. That presumption is rebutted if the soldier, having duly inquired, not only believes but also has sufficient justified confidence that the state's order to fight is unjust.[7] (Call this Renzo's *special epistemic exception*.) Barring that, even if the soldier believes, even correctly (but not sufficiently confidently), that the order is unjust, the all-things-considered duty is still to obey. The soldier thereby participates in unjust killing, but the wrong is not the soldier's.

This formulation avoids the awkward idea of permissible wrongdoing. On my best interpretation, then, Renzo's view is more opposed to the revisionist view—the view that such obedience is a moral violation, all things considered—than might appear from his saying that an obedient soldier wrongs the victim, etc. It is, it turns out, a qualified version of Permitted Obedience.

There is a second way in which Renzo's view is not what it might seem to be. Renzo appears to endorse what I call a certain *complicity principle*, but the

6 This is consistent with W. D. Ross, who first introduces the terminology of *prima facie* duties. They are not a kind of duty but a kind of reason. *Prima facie* duties are nothing but a certain set of *pro tanto* reasons that have lasting weight in the direction of what might or might not, all things considered, be a duty of any kind. This point about Ross's view is explained by Anthony Skelton, "William David Ross."

7 Renzo states the condition as depending on the soldier being "justified in believing that the authority has made a mistake that will lead to serious injustice and we are sufficiently confident about this belief" ("Political Authority and Unjust War," 349).

appearance is misleading.[8] He says that we have a duty to "disobey if complying would lead to the perpetration of serious injustice" (349). I believe he has not stated his position carefully. That would be Forbidden Obedience, which Renzo actually rejects. His view, which is laid out at length in the article, is better conveyed by these compact formulations:

> In those cases where the epistemic threshold is not met … our reasons to act as the authority requires are undefeated, and thus we are under a duty to obey, even if the authority has made a mistake and complying will lead us to perpetrate a serious injustice. (350)

> My view is close to the orthodox account [i.e., roughly, Permitted Obedience] in that it acknowledges that there will be at least some combatants who will have an all-things-considered duty to obey the order to fight an unjust war. (354)

What explains the several statements seeming to imply Forbidden Obedience? It appears to be an artifact of the formulation I criticized earlier—that the same act can be wrong and permissible and, relatedly, that we can have a duty not to do something and be permitted or required to do that same thing. In particular, the problem stems from the problematic "wrong and permitted" formulation we have just discussed and adjusted—in particular, the very idea of *pro tanto* rather than merely *prima facie* duties. So when Renzo says we have a "duty" not to obey when that would be to participate in serious injustice, he does not (even though he calls it "wrong") take this to imply that obeying is impermissible. On his view, it might be morally permissible or even required.

Here, then, is a complicity principle that he might have seemed to endorse but does not and that does indeed imply Forbidden Obedience, which he rejects:

> *Forbidden Complicity Principle*: We are morally required never to participate in the perpetration of serious injustice.

This is a tempting and widely held view in its own right and serves as a foil for Renzo's below.

3. AUTHORITY-FAVORABLE CONDITIONS

Before considering a family of objections to arguments from political authority such as Renzo's, let us specify the substantive—not merely formal—question

8 I use the term 'complicity' in its ordinary, nontechnical sense, and it picks up on Renzo's use of the cognate term 'accomplice'. The *Merriam-Webster Dictionary* gives a typical definition: "association or participation in or as if in a wrongful act."

between Permitted Obedience and Forbidden Obedience. Some may hold that a soldier's actions are covered by Forbidden Obedience even under the following conditions but laying them out exposes several attractive aspects of Permitted Obedience. We limit ourselves to cases that meet the following authority-favorable conditions:

1. *Conscientious Authority*: The commanding state authority has done all that can be expected (e.g., taken some procedural prelude) to refrain from prosecuting war unjustly, and the mistake is, in that sense, an honest one, or in "good faith."[9]
2. *Informed Authority*: The authority has access to and has taken the measure of whatever relevant and significant arguments or evidence are available to the commanded soldier. (The soldier has no extra reasons in this sense.)
3. *Credibly Wrongful and Lethal Threat*: The enemy soldier is believed by the defending soldier and is justifiably (even if erroneously) believed by the commanding authority to be in the wrong and to be part of a great threat.
4. *Only Death Removes the Threat*: It is known by all that the only way to prevent that ostensible danger is to incapacitate many enemy soldiers and that this can be done only by seeing to it that they are killed.
5. *Permitted when Accurate*: Killing enemy soldiers to remove the threat would be permissible by the commanding state if, as the state mistakenly believes, the enemy state were unjustly in the war.

Thus, the substantive question, going beyond the two formal points is this: is it morally permissible for a soldier to obey lethal orders in war in these conditions?

4. Killing's Weight

As we have seen, Renzo argues that except in the case of the special epistemic exception whereby due inquiry gives the soldier sufficient warrant and confidence that the command is unjust, the reasons that the authority provides can prevail even when the command in question is understood by the commanded soldier to be wrong. The weight of the reasons against the permissibility of the commanded action is not to be brought to bear. That is in the formal nature

9 Renzo uses the term 'good faith' throughout (346), interchangeably with the term 'honest mistake', which I introduce in Estlund, "On Following Orders in an Unjust War."

of authority, Renzo argues. They do have countervailing weight as reasons, we might say, but they are to be left off the scale.

However, even if this is correct about the structure of authority, that cannot by itself show that a command from a legitimate authority can be morally binding, all things considered, regardless of how severe the wrong would be that one would thereby be contributing to. Renzo's epistemic exception gives no antecedent weight to the severity of the wrong—only to warrant and confidence in the soldier's belief that it is wrong. What if the allegedly bracketed consideration was instead the fact that if the soldier were to obey, they would be contributing to the imminent collapse of civilization? Renzo's view suggests that even if there were little risk or cost from disobeying, unless the soldier not only believes the truth with good reason but also has some additional quantum of confidence, then obedience is required. No one should bite that bullet just because authority has a certain formal bracketing structure. Even if it is correct, that whole structure of authority must surely be within the scope of certain exceptions, at least partly on grounds of great severity. And then the revisionist can fairly ask: Why would the killing of innocent people not be within the class of exceptions (and antecedent severity threshold), thus preempting Renzo's epistemic exception requiring warrant and confidence? This is an important concern, as far as it goes.

The question is how far it does go. The central claim at issue here is what I call the Permitted Obedience principle, discussed in section 1 above. Going forward, it is useful to call the principle the Authority View.[10]

> *The Authority View* (hereafter *Authority*): In certain cases, a military command can render permissible (i.e., not wrong) a soldier's act of killing that otherwise would have been a wrong by the soldier, even if the military command itself is wrong and the target ought morally not to be killed.

Authority implies that someone may, in specific circumstances, permissibly kill or injure another on the basis of a command known to be erroneous. That, as stated, intuitively rankles. My modest aim is not to fully address arguments against Authority. Rather, it is to show that even though it has intuitive costs, so does the salient alternative. Authority might, given the choices, be a reasonable position in spite of its intuitive cost. The relevant more restrictive alternative view might also be a reasonable (and questionable) position in light of its cost, given the intuitive cost of Authority. If an intuitive price must be paid on either

10 This is because the question of authority takes center stage and because its foil is no longer Forbidden Obedience as before but what I will shortly define as the Restrictive View, which is more specific.

view, neither view decides the case in its own favor by trumpeting the other view's particular intuitive cost.[11]

A first step is to challenge a natural and more specific version of the thought that authority cannot be enough because the following may seem obvious: nothing can have the weight to justify inflicting death and injury except countervailing appeals to death and injury of certain kinds.

> *The Restrictive View* (hereafter *Restrictive*): Given the great moral weight of killing or severely injuring someone, nothing but appeal to death and injury can weigh in its defense, such as legitimate self-defense/other defense against death or injury, or the attacker's guilt or responsibility for contributing to wrongful death or injury, or the prevention of a lot more death or injury down the road.[12]

That last clause is ambiguous. It would seem especially in the spirit of the rest if it meant that killing or injuring that one targeted soldier would prevent those later bad consequences. It could be that for some set of combatants, while each of them is more likely than not to kill no one, it is virtually certain that between them, they will kill many. For example, to take a number we will come back to, it could be that each soldier only has a .45 chance of killing anyone, but a set of one hundred such soldiers will almost certainly kill dozens of people. So in a relaxation that is favorable to Restrictive in the present dialectical situation, suppose it allows that a soldier may be targeted not only if they are themselves liable or an imminent threat but even if they are (in certain ways) merely part of such a set whose removal (but no smaller subset's removal) from the battlefield would prevent a lot more death and injury down the road.[13]

11 I am not suggesting that critics of the Authority sort of view never say more than that it violates the strong intuition about the moral weight of killing. That is far from true. But more fulsome arguments do seem to me often still to bottom out in appeal to roughly that same broad intuitive assumption. I cannot take up a full assessment here, so I settle for calling attention to that possibility in light of what I argue here would be its dialectical implications.

12 This is not meant to reflect every view that is called revisionist, which is a family of views rather than a single view. See Lazar, "Just War Theory." Rather, this is meant to be what is at first a highly intuitive position as stated, based on the moral weight of death and injury. If it has counterintuitive implications after all, revisionists can consider whether it can, still in a principled rather than ad hoc way, be relaxed to remove the conflict. It bears pointing out that in doing so, they absorb one intuitive cost (death and injury may be permitted by considerations having nothing to do with death and injury) to avoid another, seeking, in effect, what is, in light of additional arguments, intuitive on balance.

13 Helen Frowe treats "contributors" to a threatened harm as liable to defensive harm in certain cases (*Defensive Killing*). But as I understand her view, "contributing" does not

This kind of exception to limiting permissible killings to liability or self/other defense is widely called a "lesser evil" provision. It can reasonably be asked how it is still in the spirit of Restrictive. It might be replied that the goal is a broader one than it might seem at first: to let considerations make only the same kind of moral difference they would make in ordinary non-war contexts, and the (even broad, aggregating) lesser evil provision may pass that test. It is also still in the spirit of Restrictive insofar as still it does not count anything other than causes, contributions, or complicity around harms to persons.

The following position, however, conflicts with Restrictive even so construed:

> *Due National Defense*: Sometimes intentionally killing an attacking soldier on an unjust side is permissible even if that soldier is neither relevantly responsible nor relevantly dangerous (imminently or down the road), if the attack is in defense of the survival of a just attacked society's basic sovereignty, and by way of an appropriate structure of command.[14]

Let "relevantly responsible" mean that the attacking soldier is either culpable or at least relevantly responsible for fighting on the unjust side, and let "relevantly dangerous" mean that they either represent an imminent threat of death or severe injury to someone or are a linchpin whose—or on the less restrictive version yet, are part of a set whose—removal from the battlefield might plausibly prevent a much greater evil of deaths and injuries.

Restrictive must deny what I expect many people find very implausible to deny, as illustrated in the following case. There are available credibly reported numbers about the Russia-Ukraine war that highlight the intuitive attraction of Due National Defense. But the numbers are constantly evolving and have only limited reliability anyway. For that reason and also for one other, I concentrate instead on a similar but hypothetical case of the countries Nukraine and Crushia, stipulating that the cited numbers and assumptions are accurate for them. The most important difference in the hypothetical case is that we assume that the vast majority of those fighting for Crushia are forced to do so in ways

---

cover non-difference-making agents but rather means those affecting circumstances so that they result in a direct threat from a person who will harm someone.

14 I do not mean to suggest that an unjustly attacked country is morally permitted to do whatever is necessary to save itself. McMahan is surely correct that "Ukraine and its supporters are . . . also bound by moral constraints . . . that Russia can exploit to its advantage, perhaps making it impermissible for Ukraine to fully achieve all of its just aims" ("Just War Theory and the Russia-Ukraine War"). Other authors have discussed issues around whether views roughly like Forbidden Obedience and Restrictive have implausible implications about national defense. See, for example, Rodin, *War and Self-Defense*; and Frowe, "Can Reductive Individualists Allow Defence Against Political Aggression?"

that prevent them from being culpable or otherwise responsible for fighting on an unjust side. There is evidence that many of those fighting for Russia in its war against Ukraine are forcibly conscripted and unable to refuse to fight, but it is unknown what fraction this might be.[15] We put these questions aside by moving to the case of Crushia and stipulating that most have unjustly been given no reasonable choice but to fight.

Recall that I assume that Restrictive permits killing a soldier if the soldier is, in certain ways subject to certain conditions, a part of a set of soldiers the killing of whom is sufficiently likely to prevent the evil of much greater death and injury. But let us add the assumption, again not unrealistic, that if Nukraine thoroughly obeyed the ethics of war according to Restrictive, there would be no such prospect of a great evil of death and injury perpetrated by Crushia.[16] That said, let us import into our hypothetical example some numbers similar to what has actually been reported, with some simplified rounding in a way that does not skew them in my favor. According to some estimates of the deaths and severe injuries to soldiers and civilians on the Ukrainian side, as of early 2025, rounding up for our purposes, the number may be put somewhere around 450,000.[17] How many Russian combatants have done that much harm? The total number of combatants on the Russian side over the period of the war is surely more than the number of Russian military deaths and severe injuries,

15 See Schwartz and Gorenburg, *Russian Military Mobilization During the Ukraine War*.

16 The case of "bloodless invasion," where no blood is shed if there is no resistance, is widely discussed to a similar purpose. For a recent example, see Lazar, "National Defence, Self Defence, and the Problem of Political Aggression." Some discussions focus on a question I do not address: how to take moral account of additional violence triggered by violent defense.

17 Let the relevant period of war in the hypothetical case correspond to the period from roughly early 2022 to early 2025. See Croft, "More than 46,000 Ukrainian Troops Killed Defending Against Russian Invasion, Says Zelensky." (To be clear, the article's title estimates troops "killed," which is not the question for which I am citing the article.) The article cites reports last year that suggest from 200,000 to over 300,000: "Last year, *The Economist* reported anonymous US officials as estimating that at least 70,000 Ukrainian soldiers had died and up to 120,000 wounded. In October, another US official estimated to the *New York Times* that more than 57,500 had been killed and 250,000 wounded." For a rough estimate for early 2025 and to further round up (against interest), suppose soldier casualties (death or injury) from the battlefield on the Ukrainian side are about 350,00. As for civilian casualties on the Ukrainian side, see Statista Research Department, "Number of Civilian Casualties in Ukraine during Russia's Invasion Verified by OHCHR from February 24, 2022 to May 31, 2025": "46,085 civilian casualties . . . as of May 31, 2025." Rounding up a lot, against my interest, suppose this is 100,000. Added to the 350,000 assumed Ukrainian soldier casualties, this gives us a working assumption of 450,000 Ukrainian military and civilian deaths or injuries. These empirical observations guide our hypothetical example of Nukraine and Crushia.

and the latter has been reported as around one million.[18] Set the estimate for total Crushian combatants at one million or more.

On those suppositions, the expected killings or severe injuries of combatants or noncombatants by a Crushian combatant over the course of a violently contested war comes to something less than .45. Now, if you were confronted with an aggressor with a known .45 chance of intentionally killing or severely injuring you, lethal force may or may not be permitted in self-defense by the appropriate standard of sufficient certainty of the threat. Suppose that it would be. But we are supposing the Crushian soldiers are mostly forced to fight and kill, which raises further doubts about whether self-defense against them is permitted.[19] But even bracketing all that, Crushian soldiers are permissibly killed or injured by Nukrainian forces in self/other defense only at such time as the Crushian soldiers are an *imminent* threat—roughly, reasonably likely to directly kill or harm more or less immediately. Assuming that no such limits on targeting are in place in practice in Nukraine's war against Crushia (as they are not in Ukraine's war against Russia), we assume that any combatant is, in practice and erroneously, treated as a legitimate target at virtually any time (subject to other rules of *jus in bello*). Suffice it to say that, with that .45 chance of killing or injuring anyone, the average Crushian soldier in combat is likely to be an imminent threat to life and limb only a tiny fraction of their time in combat. The majority of Crushian military casualties inflicted by Nukrainian soldiers is thus mostly if not overwhelmingly impermissible. Almost none of the cases on its own is a permitted killing according to Restrictive. Given those not unrealistic numbers, so long as Crushia's soldiers are forced to fight, Nukrainian soldiers are effectively required to surrender their country to Crushia without a fight. I am claiming no more than that, for many, this implication counts intuitively against Restrictive View, just as Restrictive View may have intuitive appeal that counts against Due National Defense.

Even if it has weight against Restrictive's plausibility, the national defense case might not seem helpful to the case for Authority in particular, since it

18 The number of Russians in combat roles over the course of the war is surely greater than their military casualties, for which estimates in the first part of 2025 range from 750,000 to one million. I conservatively estimate the total number of combat soldiers at one million or more, though it could well be much more, which would favor my argument. According to *Nesweek*, "more than 800,000 Russian soldiers have been killed or severely wounded" (deJong, "How Many People Have Died in the Russia-Ukraine War?"). See also Edwards, "Russia Nears 1 Million War Casualties in Ukraine, Study Finds." The text clarifies that these do not include civilian casualties but "Russian soldiers . . . killed or injured." See also Office of the Director of National Intelligence, *Annual Threat Assessment of the US Intelligence Community*: "750,000-plus [Russian military] dead and wounded" (22).

19 For more on these matters see, Frowe and Parry, "Self-Defense."

seems only to introduce a kind of lesser evil justification for avoiding bad consequences other than death or injury—perhaps framed in terms of massive violations of individual rights to participate in collective sovereignty, for example—which can permit targeting innocent nonthreatening combatants.[20] That possible further relaxation of Restrictive raises important questions. Not only does it jettison the focus on personal harms; it is also far from clear that the test of not going beyond ordinary interpersonal morality is being respected. Where in ordinary interpersonal ethics are we to look for a similar significance for collective sovereignty? And even if there were a good answer, what prevents the relevant moral weight of authority from also meeting that test of being found in nonwar contexts, such as in the case of a legal system of punishment?

However, letting those concerns pass, consider, with an eye to the question of authority, the question whether a person acting as a vigilante—either a nonsoldier or a soldier without any relevant orders—may also kill innocent Crushian combatants on those expansive "lesser evil" grounds around national defense? It may be that defending oneself or others from death or severe injury does not depend on an official command structure, but that is less clear in this case of defending a society's independence and basic sovereignty, where no death or injury is at stake. One reason might be that in that case, the society or state should get to decide whether a violent defense is to be mounted or not. If that is right, and a vigilante Nukrainian partisan would not be permitted to kill or severely injure the same Crushian soldiers that duly commanded Nukrainian soldiers may kill or severely injure, the fact of following the official command would make a signal moral difference. In the circumstances, for the Nukrainian soldier, the ground of their act's permissibility, even given the moral gravity of killing an innocent and nonthreatening person, is, in short, that they are following orders.

This is enough, or so I believe, to put into some question Restrictive (which is far from refuting it, which I am not attempting)—and on grounds that are congenial to Authority. But so far, of course, this is not fully the case we want to consider. In this case, the defending side is correct in thinking that standards of defensive *jus ad bellum* are met—especially that there is in fact an unjustified effort to invade and annex the country. By contrast, our topic concerns a side

20 Lesser evil justifications in the ethics of war normally concentrate on avoidable harms to persons, though general statements of the idea normally do not build that in as a constraint. Helen Frowe, for example, writes, "Lesser-evil justifications invoke the disparity between the harm that one will inflict on an agent, and the harm that one will prevent by harming her" ("Lesser-Evil Justifications for Harming," 460). She then goes on to consider only preventable harms to persons. But we can easily understand a broader idea of lesser evil as well, such as committing a wrong to prevent far greater wrong, though not necessarily harm.

that is mounting a war in defense against what it reasonably but mistakenly takes to be an effort at unjust annexation. So we suppose that the principles of *jus ad bellum* are not in fact met. What we carry over from the discussion above is the crucial point that because the answer to the vigilante question (in the nonerror case) is not obvious, it is not obvious that Authority is false—that authoritative commands can never convert from wrongful to permissible what would otherwise be a wrongful killing of a relevantly innocent person. Allowing that authority might have that perhaps surprising moral weight in some cases, we now turn specifically to the case of erroneous command. In changing over now to that error case, we can drop the assumption that the attacking soldiers are nonliable and nonresponsible, since we want a case in which the defensive counterattack, which is objectively wrong, *would be* permissible if conditions were as the defending state reasonably takes them to be. So I now step back from the case of Nukraine and Crushia.

Here is an example designed to have all the elements we need:

> *Credible Hapless Invaders*: Suppose invading soldiers are willingly carrying out a war of aggression, and the invasion begins as a patently unjust but nonviolent incursion with infantry, tanks, and drones. Commanding agents on the attacked side have significant reason to believe that without defensive military response, they will lose their country. Suppose, though, that this is incorrect, and the invading country, even if not resisted with violence, would find itself too incompetent or ill supplied to remain for long. Any deaths or injuries ensuing from a violent defensive attack would be wrongs by (at least) the commanding agents, if perhaps excusable.

Consider, now, a soldier who is commanded to use lethal force against the invaders but who disagrees with the commanders and believes (correctly) that this defensive war effort is unnecessary. If the command lacks justifying authority because it is erroneous, in which case those enemy soldiers in fact ought not to be killed, then by obeying, the soldier commits murder morally speaking.

There are two ways an anti-authority advocate might try to soften this harsh verdict under pressure from contrary intuitions, but neither of them accomplishes much in the end. The first way to accommodate that thought is to say that the erroneous command is still wrong to obey, but the command partly excuses compliance—the nonvigilante soldier is still wrong but less blameworthy than the vigilante would be. But what might the excuse stemming from being commanded be? As always, if disobeying is very risky or burdensome, that might be excusing, but a command itself does not impose any obstacle, cost, or burden for disobeying. If the soldier reasonably believed that the chain

of command is their best evidence about the justice of the defensive war and so of the legitimacy of this target, then their nonculpable ignorance would be an obstacle out of their control and potentially excusing. But our example is defined by the assumption that there is no such ignorance and so no such excuse available to the soldier. It is unclear, then, what excusing but not justifying circumstance is generally available to soften the implication for the anti-authority view that the soldier following the good-faith erroneous command is guilty of murder and fully blameworthy for it. Some are comfortable with this, but that, of course, is not an argument that one ought to be.

A second way to soften the harsh verdict against the mistakenly commanded soldier is to grant that the erroneous command does have some moral force in the direction of permitting—presumably from moral force of authority—but that it is outweighed or overridden by the severity of the wrong that would result. As a result, the commanded killing is a lesser wrong than the uncommanded one. Blameworthiness is indeed lessened, not from any excuse but only from the blameworthy act being a lesser wrong. However, this tells us nothing about whether or not it is any less severely wrong than murder and indeed whether it might even still be a case of murder, leaving the harsh verdict in place. On the other hand, the permitting power of the command might in principle nearly reach permission. Pending some argument (beyond reporting an intuition) that this is not so, the authority view could grant the general point and allow (if only for the sake of argument) that the killing would be somewhat more wrong and blameworthy if there is no command. This leaves the dialectical situation unchanged. Neither the possibility of excuse nor of lessened degree of wrong is enough to show that the duly commanded soldier is not at least nearly as wrong and blameworthy as if there had been no command at all. Some are comfortable with that, as we know, but that difference from those who find that implication highly implausible is just that—a difference, not an argument against them. I return to this point in the next and final section.

5. Intuition and Argument

After clearing up some terms and distinctions about duty and wrong, I have argued in defense of what I think Renzo's view implies—that the moral severity of someone being wrongly killed does not obviously trump the moral authorizing weight of properly conducted commands. He may well be right that unless a soldier is in some further special epistemic situation, they are not permitted to disobey orders simply because their best evidence suggests the command is mistaken. They should (and so at least may) impose those great but erroneous burdens as commanded. I have not insisted that this or any other

argument from authority prevails over contrary views like Restrictive. Rather, the modest aim has been to guard against a temptation to exaggerate the force of objections anchored in the great moral weight of killing an innocent and nondangerous person. In this final section, I pull together some main points in a summary way, now with an eye especially to my methodological subtheme about competing intuitions.

It is helpful to introduce an analogy between a jailer and a soldier, both of whom receive commands from agents participating in legitimate good-faith procedures, to impose terrible burdens.[21] Each believes with reason that the orders are mistaken, aimed at people who ought not to have these burdens imposed. In the jailer's case, assuming a fair and procedurally legitimate trial and assuming that legally available punishments are morally appropriate for the crimes in question, it seems to many people (which is not to say they are correct) that a jailer ought not to set a convict free (if they somehow could) because the jailer believes the verdict is mistaken. As it might be commonly expressed, the jailer is not to, in effect, appoint themselves judge and jury as well as jailer. Even if the verdict is mistaken, the error is, morally speaking, the judge's and/or the jury's error. Those agents might be morally wrong to convict and sentence the defendant (even if this might be excusable given their best efforts). But, so this common view says, the jailer would be misusing their position if they were to disobey and preempt the judgment of the court and free the convicted person. It is fair to suggest, then, that similar reasoning should apply in the case of a good-faith procedure that yields a military order, eventually reaching the soldier, to engage in lethal violence to defend against what is reasonably thought to be unjust aggression. This common view embeds an argument from authority.

The cases might be distinguished by arguing that the jailer has this duty to obey only if the trial procedure is not only fair and so on but also so accurate, at least in cases like the present one, that the best evidence overall is that, contrary to the jailer's initial view, the convicted party is guilty after all. That is a fair distinction, but it may not satisfy the purveyor of the analogy. The view of many, I suspect, is that it is simply not morally up to the jailer to act on their own assessment of guilt, and this applies even if their overall evidence, even after factoring in whatever epistemic value the trial procedure has, is that the person is in fact not guilty.

21 I have discussed this analogy previously in "On Following Orders in an Unjust War." For my purposes here, I follow Renzo in not relying on any particular account of the ground of political authority such as the one I sketched there, having to do partly with the potential epistemic value of a well-functioning democratic decision to go to war.

This reasonable view has a form that Renzo's view of the weight of authority purports to vindicate: the considerations for or against the imposition of the burdens are, in effect, morally removed from the commanded party's legitimate deliberations and remain the business of the commanding deciders. The parallel in war is a legitimate procedure for a society's decision to go to defensive war, conducted in good faith, and concluding that defensive war is necessary and justified. The military order reaches the soldier; in light of all of his evidence, including whatever epistemic value the decision procedure has (though neither Renzo's view nor my present argument relies on any assumptions about that), the soldier's evidence might still, on balance, indicate that the defensive war is wrong because, say for reasons sketched above, it is actually unnecessary. But even so, on this view, it is not the soldier's legitimate place to preempt the society's good-faith decision simply because the solider judges it to be a mistake.

Now, as against this argument from authority, others might find it morally implausible that either the jailer or the soldier is morally permitted, simply by the fact of good-faith command or even the supposed bracketing structure of practical authority, to jail or even kill an innocent person. That is also an understandable position. However deeply that conviction is held, it cannot be treated as obvious if that means any right-thinking person would agree. Each side is only as obvious as the opposing side is absurd, which is to say, not very. Renzo does grant, however, that it would go too far to say that the jailer or soldier may obey even if their belief is not only supported by the best epistemic reasons but in addition obvious to them. He carves off a special epistemic exception whereby the soldier or jailer not only believes, all things considered, that the order is a mistake but also has an objectively warranted extra degree of confidence about it—some significant degree of warranted confidence over and above what is entailed by believing it. In simple terms: if, despite a procedure's good-faith efforts, it has patently made a mistake, then the procedure's authority may lapse, and the solider/jailer is permitted or even required to disobey.

Restrictive must deny this overall position. What forbids the jailer and soldier to follow the orders is simply whether the orders are in fact mistaken or not; the best they can do is assess the overall balance of evidence. Authority, which says by contrast that this would be an abuse of their position, is not compatible with that. Restrictive might ask how obviousness, or warranted confidence, can be allowed to make this difference unless the aim is actually to simply figure out as best one can whether the command is correct or not. I do think this feature of Renzo's view requires some explanation, but there are a number of ways to provide it. I cannot speak for Renzo, but one explanation might be that the only considerations that can permit disobedience are

ones that the commanded agent has reason to believe that the procedure did not consider (maybe because it was not available to them), *plus*, if they had been able to consider them, they would have had no justification or excuse for making the decision they did. Short of that, even if the commander or judge/jury is mistaken, it remains their decision to make, and the soldier or jailer is not permitted to take over their role. There might be other ways to explain the nature of the special epistemic exception to similar effect.

With the special epistemic exception built in, Authority is neither obvious nor absurd. From the camp of Restrictive, some will still not think authority can have that much moral weight. My modest point is that such a report has zero weight as an argument against the other side's own argument. They can themselves report that, even on such grave matters, they do not think that a soldier or jailer may interfere with a legitimate and good-faith procedural effort to implement urgent national defense or justified criminal punishment. Restrictive is itself not somehow an argument that every reasonable person should take to refute that intuition. Restrictive is, it seems to me, just the working out of certain contrary intuitions: ones that the Authority side does not share—but also, it seems to me, ones that it does not have arguments against.

My point is not the frivolous one that neither side can establish itself conclusively or with certainty. It is that it is not clear that either side even so much as favors itself—never mind conclusively establishes itself—by argument against the other side's argument. Rather, each argues against the other side's conclusion but does so based on intuitive premises that are not shown by argument to be favored over those of the other side. Doing better than this is not impossible in principle, whether or not there is a way here. In principle, one side could show that the other side's view has implications that would lead even the other side to withdraw, at least unless they are irrational or morally obtuse. Sometimes views similar to Restrictive in particular are thought to have that kind of force: upholding the clarion moral principle that, since war is not special, innocent and nonthreatening people may not be killed (well, except as a lesser evil of death or injury … so, clarion-ish). I have argued that this would seriously misrepresent the dialectical situation. That principle is not shown to be relevantly clarion—clear to all right-minded people—unless no right-minded person could think that such killing is sometimes permitted, for example in Due National Defense. But that further thing is not offered any support; the principle, rather, is merely reported to have some intuitive force, which is hard to deny but also entirely different.

I anticipate this response: there are admittedly good and hard philosophical questions for both Authority and Restrictive, but in the face of reasonable uncertainty, with life and death at stake, surely a soldier should err on the side of

caution and refuse to obey.[22] Whether or not that is so, that putatively cautionary approach changes the subject. It does not offer any help with the question to which Renzo's article and this comment are primarily addressed and which the morally serious soldier (and jailer) would care greatly to answer. While addressing the different and fine question of what one ought to do if one does not know what one ought to do (so to speak), it declines to address our more fundamental first-order question of what one ought to do—that is, whether and when it is permitted or forbidden to follow orders to kill in an unjust war.[23]

Brown University
david_estlund@brown.edu

REFERENCES

Croft, Alex "More than 46,000 Ukrainian Troops Killed Defending Against Russian Invasion, Says Zelensky." *Independent*, February 18, 2025. https://www.independent.co.uk/news/world/europe/ukrainian-soldiers-killed-russia-putin-zelensky-b2699903.html.

deJong, Jenna. "How Many People Have Died in the Russia-Ukraine War?" *Newsweek*, February 19, 2025. https://www.newsweek.com/people-died-casualties-russia-ukraine-war-update-2033371.

Edwards, Christian. "Russia Nears 1 Million War Casualties in Ukraine, Study Finds," CNN, June 4, 2025. https://www.cnn.com/2025/06/04/europe/russia-war-casualties-1-million-ukraine-intl.

Estlund, David. "Moreau on Discrimination and Wrong." *Jerusalem Review of Legal Studies* 25, no. 1 (2022): 84–95.

———. "On Following Orders in an Unjust War." *Journal of Political Philosophy* 15, no. 2 (2007): 213–34.

Frowe, Helen. "Can Reductive Individualists Allow Defence Against Political Aggression?" In *Oxford Studies in Political Philosophy*, vol. 1, edited by David Sobel, Peter Vallentyne, and Steven Wall. Oxford University Press, 2015.

———. *Defensive Killing*. Oxford University Press, 2014.

———. "Lesser-Evil Justifications for Harming: Why We're Required to Turn the Trolley." *Philosophical Quarterly* 68, no. 272 (2018): 460–80.

22 For example, McMahan writes, "When a soldier is uncertain about the morality of a war, the presumption should be that the morally safer course is not to fight" (*Killing in War*, 145).

23 I am grateful to Massimo Renzo for valuable discussion about his piece and the issues taken up here, and to Erin Kelly for helpful conversation in the final stages.

Frowe, Helen, and Jonathan Parry. "Self-Defense." *Stanford Encyclopedia of Philosophy* (Summer 2022). https://plato.stanford.edu/entries/self-defense/.

Lazar, Seth. "Just War Theory: Revisionists vs. Traditionalists." *Annual Review of Political Science* 20 (2017): 37–54.

———. "National Defence, Self-Defence, and the Problem of Political Aggression." In *The Morality of Defensive War*, edited by Cécile Fabre and Seth Lazar. Oxford University Press, 2014.

McMahan, Jeff. "Just War Theory and the Russia-Ukraine War." *Justice Everywhere* (blog), October 31, 2024. https://justice-everywhere.org/international/just-war-theory-and-the-russia-ukraine-war/.

———. *Killing in War*. Oxford University Press, 2009.

Moreau, Sophia. *Faces of Inequality: A Theory of Wrongful Discrimination*. Oxford University Press, 2020.

Office of the Director of National Intelligence. *Annual Threat Assessment of the US Intelligence Community*. Report, March 2025. https://www.dni.gov/index.php/newsroom/reports-publications/reports-publications-2025/4058-2025-annual-threat-assessment.

Renzo, Massimo. "Political Authority and Unjust War." *Philosophy and Phenomenological Research* 99, no. 2 (2019): 336–57.

Rodin, David. *War and Self-Defense*. Oxford University Press, 2002.

Schwartz, Paul, and Dmitry Gorenburg. *Russian Military Mobilization During the Ukraine War: Evolution, Methods, and Net Impact.* Center for Naval Analyses, 2024. https://www.cna.org/reports/2024/10/Russian-Military-Mobilization-During-the-Ukraine-War.pdf.

Skelton, Anthony, "William David Ross." *Stanford Encyclopedia of Philosophy* (Spring 2022). https://plato.stanford.edu/entries/william-david-ross/.

Statista Research Department, "Number of Civilian Casualties in Ukraine during Russia's Invasion Verified by OHCHR from February 24, 2022 to May 31, 2025." Statista, June 16, 2025. https://www.statista.com/statistics/1293492/ukraine-war-casualties/.

Journal of Ethics and Social Philosophy
VOL. 30, NO. 5 · SEPTEMBER 2025
https://doi.org/10.26556/jesp.v30i5.4416

FIGHTING UNJUST WARS

POLITICAL AUTHORITY, TRAGIC CHOICES, AND THE VALUE OF OBEDIENCE

Massimo Renzo

POLITICAL AUTHORITIES are not perfect. Hopefully, we can agree about this much. Even legitimate authorities acting in good faith—those that diligently follow procedures designed to uphold justice and protect moral rights—occasionally issue mistaken directives. When this happens, we face the question of whether we ought to obey such directives. Do we have a duty to do so?

In many ways, this question goes to the very heart of the problem of political authority. After all, whenever legitimate authorities get things right, the duty to obey them is typically not in tension with other important moral duties we have. Interesting philosophical questions remain, of course—for example, whether we ought to *obey* the law (i.e., act as the law requires because the law requires it) or merely *conform to* it (i.e., act as the law requires but for other reasons).[1] But even philosophical anarchists will not lose any sleep over the question of how to act in these cases, nor should they. By contrast, when legitimate authorities issue mistaken directives that conflict with other moral obligations we have, the question of obedience becomes something more than a philosophical puzzle.

The question is especially pressing in a set of cases that have two features:

1. Obeying the mistaken directive would lead us to take part in the perpetration of a serious injustice.
2. The victims of the injustice in question are not themselves subject to the authority that issued the directive.

The significance of the first feature should be obvious. When obeying a mistaken directive would not lead to the perpetration of a serious injustice, it is easier to justify obedience as one of the costs of having political authorities that, necessarily, are not infallible. What is the significance of the second feature?

1 See, e.g., Wolff, *In Defense of Anarchism*; Simmons, *Justification and Legitimacy*, ch. 6; and Raz, *Ethics in the Public Domain*, ch. 15.

The idea here is that when the victims of the injustice are themselves subject to the authority responsible for the injustice, it is easier to justify obedience by appealing to the thought that they are themselves under a duty to support the authority in question. After all, if the authority has the power to impose on us obligations to comply with directives that will lead us to contribute to serious injustices, we might think that it also has the power to impose on the victims an obligation to bear the costs associated with those injustices.

To see this, consider a case in which a legitimate court acting in good faith wrongfully convicts an individual. While the conviction is in one sense unjust, since innocent people do not deserve to bear the costs of punishment, we might think that the defendant is nonetheless liable to bear those costs. This is because a duty to obey the law of one's country includes a duty to accept the decisions of its legitimate institutions, even when they are mistaken.[2] Notably, here it is the same justification that explains why, say, a prison guard has a duty to detain a wrongfully convicted person that is meant to also explain why that person has a duty to accept their own imprisonment. This is hardly surprising, since we are assuming that both the guard and the defendant are equally bound by the directive issued by the court.

The same line of argument, however, is not available in cases where the costs produced by a mistaken directive fall on someone who is not liable to bear them by virtue of being themselves bound by the directive. Suppose, for example, that state *A*, acting in good faith, mistakenly orders its citizens to wage a war against state *B*. Since the war is in fact unjust, obeying this order involves intentionally contributing to the killing of innocent people in *B*—i.e., people in *B* who have not done anything to waive their right not to be killed.[3] The members of *B*, however, have no duty to bear the costs of *A*'s mistaken decision, since they are not subject to *A*'s authority. Here, then, unlike in the wrongful conviction case, the justification that is meant to ground the duty to take part in perpetrating an injustice cannot also be invoked to ground a duty for those targeted by the injustice to bear its costs.

In addressing the question of whether we have a duty to obey unjust directives (call this *the question of obedience*), I will here restrict my attention to cases of unjust directives that have these two features.[4] Specifically, I will focus on the case of orders to fight wars that are unjust. But it is worth noticing that the

2 For a prominent version of this argument, see Estlund, "On Following Orders in an Unjust War."

3 This includes not only *B*'s civilians but also *B*'s enemy combatants. See McMahan, *Killing in War*, 10–15.

4 I will use the terms 'directive' and 'order' interchangeably throughout. Likewise, I will treat 'duty' and 'obligation' as synonymous.

question of obedience arises not only in the context of war. For example, state *A* might require its citizens to dispose of certain toxic substances in a way that will inflict unjust harm on the citizens of state *B*. As I see it, the moral issues raised by this directive are the same as those raised by the order to fight an unjust war. So, although my focus here is on the case of war, the view I defend extends to other cases of mistaken directives.[5]

In answering the question of obedience, philosophers have adopted one of two strategies. Some have argued that we have a duty to obey the order to fight even when the order is in fact unjust. The question of the justice of the war we are asked to fight is simply one that is not for us to answer. So even when our state makes a mistake in issuing the order to fight, our participation in the wrongs suffered by our victims is "sanitized because [we are] following orders."[6] Call this the *orthodox view*.[7]

The problem with this view is that insofar as the victims of our attack are morally innocent, they retain their right not to be attacked. For how could the existence of this right be compromised by an order issued by another state, no matter how well intentioned? This has led others to conclude that when a legitimate state orders us to fight a war that is unjust, our duty is to disobey, since even well-intentioned orders cannot change the fact that violating the rights of innocent people is morally impermissible. Call this the *revisionist view*.[8]

The reason to reject both these views, as I see it, is that they each neglect a morally significant aspect of the problem. If the war we are asked to fight is unjust, surely that has to make some difference as to whether we have a duty to obey. For even if we have a duty to obey the law, that is not the only duty we have. And it would be surprising if not contributing to the killing of innocent people did not figure in the list of our further duties. This is the aspect of the problem neglected by defenders of the orthodox view. On the other hand, the fact that we have a *pro tanto* duty to obey the law presumably does make some

5 I stress this point because arguments for the duty to obey orders to fight an unjust war sometimes rest on the idea that the morality of war is exceptional. According to this approach, the fact that in the context of war, we are permitted to act in ways that would normally be impermissible is explained by the thought that the morality of war is somehow special. The view I defend does not rest on this assumption.

6 Estlund, "On Following Orders in an Unjust War," 213.

7 Versions of this view have been defended by Walzer, *Just and Unjust Wars*; Estlund, "On Following Orders in an Unjust War"; Kutz, *On War and Democracy*; Shue, *Fighting Hurt*; and Buchanan, *Institutionalizing the Just War*. There are importance differences between them, but they need not concern us here.

8 Prominent defenses of this view include Rodin, *War and Self-Defense*; McMahan, *Killing in War*; Fabre, *Cosmopolitan War*; Frowe, *Defensive Killing*; and Tadros, *To Do, To Die, To Reason Why*.

difference to how we ought to act when a legitimate authority acting in good faith issues a binding directive.[9] This is the aspect of the problem neglected by defenders of the revisionist view.

My aim in "Political Authority and Unjust Wars" was to offer an answer to the question of obedience that does justice to both these aspects of the problem. I am very grateful to Thomas Christiano, David Estlund, and Christopher Kutz—three philosophers from whom I have learned a great deal in thinking about these issues—for their thoughtful engagement with that answer. Their contributions probe my view in important ways, raising insightful challenges from different directions. I particularly welcome the opportunity to address those challenges because though Christiano, Eslund, and Kutz use them to support answers to the question of obedience that are different from mine, they do so while recognizing the importance of meeting the two desiderata above. Engaging with their objections is thus a crucial way of strengthening the case for my account. For the plausibility of the account obviously depends not only on how it compares to traditional approaches, which neglect those desiderata, but also on how it compares to competing views that seek to meet them in different ways.

1. THE MODEL OF PRESUMPTIVE REASONS

Let me start by sketching in broad strokes my answer to the question of obedience. I set aside finer details for now, though we will need to come back to some of them later. My aim here is not to present a comprehensive articulation of my view but rather to provide an outline that is sufficiently informative to render the objections discussed below intelligible to someone who is not already familiar with it.

The core idea in "Political Authority and Unjust Wars" is simple. When a legitimate state acting in good faith orders us to fight an unjust war, we acquire a *pro tanto* duty to obey. However, this duty, being *pro tanto*, can be overridden under certain conditions by a weightier duty we have not to contribute to the unjustified killing of innocents. To defend this view, I offer an account of the *presumptive nature* (as I call it) of the reasons that legitimate authorities generate for us when they issue a valid directive.[10] This is where the complications begin.

9 To simplify my prose, I will assume that the legitimate state issuing the order to fight is acting in good faith. I will also use the expression 'unjust wars' to refer to wars that are in fact unjust, despite the fact that the state waging them is unaware of this. So the predicate 'unjust' is to be understood *de re*, not *de dicto*.

10 As I explain in "Political Authority and Unjust Wars," the notion of presumption I work with is practical, not epistemic (351). To see what I mean, consider the notion of

Unlike in the classic model of exclusionary reasons defended by Joseph Raz, according to which any competing reasons for action the authority was meant to consider in issuing its directives are to be excluded from our deliberation, I argue that authoritative directives create only a presumption in favor of such exclusion.[11] That presumption can be overridden when we are (1) justified in believing that the authority has made a mistake and (2) sufficiently confident about this belief. This is because when such an epistemic threshold is met, we are permitted—and sometimes required—to balance the new reasons created by the authority against the reasons the authority intended to exclude. And of course, when we do that, the duty to obey will be normally overridden when we are ordered to fight unjust wars since, barring exceptional circumstances, the reasons we have not to target innocent people (including enemy combatants who have done nothing to forfeit their right not to be attacked) will typically outweigh the reasons we have to obey the authority. Importantly, however, when the epistemic threshold is *not* met, we are not allowed to balance the new reasons created by the authority against those the authority intended to exclude. In this case, the duty to obey the order will remain undefeated.[12]

Does that mean we must scrutinize every directive issued by the authority for errors? No, that would defeat the purpose of having the authority. (More on this below.) Generally, it is enough to remain open to the possibility that the presumption might be overturned, occasionally conducting a quick

---

presumption of innocence, as it is used in criminal law. The presumption of innocence does not tell us what we should believe about the defendant. It tells us how the defendant should be treated. The defendant should not be punished unless their guilt has been proved beyond any reasonable doubt. I return to this idea below.

11 For this formulation of Raz's view, see Owens, "Rationalism About Obligation," 413–14. Strictly speaking, deliberating about excluded first-order reasons is allowed, according to Raz, as long as we do not act on them. But this is only because Raz understands deliberation as merely thinking about reasons rather than thinking about them with the aim of deciding what to do. This latter possibility is precisely what exclusionary reasons are meant to rule out. When exclusionary reasons are present, we are not allowed to balance competing first-order reasons against them in deciding what to do. I find Owen's formulation clearer, but nothing hangs on this choice. For Raz's classic discussion of exclusionary reasons, see his *Practical Reason and Norms*.

12 Since the question we are considering here is whether there is a duty to obey the order to fight in unjust wars, my discussion will focus on cases where the mistake made by the authority would lead us to act unjustly. Notice, however, that the model of presumptive reasons is meant to cover all cases of mistaken authoritative directives. In cases where obedience would lead us to act unjustly, whenever we meet the epistemic threshold, we are morally required to balance the reasons created by the authority against those the authority intended to exclude. In other cases (such as the one discussed in note 29 below), we are merely permitted to do so.

preliminary check rather than always undertaking a thorough investigation.[13] In some cases, the mistakes will be immediately apparent; in others, they will not, but this will not be a problem since, as far as we can tell, not much is at stake. Here, as we have seen, obedience can be justified as one of the costs of dealing with fallible political authorities. However, when we have reason to suspect that a political authority may be biased or prone to error, and when such errors could lead to grave injustices, we have a duty to actively scrutinize the authority's decisions. In these cases, the amount of scrutiny required is a function of both the likelihood that a mistake was made and the seriousness of the injustice in question.

The motivation for this view is twofold. On the one hand, our duty to obey the authority is primarily grounded in the idea that doing so is the best way for us to discharge our natural duty not to expose others to the dangers of the state of nature.[14] This is because authorities provide us with vitally important benefits—most notably by enforcing authoritative standards of conduct that, given reasonable disagreement about justice, are necessary to coordinate our behaviors so that we can respect each other's rights. Authorities, however, could not perform this function effectively if every time they issue a directive, each person's duty to obey was conditional on their own assessment that the directive is sound. On the other hand, the claim that the benefits of authority would be unattainable if we had to scrutinize every directive is compatible with the claim that, at times, we ought to do so. This is why the benefits provided by the authority are not lost if we scrutinize its directives in a restricted number of cases—namely, when we are justified in believing that the authority has made a mistake and we are sufficiently confident about that.

Importantly, this argument is premised on the idea that the authority in question is both *legitimate* and *acting in good faith*. Illegitimate authorities cannot create an obligation to fight unjust wars because they lack the right to impose obligations at all. This matters because, on my view, legitimacy requires that an authority be at least minimally just. Consequently, states that are seriously unjust lack the power to create duties to obey.[15]

As for legitimate authorities that do not act in good faith (i.e., those that intentionally fight unjust wars), they cannot create an obligation to fight

13 This idea comes from Schauer, "Rules and the Rule of Law" and *Playing by the Rules*.

14 I defend this view in "State Legitimacy and Self-Defence." Although this is the primary justification for the duty to obey the law, it is not the only one. See Renzo, "Associative Responsibilities and Political Obligation," "Fairness, Self-Deception and Political Obligation," and "A Pluralist Theory of Political Obligation." Still, the arguments developed here (as well as in "Political Authority and Unjust Wars") are grounded exclusively on this justification.

15 A seriously unjust state is one that fails to be minimally just.

because intentionally creating obligations to perpetrate serious injustices falls outside the scope of any normative power.[16] Just as we cannot generate obligations, even *pro tanto*, for ourselves to commit serious injustices by promising, an authority cannot generate obligations, even *pro tanto*, for us to perpetrate serious injustices by issuing an order.

2. RETHINKING LEGITIMACY

One way to put pressure on my view is by challenging the notion of legitimacy on which it rests. This is what Christiano does when he writes:

> Renzo … seems to say that a state is either legitimate or illegitimate. Legitimacy must depend on possessing a complex set of properties simultaneously. Hence, it appears that the legitimacy of the state cannot be piecemeal; it must be holistic. This approach is not adhered to on all dimensions, as far as I can tell, because Renzo does say that different people with different amounts of knowledge or access to knowledge might have different obligations. So, there is a certain amount of piecemeal legitimacy regarding who is under a duty. But Renzo seems to think that if a state is legitimate in one area, it must be in most or all other areas. But it is not clear that this is consistent with a thoroughgoing instrumentalism. Why not think that the state's commands enable you to act better in accordance with duty in some areas but not others?[17]

The first thing to say here is that my view of legitimacy is not piecemeal in the way Christiano suggests. Within my view, all the members of a legitimate authority are subject to the same *pro tanto* duty to obey. The reason some members are required to disobey while others are not is simply that our all-things-considered duties result from the interaction between the duty to obey and other moral considerations (most notably, competing duties) that apply to us, and these will be different for different people. This is a feature of all normative powers that generate obligations. Promises, for example, function this way. Even if you and I make the same promise, our all-things-considered duties may differ depending on which further moral considerations apply to us. It may well be that your all-things-considered duty is to keep the promise, and mine is not (if, say, doing so would be incompatible with discharging a more

16 I leave open here the possibility that a state can retain its legitimacy despite intentionally fighting unjust wars, though such cases, if they exist, are presumably uncommon.

17 Christiano, "War, Legitimacy, and Democracy," 753.

stringent duty of mine). This feature, however, does not render an account of promises piecemeal in any interesting sense.

The same holds for the authoritative directives issued by legitimate authorities. The distinctive feature of the model of presumptive reasons is that when the relevant epistemic threshold is met, stringent competing duties that were previously excluded from the scope of our deliberation are now capable of defeating the duty to obey authoritative directives.[18] However, this is a feature of the duty to obey as it applies to *all* those subject to the legitimate authority. It is true for all of them that when they meet the epistemic threshold, they ought to balance the duty to obey with competing duties. Thus, it would be misleading to view this account as piecemeal.

I mention this point to dispel the worry that there might be an internal tension within my view. Christiano's challenge, however, stands. For the challenge targets precisely the "holistic" understanding, as he puts it, of legitimacy that I work with. The challenge is: Why not take a more piecemeal approach, which would enable us to make more fine-grained judgments in determining who is under a duty to obey the order to fight unjust wars and who is not?

One way to do that, Christiano suggests, is to reject the idea that in order to be legitimate—and thus have the right to impose moral obligations—an authority must be minimally just. Within this approach then, even seriously unjust states—those that fail the legitimacy test in fundamental ways (say, by persecuting internal minorities or failing to maintain the rule of law)—can retain the right to impose obligations in relation to war-related matters. Their injustice does not undermine their legitimacy *tout court*, only their legitimacy with respect to the domains in which they act unjustly. Thus, if they do not act unjustly within the domain of war-related activities, they can enjoy legitimacy within this domain.

This is an intriguing suggestion, and piecemeal notions of legitimacy are certainly worth exploring. Still, I believe we should resist Christiano's proposal. This is because complying with the directives of a state is a way of providing support to its existence and its functioning. But when a state is seriously unjust, we have a duty not to provide this kind of support (barring exceptional circumstances).[19] Here Christiano might reply that insofar as our obedience is restricted to the orders falling within a domain in which the state is not acting unjustly—in this

18 Some have objected to my view that it is unclear how meeting the epistemic threshold could generate new duties. See Venezia, "Mistaken Authority and Obligation," 347–48. That is a misunderstanding of the view. Meeting the epistemic threshold does not generate any duties; it only allows us to balance preexisting duties that are normally excluded from our deliberation.

19 The sort of circumstances I have in mind are those discussed in note 35 below.

case, the domain of war-related matters—it is a good thing that it can count on our support. Thus, it is simply not true that our duty is not to support a seriously unjust state within this domain. This reply, however, is unconvincing.

To begin with, it is implausible that a state that regularly engages in human rights violations will set such violations aside in the context of war. If, for example, it systematically targets or discriminates against a religious minority, that persecution will likely not stop during the course of a military effort. Thus, it is hard to imagine a realistic scenario in which supporting the unjust state in its military efforts, even only those aimed at pursuing just wars, does not also involve supporting the unjust treatment of the minority in question. This is because it is unrealistic to conceive of the various domains in which states make decisions as insulated from each other in the way that Christiano's suggestion requires. Specifically, decisions about war-related matters are not completely disconnected from decisions about economic, cultural, and religious matters. Indeed, if we try to articulate what the boundaries of the domain of war-related activities are, it quickly becomes apparent that they cannot be drawn so as to neatly exclude other domains.

To take an obvious example, war efforts are normally funded by taxes. Does that make the tax system part of the domain of war-related activities to which, according to Christiano's suggested piecemeal account, we owe obedience? The answer cannot be no, since without financial resources there could be no war-related activities. But if the answer is yes, then it is hard to see how the domain of war-related matters can be insulated from the unjust goals pursued by the state in other domains. This is partly because the revenue secured through taxation will also be used to pursue such goals (say, the funding of segregationist policies), partly because in a profoundly unjust state, the tax system itself will likely reflect and entrench unjust social relationships within the political community in question.

Suppose now that we were to concede that there can be a clear separation between the various domains within which a state operates so that the injustices perpetrated within one domain do not necessarily have an impact on other ones. In this case too, I believe we should resist the idea that an unjust state can enjoy legitimacy within specific domains. To see this, consider the case of a mob family that tightly controls the criminal activities in a big city. Suppose that the family provides valuable services within certain domains to those living in that city, such as protecting them from other forms of criminality or providing food to families in need. Should we say that the citizens have a duty to obey the directives of the mob in these domains? That seems implausible. Even if we stipulate that in performing these valuable functions, the mob is not disrupting the activity of another legitimate authority (say, because the existing

institutions are inefficient or corrupt), it seems clear that the mob family lacks the right to impose genuine moral obligations. People might have reasons, even duties, to act in conformity with what the family requires, if doing so happens to be the best way to discharge independent moral duties they have. But these duties are not grounded in the fact that the mob enjoys anything resembling legitimacy (i.e., the right to impose binding moral obligations), not even within a restricted number of domains.

We should think about profoundly unjust states in the same way. The fact that they are engaged in the pursuit of serious injustice intuitively undermines any claim they might have to impose genuine moral obligations on us. We might have duties to act in conformity with what they require if doing so is the best way to discharge independent moral duties we happen to have. These duties, however, are not justified by the fact that such states enjoy legitimacy, albeit within restricted domains.[20]

3. RETHINKING AUTHORITY

As Estlund notices, the central idea on which my account rests is a formal point about the nature of authority—namely, the idea that when a legitimate authority issues an order, that order has *pro tanto* force even if it is mistaken, provided that the authority in question acts in good faith. So, although a state cannot generate a *pro tanto* duty to obey an order to ϕ if it knows that ϕ-ing would be unjust in the circumstances at hand, it can generate a *pro tanto* duty to obey if in issuing the order to ϕ, the state has made an honest mistake. This duty may be overridden by competing ones if the epistemic threshold is met; nevertheless, it retains its normative force.

This is what opens up the possibility of a duty to obey a mistaken order, even when doing so involves perpetrating a serious injustice. When the epistemic threshold is *not* met, our all-things-considered duty is to obey the order to ϕ, even if ϕ-ing is in fact unjust, simply because in this case, we are not allowed to balance the duty to obey against any competing duties.

A natural way to reject my view is to challenge this way of understanding the notion of authority—a strategy independently pursued by Christiano and Estlund in their contributions. This is an important objection, which, as we will see, can take different forms. And the idea behind it is appealing in its simplicity: Rather than holding that legitimate authorities generate *pro tanto* duties whenever their orders are issued in good faith, why not adopt a view in

20 One way to see this point is to notice that philosophical anarchists have no problem acknowledging the existence of such duties in the context of illegitimate states. This is because those duties are not grounded in legitimacy.

which unjust orders lack binding force altogether? On this alternative picture of authority, whenever the state commands us to fight an unjust war, we acquire no duty—not even *pro tanto*—to obey, and thus, there is no need to meet any epistemic threshold to justify disobedience. Within this view, the duty to obey does not exist in the first place. Since the state "has no authority to engage in unjustified war, . . . it cannot impose duties on citizens to participate."[21]

One way to defend this approach is to argue that the directives of a legitimate authority, at least in the domain of war, are binding only insofar as they are correct. And this view might seem especially attractive if, as I contend, the duty to obey the law ultimately rests on instrumental considerations—namely, the fact that obedience enables us to discharge our natural duties of justice.[22] Tempting as it is, however, the view ultimately rests on an implausible understanding of the notion of authority.

To begin with, if authoritative directives were binding only when correct, then every time we receive an order to ϕ, in order to establish whether the directive binds us, we would first need to determine independently whether ϕ-ing is what we are already required to do. But this would defeat the purpose of having the authority, since the justification of authority rests on the idea that within its domains of competence, a legitimate authority is generally better placed to determine how we ought to act. This is not only because the authority, given its resources and expertise, is normally best placed to assess what we have reasons to do in cases where those reasons are fully determined independently of its decisions, but also because there are cases in which what we have reason to do is *not* fully determined in this way. In these cases, what is required instead is some sort of coordination among those subject to the authority. But since multiple forms of coordination are possible, and reasonable disagreements can arise over which to adopt, the authority becomes necessary to select and enforce a common system.

For example, we know that disposing of industrial waste in certain ways can be extremely harmful. So, are the authority's directives requiring us to employ particular methods of industrial waste disposal binding? If answering this question requires establishing by ourselves whether the authority's chosen methods are indeed the right ones, the advantage of relying on the authority's judgment is lost. We also know that certain failures to coordinate can lead to "tragedy of the commons" scenarios in which natural resources become depleted. Are the directives of a legitimate authority aimed at preventing such outcomes binding?

21 Christiano, "War, Legitimacy, and Democracy," 755–56.

22 For the worry that my "good faith" condition is hard to justify within the sort of instrumentalist picture I favor, see Tadros, *To Do, To Die, To Reason Why*, 44.

The challenge here is even more acute for two reasons. First, assessing the directives in question is an especially complex task, since their effectiveness will depend on how multiple parties are meant to interact. Second, a uniformly followed albeit imperfect directive often proves more effective at averting tragedy of the commons scenarios than superior directives that lack universal compliance.[23] This is why Raz writes that "there is no point in having authorities unless their determinations are binding even if mistaken (though some mistakes may disqualify them). The whole point and purpose of authorities . . . is to preempt individual judgment."[24]

Although I cannot fully develop this point, it is worth noticing that there is nothing special about authority here. This is how we should generally think about normative powers, at least duty-conferring ones. Consider, once again, promising. Above, I said that we cannot generate *pro tanto* obligations for ourselves to commit serious injustices by promising. By that, I only meant that we cannot do so *intentionally*. We cannot incur a *pro tanto* obligation to ϕ by promising to ϕ if we know that ϕ-ing would constitute (or contribute to) a serious injustice. But we can incur a *pro tanto* obligation to ϕ by promising to ϕ if we are not aware of that. Here, too, the good faith condition applies.[25]

To see this, suppose you promise me that tonight you will turn on the irrigation system in my garden while I am away. In doing so, you acquire a *pro tanto* duty. Of course, if when you are about to flip the switch, you realize that doing so will cause a short circuit that will electrocute my neighbor, your promissory duty is overridden by your duty not to harm poor old Mrs. Clarke.[26] It is overridden in the same way it would be if activating the irrigation system conflicted with your duty to take an injured person to the hospital. By contrast,

23 Victor Tadros challenges this view by arguing that machines can in principle solve coordination problems by issuing random coordination points, as long as we agree to abide by them, whatever they may be. But since machines are not authorities and do not issue orders, Tadros argues, this shows that solving coordination problems does not require binding orders issued by authorities (*To Do, To Die, To Reason Why*, 81). The reason this argument fails is that once we assume the existence of an agreement to adhere to the machine's coordination points, the coordination problem no longer exists. Effectively, Tadros's case is one in which we have agreed on a specific way to coordinate, i.e., by complying with the machine's instructions. Genuine coordination problems arise precisely when no such agreement is in place. Further arguments offered by Tadros in support of his view raise similar issues, as they fail to address the challenges raised by disagreement, even among well-intentioned parties, in achieving successful coordination.

24 Raz, *The Morality of Freedom*, 47–48.

25 Venezia's attempt to reject my view by appealing to the parallel with promises (in "Mistaken Authority and Obligation," 350–51) fails because he neglects this distinction.

26 This is based on a case in Thomson, *The Realm of Rights*, 229.

you could not have incurred a *pro tanto* duty to turn on the irrigation system if you knew that doing so would electrocute Mrs. Clarke. Intentionally generating for yourself a duty to do something profoundly wrong falls outside the scope of the power to promise.

Here one might be tempted to raise the same objection that Christiano and Estlund raise about authority. Why not say that even when made in good faith, the promise to activate the irrigation system is not binding if doing so would result in Mrs. Clarke's demise? The answer is the same. If that was the case, promisor and promisee could not rely on the fact that a *pro tanto* promissory duty is created until they have first established that no serious injustice will follow from keeping the promise. But that would be incredibly burdensome. Knowing that any promise involving flipping a switch is binding would require first checking that the switch in question will not electrocute someone if flipped. Knowing that any promise to meet someone for lunch is binding would require checking first that this would not be incompatible with discharging our duty to take injured people to the hospital. That would render promise making highly inefficient, thereby undermining the very value of the institution of promising.[27]

Consider now a weaker revision to the notion of authority my account presupposes. Instead of arguing that states cannot generate a duty to take part in an unjust war by appealing to the idea that mistaken directives always fail to be binding, we might do so by appealing to the idea that only some mistaken directives fail to bind—namely, those that, if obeyed, would lead us to the perpetration of serious forms of injustice. This is the view that Estlund has in mind when he writes that even if we agree that the formal features of the structure of authority I have identified are substantially correct, "that cannot by itself show that a command from a legitimate authority can be morally binding, all things considered, regardless of how severe the wrong would be that one would thereby be contributing to." Estlund continues:

> Renzo's epistemic exception gives no antecedent weight to the severity of the wrong—only to warrant and confidence in the soldier's belief that it is wrong. What if the allegedly bracketed consideration was instead the fact that if the soldier were to obey, they would be contributing to the imminent collapse of civilization? Renzo's view suggests that even if there were little risk or cost from disobeying, unless the soldier not only believes the truth with good reason but also has some additional quantum of confidence, then obedience is required. No one should bite

27 Unlike in the case of authority, I do not go as far as to say that it would undermine our capacity to achieve the goods that justify the institution of promising. This is only because I do not have set views about what those goods are.

> that bullet just because authority has a certain formal bracketing structure. Even if it is correct, that whole structure of authority must surely be within the scope of certain exceptions, at least partly on grounds of great severity. And then the revisionist can fairly ask: Why would the killing of innocent people not be within the class of exceptions . . . thus preempting Renzo's epistemic exception requiring warrant and confidence?[28]

Despite calling for a weaker revision to the notion of authority, this argument faces the same objection discussed above. If we were to adopt Estlund's suggestion, whenever we receive a directive, we would need to determine whether obeying it would lead us to perpetrate a serious injustice. Doing so, however, would defeat the purpose of having the authority, since legitimate authorities are better placed to make that judgment, especially in contexts that require coordinated action. Estlund's view is weaker than Christiano's because it restricts the set of cases in which directives fail to bind us to situations in which obeying would lead us to perpetrate some serious injustice. But we are still left with the problem of having to assess every directive we receive in order to determine which ones fall within this set. There is, however, a different objection, which is in the neighborhood of Estlund's but avoids this difficulty.

While having to evaluate every directive for its potential to lead us to cause severe injustice would defeat the purpose of having the authority, doing so in a restricted number of cases—i.e., when it is especially clear that such an outcome might be produced—does not. And indeed, this is precisely what my view calls for. In cases where we have reason to believe that obeying might lead us to the perpetration of serious injustice, my view requires that we scrutinize the directive of the authority. Still, as we have seen, the directive remains binding in these cases too, according to my view, and this is why our duty is to obey it unless we meet the epistemic threshold. Here one might object, in the spirit of the worry raised by Estlund, that this is implausible. Would it not be better to adopt a picture of authority according to which when we have some reasons to believe that obeying the directive might lead to severe injustice—even if those reasons are not strong enough to meet the epistemic threshold—the directive is not binding at all? An advantage of this view, we might think, is that when the epistemic threshold is not met, disobedience rather than obedience is required.

The problem with this picture of authority is that although it does well with false positives (cases in which obeying would lead to severe injustice), it does terribly with false negatives (cases in which failing to obey would lead to severe injustice). Avoiding false negatives, however, is surely as important. After all,

28 Estlund, "The Weight of Authority in War," 782. At times, Christiano also seems to have in mind this formulation of the objection.

if cases in which states order us to fight wars that will lead to the collapse of civilization are certainly possible, so are cases in which they order us to fight wars that are necessary to prevent the collapse of civilization. So, what happens if we have some reasons to believe that obeying the order to fight a given war might lead to severe injustice, but in fact, the war in question is necessary to prevent the collapse of civilization? Adopting a view of authority according to which such an order is not binding unless we are sufficiently confident that the order is just would render authorities ineffective in preventing the collapse of civilization since they would be unable to count on our obedience.

This provides further support for the central idea on which my account rests. Since legitimate authorities are generally best placed to assess what we have reasons to do, as well as to coordinate collective behavior, this gives us reasons to treat their directives as *pro tanto* binding, so that the default, when we fail to meet the epistemic threshold, is obedience. A model that treats these directives as not binding (such as the one just considered or the one adopted by revisionist approaches) makes disobedience the default whenever we fail to meet the epistemic threshold. This, however, would undermine our capacity to achieve the very goals that the authority is meant to help us secure.[29]

4. War and Tragedy

Suppose I am right that a legitimate authority acting in good faith can generate a *pro tanto* duty to obey an order to fight an unjust war. Another way to criticize my account is by rejecting the idea that when this duty conflicts with the duty not to contribute to the killing of innocents, we should expect the latter to prevail. Kutz pursues this line of argument, objecting that the very thought of resolving the

29 This provides a natural account of cases that traditional theories of authority notoriously struggle with. Consider, for example, the case of the red traffic light on a deserted road late at night, where visibility is perfect, and it is clear that no one is approaching. The idea that one has a duty to obey the directive to stop in such a scenario seems so implausible that many take this case to show that there can be no legitimate authority (or at least no duty to obey legitimate authorities). If we adopt the traditional understanding of authority, the only available reply here is to bite the bullet and accept, as Raz does, that one ought to "blindly" follow the directive and stop at the red light (Raz, *The Authority of Law*, 24). By contrast, the model of presumptive reason treats this as a case in which the epistemic threshold is easily met. This explains why, at least in some circumstances, the duty to stop will be easily overridden by competing considerations. Even in such cases, however, it is important that obedience be treated as the default from which departure must be justified on a case-by-case basis by establishing that the epistemic threshold is met. A system in which disobedience is treated as the default and obedience has to be justified on a case-by-case basis (by identifying good reasons to stop) would fail to adequately serve the purposes that justify the existence of traffic law for the reasons discussed in the text.

tension between these duties is misguided.[30] Instead, we should acknowledge that the tension is irresolvable and that what we are faced with is a tragic choice.

This approach has two advantages, according to Kutz. First, it explains why it would be appropriate for soldiers who obey the order to fight an unjust war to feel regret for their actions. Second, it explains why their victims have the power to forgive them. And surely both ideas are plausible. To see this, consider Jose, the soldier (described by Kutz) who, having failed to meet the epistemic threshold, fights in an unjust war. If Jose later feels no regret for his conduct, even after new evidence has shown that the war was in fact unjust, that would be deeply troubling. Likewise, it seems clear that the victims of Jose's actions have the right to forgive him. But if Jose was all-things-considered justified in acting as he did, Kutz argues, there is no place for regret or forgiveness.

The idea of conceiving Jose's situation as a moral dilemma is an intriguing and, in many ways, natural suggestion. It also has the advantage of capturing the spirit of my own view (the thought that an adequate answer to the question of obedience requires acknowledging the force both of our duty not to contribute to the killing of innocents and of our political obligations) while avoiding complicated discussions about which reasons are excluded from our practical deliberation under which conditions. Instead, we simply have two competing duties and no resolution about which one should prevail. This means that no matter what we do, we will do something wrong. But this is a feature, not a bug of the view, Kutz argues. Indeed, it is a welcome feature since it explains both why Jose ought to regret his decision to fight and why his victims have the power to forgive him. Given that Jose's duty to obey cannot be said to override his duty not to contribute to the killing of innocents, it is the violation of this latter duty that warrants his regret, as well as the power of his victims to forgive him. Similarly, had Jose decided not to fight, he would have had to regret that decision, and his fellow citizens would have had the power to forgive him.

The reason to resist this suggestion is that it seems implausible to claim that the following two duties constitute a genuine moral dilemma:

1. The duty not to contribute to severely harming innocent people.
2. The duty to obey a mistaken order that we are sufficiently confident (having met a suitably robust epistemic threshold) will lead us to contribute to a serious injustice.

To see this, let me propose the following question as a heuristic for determining whether a given conflict of duties constitutes a genuine moral dilemma: *Would*

30 Christiano also raises this worry, albeit more briefly. See Christiano, "War, Legitimacy, and Democracy," 758.

both parties whose claims are at risk of being violated by the agent's decision have a legitimate complaint if their claims were indeed violated?

The idea behind this test is that if one party intuitively lacks a legitimate complaint, this is plausibly explained by the fact that the competing duty had overriding force after all. So, if Sartre's soldier decides to stay with his mother, intuitively, his fellow soldiers can legitimately complain that the value of fighting the Nazis should not have been subordinated to the value of caring for a family member; and if he opts to join the Free French Forces, intuitively, his mother can legitimately complain that the value of caring for her should not have been subordinated to the value of fighting the Nazis. This shows that the soldier faces a genuine moral dilemma. But suppose that the soldier had faced instead a choice between joining the Free French Forces and keeping his promise to regularly water his neighbor's plants. Any complaint by the neighbor that the soldier ought to have kept his promise would plainly lack force. This is evidence of the fact that in this version of the story, the soldier was not facing a dilemma.

Let us now examine the case in which, having met the epistemic threshold, we face the question of obedience. If we obey the order to fight an unjust war, the victims of our actions clearly have a legitimate complaint about our decision to contribute to the killing of people who we are sufficiently confident are innocent. If we do not, do our fellow citizens have a legitimate complaint that we should have obeyed an order that we were sufficiently confident (having met a suitably robust epistemic threshold) would have led us to contribute to a serious injustice? This seems implausible.

One way to see this is to notice that while in the case at hand, obeying the mistaken directive conflicts with our duty not to inflict unjust harms on members of another political community, there will be also instances when obeying a mistaken directive conflicts with our duty not to inflict unjust harms on members of our own political community. (Think, again, of a scenario in which we are required by law to dispose of industrial waste in a way that will significantly harm our fellow citizens.) Plausibly, in these cases, our fellow citizens would not insist that our duty is to obey the directive if we are sufficiently confident that it is mistaken. If so, this shows that we do not face a genuine dilemma when our choice is between the duty not to contribute to severely harming innocent people and the duty to obey mistaken orders that we are sufficiently confident will lead us to contribute to a serious injustice.

5. Justification and Regret

A worry raised by both Kutz and Estlund is that my view seems to be incoherent or paradoxical. For Kutz, this is a worry about the psychology of combatants.

Soldiers will experience their own position as incoherent insofar as they will have to think of themselves as justified in fighting (since they are under a binding obligation to do so) and, at the same time, as unjustified in doing so (since they cannot rule out that the war is unjust). For Estlund, the paradox concerns the very idea that a given conduct can be morally unjust and morally required at the same time. In this section, I consider Kutz's objection. Estlund's is the focus of the next one.

In addressing Kutz's objection, the first thing to notice is that it needs to be reformulated. Surely, if there is a view that raises the worry of incoherence or of generating paradoxical results (though I would not myself put the point quite this way), it is the one that sees soldiers as facing a moral dilemma. For if they are facing a moral dilemma, they are—by definition—in the position of having to see themselves as having a duty to do something and, at the same time, having a duty not to do it. According to my view, by contrast, regardless of whether soldiers end up obeying or disobeying, they can see themselves as acting for an undefeated duty. Although they are aware that even legitimate states acting in good faith sometimes wage unjust wars by mistake, soldiers who obey may take comfort in the fact that, when they are not sufficiently confident that the state has made a mistake, obedience is the most reliable way to discharge their natural duties of justice.[31]

Now, this is precisely the idea that Kutz finds problematic. His point is that this way of thinking fails to do justice to the complexity of the situation. As he puts it, "this tragic dimension of individual political membership . . . may simply be an irreducible fact of life even in a generally legitimate state, so long as that state exists in a confusing and unjust world. We disguise it by pretending that the duties can be ranked and thus reconciled, as Renzo does. But they cannot be."[32] Perhaps. But even if this concern has merit, the worry is not that my account is paradoxical or incoherent but rather that it mischaracterizes our normative situation. Soldiers might be delusional in understanding their situation as my account recommends, but they would not be facing any paradox or incoherence. Still, addressing Kutz's arguments in support of the incoherence objection is helpful, as they rest on a misunderstanding of my position that needs clarification. So let us look at them more closely.

Kutz understands my view as involving a form of justification that is neither fact-relative nor evidence-relative but "hybrid," in the sense that it makes the permissibility of fighting depend on the actual justice of the war and, at

31 If they are shaken by the incoherence of two conflicting duties, one of which is overridden by the other, wait until they hear that, in fact, neither duty is overriding, and they will be acting wrongly no matter what they do!

32 Kutz, "Soldiers and Moral Tragedy," 769.

the same time, on whether we have epistemic access to this information.[33] According to this interpretation of my view, if after the war, Jose acquires further evidence that the war was unjust, he should somehow look at his conduct as being justified and unjustified at the same time, and herein lies the alleged incoherence. To support this point, Kutz considers cases in which we find ourselves having to gamble or to guess at some future outcome. In such cases, we will try to make the best choice we can, given the evidence available to us at the time, Kutz concedes. But this cannot change the fact that our choice lacks fact-relative justification if, as it turns out, we choose wrongly. Which is why we should feel regret for our choice, even if we know that it was justified from the evidence-relative perspective at the time of the action.

I should say that, even if we stick to cases of gambling and guessing, I do not find the sort of psychological state described by Kutz problematic. These are simply cases in which, looking back, we should think that at the time we acted, we had an evidence-relative justification for our conduct but, as it turns out, not a fact-relative one. Regret is certainly an apt response, but the worry of incoherence seems once again misplaced.

More importantly, my view does *not* treat Jose's case along the lines of the gambling and the guessing scenarios mentioned by Kutz. For my view relies exclusively on fact-relative considerations to justify Jose's duty to fight, not on evidence-relative ones. Jose's justification for fighting is that the fact-relative reasons that ground his duty to obey remain undefeated. It is true that evidential considerations play a role in determining when a further set of fact-relative reasons can be admitted in Jose's practical deliberation, and that when that happens, these further reasons (which ground his duty not to kill innocents) can defeat his reasons to obey. This, however, is not enough to render the justification evidence-relative. The reasons on which the justification of Jose's conduct rest remain exclusively fact-relative.[34]

To see this more clearly, consider how my view would handle cases in which we do meet the epistemic threshold. In such cases, my view says that our all-things-considered duty is to disobey, barring exceptional circumstances.

33 Kutz prefers to distinguish between 'objective' and 'subjective' justifications, but I will stick to the distinction between 'fact-relative' and 'evidence-relative' justifications, which I use in "Political Authority and Unjust Wars." Nothing hangs on this choice.

34 Luciano Venezia and Rodrigo Sánchez Brígido argue that my account cannot vindicate the idea that Jose has a fact-relative justification for obeying, since in "kill(ing) innocent persons without justification," he is acting against what he has reasons to do ("Between Traditionalism and Revisionism," 362). But this argument begs the question against the view that the duty to obey can provide such a justification. My account is precisely meant to explain why it does.

This is not, however, because given the evidence available to us, we are justified in believing that fighting would be unjust. Rather, it is because our fact-relative reasons to obey are overridden by the fact-relative reasons we have not to contribute to the killing of innocents. So once again, this is a fact-relative justification through and through.[35]

The comparison with how we normally think about the justification for the conduct of other officials is helpful here. Take judges, for example. If the prosecution fails to prove a defendant's guilt beyond a reasonable doubt, the judge must acquit. That is the case even if the only reason the standard of proof was not met is that crucial evidence was thrown out on procedural grounds, and thus the judge knows that the defendant committed the crime (and did so with the requisite *mens rea*). Clearly, the justification for the judge's conduct here is not that they have evidence-relative reasons to believe that the defendant did not commit the crime. By hypothesis, they know that the defendant did. Rather, the justification is that correctly following criminal procedure gives

35 In writing that our all-things-considered duty is to disobey *barring exceptional circumstances*, I have in mind two types of cases. First, there might be situations in which, despite meeting the epistemic threshold, we are nevertheless mistaken about the injustice of the war. I might have collected sufficient evidence to justify both my belief that the war is unjust and my confidence in this conclusion—and yet still be mistaken. Against all the evidence, the war is in fact just. In this case, because the epistemic threshold is met, I am required to balance the fact-relative reasons I have to obey with the fact-relative reasons I have not to contribute to the war. But since, by hypothesis, the war is not unjust, I do not have any fact-relative reasons not to take part in it. Thus, my all-things-considered duty is to obey the order, despite meeting the epistemic threshold. If I do not obey because I reasonably believe that my duty is to disobey, my conduct might be excused. (Blame is justified by how we respond to our evidence-relative reasons for action rather than to our fact-relative ones.) Still, I will have failed to discharge my all-things-considered duty.

The other kind of exceptional circumstance I have in mind concerns cases in which, despite meeting the epistemic threshold, we may nonetheless have an all-things-considered duty to take part in an unjust war on lesser evil grounds. Consider:

> *Lesser Evil*: State *A* is fighting (in good faith) an unjust war against state *B*. *A* should stop the war, but it is not going to because it believes the war to be just. If *A* loses the war, the consequences would be disastrous (say, because *A*'s defeat would destabilize a vast region, leading to far more innocent deaths than those produced in defeating *B*).

Here, *A*'s citizens have an all-things-considered duty to fight. Importantly, however, this is not because the duty not to contribute to the killing of innocents in *B* is outweighed by the duty to obey. It is because the duty not to contribute to the killing of innocents in *B* is outweighed by the duty to prevent disastrous consequences. So, while *A*'s citizens do have a fact-relative justification, all things considered, for fighting, this is not strictly speaking a justification for obeying the order. Rather, it is a justification for conforming to it, in the way that philosophical anarchists believe we ought sometimes to conform to what the law requires.

the judge undefeated fact-relative reasons to acquit. Similarly, our duty to obey (or disobey) is the product of the fact that following the correct deliberative procedure generates undefeated reasons to do so, not the fact that this is the right choice in light of the evidence available to us at the time of the action.

Importantly, none of this undermines the aptness of feeling regret if it turns out that the order we obeyed was mistaken. Consider again Jose's case. That he had a fact-relative justification for obeying means only that he is not at fault for doing so. This is compatible with the presence of regret, however. Regret requires Jose to "wish that things had been otherwise," as well as some sort of first-personal reaction to his own conduct, such as willingness to compensate his victims or apologize to them.[36] But if it is apt for the lorry driver in Williams's famous case to have such responses despite not being at fault for his conduct, the same holds for Jose. And if it makes sense for Jose to apologize, it also makes sense for his victims to forgive him.

My view thus has no problems vindicating either of the ideas Kutz uses to motivate the claim that the tension between the duty to obey and the duty not to contribute to the killing of innocents is best understood as a tragic choice. His mistake is inferring that Jose faces a tragic choice from the aptness of Jose's regret. This inference is unwarranted because even if we agree that regret is always present in tragic choices, it is not present *only* in tragic choices (as Williams's case illustrates).

6. *PRIMA FACIE* VERSUS *PRO TANTO* DUTIES

While for Kutz, my view is paradoxical in the sense that its adoption would be psychologically unstable, for Estlund, it is paradoxical in the sense that it involves a sort of conceptual incoherence. His worry concerns my claim that soldiers who have an all-things-considered duty to obey, despite being on the unjust side, can be said to wrong their victims and thus perpetrate a serious injustice, despite the fact that they act permissibly. This is because on Estlund's view, an action cannot be wrong (or unjust) and permissible at the same time.

Estlund understands, of course, that my claim is that Jose's action is only *pro tanto* wrong. But it is precisely the claim that an action can be *pro tanto* wrong and all-things-considered permissible that he regards as "unnecessarily paradoxical." To get around this issue, he reformulates my view in terms of *prima facie* rather than *pro tanto* duties. Unlike *pro tanto* duties, *prima facie* duties do not possess genuine normative force. Rather, they are best understood as initial

36 Williams, *Moral Luck*, 31. I acknowledge the aptness of these responses ("Political Authority and Unjust Wars," 356).

appearances of a duty—appearances that may not survive closer scrutiny and thus might turn out not to constitute actual duties in the end.

How does this help with the alleged paradox? Once reformulated in this way, my view would say that whenever the epistemic threshold is not met, our *prima facie* duty to obey the order becomes an actual duty (or "duty proper"), and our duty not to kill the innocent people we will be targeting is thereby revealed to be just apparent. But if our duty not to kill them is just apparent, then we do no wrong in killing them. To be sure, our victims still suffer a wrong, insofar as they have not lost their right not to be killed. But the wrong is not inflicted by us, Estlund argues. Rather, it is inflicted by the state that has issued the order to fight. It is in this sense that our participation in the unjust war can be "sanitized" by the fact that we are following orders.

Let me start by noticing that here, too, the charge of being paradoxical seems misplaced—not so much because the distinction between *pro tanto* and all-things-considered force is widely accepted in contemporary moral theory, but because the point of the distinction is precisely to explain why it is a mistake to conflate the claim that a certain course of action has a genuine normative valence (such as being wrong) with the claim that it reflects what we ultimately ought to do. Conceptually, these are two different types of moral judgments, and the point of the *pro tanto*/all-things-considered distinction is precisely to dispel the air of paradox that certain moral statements have if we fail to appreciate this difference.

I do not deny, of course, that the distinction might still have a ring of paradox to some ears.[37] But if any such feelings linger even once we have grasped the distinction, we should disregard them as unwarranted. After all, we might imagine the same feelings persisting once we adopt Estlund's suggestion of formulating my view in terms of *prima facie* duties. Suppose one were to argue that there is a ring of paradox in the claim that we have a *prima facie* duty not to contribute to the killing of someone who has not forfeited their right not to be killed while, at the same time, having an actual duty to obey an order to do just that. Surely any uneasiness one might feel about this formulation would be unwarranted. The point of invoking the distinction between *prima facie* and actual duties is precisely to explain how the two are compatible once we appreciate their different nature.

Of course, there might still be good reasons to reformulate my account in terms of *prima facie* rather than *pro tanto* duties, if doing so better illuminates the moral phenomena we are trying to capture or if the resulting view is, say, more elegant or parsimonious. These reasons, however, have nothing to do

37 Though, for what it is worth, not to mine.

with the claim that the view is somehow paradoxical. It is on these reasons that we should focus to adjudicate whether we should adopt the model of *pro tanto* reasons or the *prima facie* ones, so let us look at them instead.

One problem for the view that the duties Jose must balance are *prima facie* rather than *pro tanto* is the implication that if Jose fails to meet the epistemic threshold, he cannot be said to have reasons to regret how he acted, let alone to apologize or compensate his victims. This is because within the *prima facie* model, once the duty to obey becomes actual, any competing duties lose all their normative force. (Indeed, strictly speaking, we should say that they are revealed to never have had such force to begin with.) But this cannot be right. As Kutz notices, there is a clear sense in which Jose ought to regret acting as he did, even when he has a justification for his conduct. And it also seems plausible that he might have compensatory duties and duties to apologize, though this further claim is perhaps more controversial.[38]

One might reply that the force of this objection is provisional, in the sense that its plausibility ultimately depends on how we should understand the justification of regret, compensation, and apologies. Obviously, I cannot address such questions here, but even if we concede this point and bracket the aptness of these responses, a fundamental problem with the *prima facie* model remains. The problem is that within this model, once it has been established that Jose's actual duty is to obey, there seems to be no room for the idea that there are nonetheless weighty considerations in favor of him not attacking his targets.[39] And *that* surely is implausible. That such considerations persist seems clear, regardless of the role they should play in calling for an apology, regret, or compensation. It is a serious cost of the *prima facie* model that it fails to vindicate their existence.

Estlund could perhaps reply that, strictly speaking, the *prima facie* model does not need to rule out the existence of such considerations; it rules out only the idea that such considerations render Jose's conduct wrong. What is rendered wrong by them is rather the conduct of those who gave Jose the order to fight. This reply would be puzzling, however. If the innocence of his targets is not only a consideration for Jose's state not to order their killing, but also a consideration for Jose himself not to kill them, it's hard to see why that same consideration should render the state's decision wrong but not Jose's conduct.

A further problem with the *prima facie* model is that it seems to have limited explanatory power. This is because establishing what our duty is within this model effectively requires us to already have an answer to the question of which

38 See section 5 above.

39 Kiesewetter, "*Pro Tanto* Rights and the Duty to Save the Greater Number," 202.

of the competing moral considerations we are faced with should prevail. But intuitively, the fact that we have a duty to do something is itself a consideration that should play a role in answering that question.

For example, knowing that I have a *duty* to keep my promise to meet a student and merely *reasons* to watch the football match taking place at the same time often explains why I ought to keep the meeting rather than watch the match.[40] Within the *prima facie* model, by contrast, the notion of duty seems to be essentially a heuristic device used to signal that the question of how we should act has already been solved. But how should we go about solving that question then? Not being able to rely on the notion of duty significantly reduces the tools available to us to do so. We can of course still balance the competing reasons that apply to us. But the value of the notion of duty is precisely to mark the existence of situations in which what we ought to do is not exclusively determined by the weight of the sum of individual reasons. To say that we have a duty to do something is to recognize that some of these reasons have a distinct peremptory force, which gives them a special role to play our deliberation regardless of their weight. A view that cannot vindicate this idea misses an important dimension of the sort of moral problems we are considering here.

In light of these problems, we should resist the suggestion to abandon the understanding of authoritative directives on which my argument rests.[41] The duty created by binding authoritative directives is best understood as having *pro tanto* rather than *prima facie* force.

7. THE VALUE OF OBEDIENCE

I conclude by considering a further argument raised by Estlund. As he acknowledges, this argument is somewhat modest. Its aim is to highlight that revisionists cannot rule out the possibility that the duty to obey can outweigh the duty not to fight in an unjust war simply by assuming that the only plausible

40 'Often' is not an understatement here. My point is not that duties will always prevail over competing reasons in this sort of cases. Perhaps there are situations in which the football match is so valuable to me that I am justified in breaking my promise. See Wolf, "Above and Below the Line of Duty"; and Tadros, *Wrongs and Crimes*, 30–32. The point is that the fact that I have a duty, as opposed to mere reasons, to keep the promise is meant to play an important role in determining how I should act. This is true regardless of whether the duty is defeated by competing considerations.

41 It is no coincidence that the model of *prima facie* duties raises problems that mirror some of the main worries afflicting specificationist accounts of moral rights. See Feinberg, *Rights, Justice, and the Bounds of Liberty*, 221–51; and Thomson, *The Realm of Rights*, 82–104. Both positions are ultimately motivated by the same basic idea—namely, that once the scope of application of these normative phenomena is suitably qualified, they have absolute force.

justification for killing innocent enemy combatants is preventing the killing of other innocent lives or other similarly dire consequences. This is because revisionists already acknowledge at least one notable exception to this idea—namely, wars of national defense.

A war of national defense is one in which a country is threatened with violent conquest that is aimed not at killing anyone but rather at securing control over "political goods" such as territory, natural resources, or political self-determination. Importantly, although aggressors taking part in such wars act unjustly, this does not mean that they are not innocent, as Estlund notices. This is both because, statistically, many of them do not end up posing any actual threat of serious harm and because even when they do, they might do so nonculpably if they have been forced to take part in the war.[42] So if revisionists are willing to concede that it is permissible to take part in wars of national defense, this shows that they already acknowledge the possibility of cases in which killing innocents can be permissible, despite the fact that doing so is not necessary to prevent seriously harmful consequences. (Calls this an instance of *subtle justification*.)[43] But if so, Estlund's argument continues, then revisionists cannot easily rule out that a similar permission can exist in other cases. Specifically, if the value of national defense can provide a subtle justification for killing innocents, the same could be true, at least in principle, for the value of obedience to a legitimate authority.

To be sure, those fighting in wars of national defense fight on the just side, unlike those facing the question of obedience in the sort of case we are considering here—and that, we might think, makes a difference. After all, the reason we can entertain the idea that the duty not to contribute to the killing of innocents can be defeated by the duty to protect political goods is that the existence of such duties, at least when we are unjustly attacked, does not seem controversial. But can we say the same for the duty to obey unjust orders? Estlund needs to show this in order to complete his argument. For if the idea of a duty to obey an unjust order is implausible to begin with, obviously any attempt to rely on the value of obedience to defeat the duty not to contribute to the killing of innocents is a nonstarter. To support this idea, Estlund appeals to a further analogy. He considers the case of a jailer who is ordered to incarcerate someone who has been erroneously convicted after a fair trial. Estlund argues that the order received by the jailer is intuitively binding, and for the same reasons, we should think that so is the order to fight in an unjust war.

42 As in Estlund's imaginary case of Crushia's aggression to Nukraine.

43 The label 'subtle justification' was used by Estlund himself in an earlier version of his paper to refer to a justification for defensive killing that appeals to something other than "causes, contributions, or complicity around harms to persons" ("The Weight of Authority in War," 784).

Ultimately, then, Estlund's argument is that the best way to justify the duty to obey the order to fight an unjust war is by showing that (1) the orders of a legitimate authority are binding even when they are seriously unjust, and (2) there is no principled reason to rule out that the value of obeying binding orders can be sufficiently weighty to provide a subtle justification for killing innocents, once we acknowledge that the value of national defense can provide such justification.

Now, strictly speaking, this is not an argument against my view. In principle, both my argument and Estlund's can independently succeed in justifying the duty to obey the order to fight in an unjust war. The motivation for presenting his argument in alternative to mine is Estlund's skepticism about my argument, prompted by his other objections. If my answers to those objections are persuasive, this motivation is obviously weakened, and so is the pressure to rely on Estlund's alternative strategy. Still, his argument deserves close attention. For while I think the argument ultimately fails, it raises a deep question about how the notion of authority ought to figure in a plausible answer to the problem of obedience. There is an important lesson to be learned from engaging with Estlund's answer to that question.

I will not discuss here the jailer analogy, both because I criticize that argument elsewhere and because I have already mentioned above a reason to be doubtful about it.[44] There is an important difference between cases in which the victims of the injustice produced by our obedience are themselves subject to the authority that has issued the order and cases in which they are not. In the former case, the justification that is meant to ground our duty to take part in perpetrating the injustice can also be invoked to ground the victim's duty to bear the costs associated with it; in the latter, it cannot. Once we pay attention to this point, it is far from obvious that the conclusion reached in the jailer case should be presumed to apply to the case of obedience to the order to fight in an unjust war.

Be that as it may, my focus here will be Estlund's argument that since revisionists admit that the value of national defense can provide a subtle justification for killing innocents, they cannot—without further argument—rule out that the value of obedience can do the same. The problem with this argument is that Estlund offers it without considering in sufficient detail the question of *how* revisionists justify the permission to participate in wars of national defense. He objects that appealing to the violation of political rights (such as the right to participate in collective sovereignty or political self-determination) to justify killing innocents is in tension with two core revisionist commitments: first, that lethal defensive

44 I analyze the jailer analogy in Renzo, "Democratic Authority and the Duty to Fight Unjust Wars."

force can be justified only by appealing to "personal harms"; and second, that this justification is regulated by "ordinary interpersonal morality" rather than a special morality of war.[45] Yet he does not elaborate on either objection. Instead, he pivots, somewhat unexpectedly, to a different argument. He observes that only those who have received a binding order from a relevant authority are permitted to kill innocent aggressors in a war of national defense. Someone lacking this authorization would not be permitted to do so. According to Estlund, this supports the view that, at least in wars of national defense, following orders is what grounds the permissibility of using lethal defensive force. Authority thus has "a surprising moral weight." It can "convert from wrongful to permissible what would otherwise be a wrongful killing of a relevantly innocent person."[46]

Both moves—ignoring how revisionists justify wars of national defense and pivoting to the question of authorization—are problematic. Start with the latter. The question of whether someone lacking the relevant authorization is permitted to use lethal defensive force in a war of national defense is orthogonal to the question of whether the value of national defense can provide a justification for killing innocent aggressors. Estlund runs them together when he attempts to derive an answer to the latter from his answer to the former. The problem here is that even those who regard acting under a "legitimate authority" as a genuine *jus ad bellum* criterion do not understand it the way Estlund does.[47] The criterion only states a necessary condition for being permitted to take part in a war: we may take part in a war only if we are ordered to do so by a legitimate authority. The criterion does not purport to provide a justification in the sense that meeting the criterion can be said to weigh in favor of fighting (or, as I would put it, "give us reasons to fight"). To infer the latter from the former is a mistake. Compare: in many countries, being at least eighteen years old is a necessary condition for being legally permitted to drink alcohol.[48] Being eighteen, however, does not provide a justification in the sense that meeting the minimum drinking age requirement can be said to weigh in favor of drinking alcohol. The justification for doing so must come from elsewhere—that I would enjoy drinking, that I want to drown my sorrows, that I wish to loosen up at a dinner with strangers, and so on. If it does not—say, because I do not like drinking or I am intolerant of alcohol—the fact that I am eighteen years or older does nothing to change that.

45 Estlund, "The Weight of Authority in War," 787.

46 Estlund, "The Weight of Authority in War," 788.

47 It is worth noticing that many revisionists do not accept the legitimate authority requirement. See McMahan, "Just War," 671–72; and Fabre, *Cosmopolitan War*, chs. 3–4.

48 At least when the drinking takes place in public, as opposed to in one's home.

Raising the question of authorization only muddies the waters then. Nor does Estlund's argument depend on it. For his argument, remember, is that since revisionists admit that the value of national defense can provide a subtle justification for killing innocents, they cannot easily rule out the possibility that the value of obedience can do the same. This argument, if successful, is enough to vindicate Estlund's conclusion, without any need to appeal to the issue of authorization.

The argument is not successful, however. This is because the revisionist case for the permissibility of taking part in wars of national defense is *not* in fact best understood as an instance of subtle justification. The revisionist case is, roughly, that since we have vitally important interests in enjoying goods such as collective self-determination, sovereignty rights, and political independence, the value of protecting these interests is sufficiently weighty to justify resorting to lethal defensive force.[49] (Or, as Estlund might put it, the prospect of having such interests undermined is a sufficiently dire consequence that it overcomes the prohibition against killing.) While these interests cannot be subsumed under the interest not to be killed or maimed that revisionists typically focus on, they are nonetheless interests not to suffer a sort of personal harm. And because of that, their defense is regulated by ordinary interpersonal morality rather than by a special morality of war.

For example, we might think that when we have a strong interest in making decisions together with people we are in a valuable relationship with, threatening that interest constitutes a form of personal harm that could warrant resorting to lethal defensive force (subject to necessity and proportionality constraints). This explains why, as a last resort, I might be permitted to kill you if you use violence to wrongfully prevent me from making important decisions that I have a right to make together with my family—say, about how to raise our children, which cultural identity to embrace, or where to live. If so, the same interest can be invoked to ground the permission to fight in wars of national defense. If we value making important decisions together with the members of our political community, threatening that interest constitutes a form of personal harm that could warrant resorting to lethal defensive force.[50]

49 See Hurka, "Proportionality in the Morality of War"; Frowe, *Defensive Killing*, 139–47; Fabre, "Cosmopolitanism and Wars of Self-Defence"; Stilz, "Territorial Rights and National Defence"; Moore, "Collective Self-Determination, Institutions of Justice, and Wars of National Defence"; and Renzo, "Political Self-Determination and Wars of National Defense."

50 This argument is subject to a number of caveats I cannot discuss here. See Renzo, "Political Self-Determination and Wars of National Defense," especially 717–22. At one point, in explaining why an unauthorized attempt to defend political goods would be impermissible,

Indeed, some revisionists qualify this case in a way that further weakens Estlund's argument. Cécile Fabre and Anna Stilz, for example, deny that it is permissible to take part in wars of national defense to protect purely political goods. They argue that the permission exists only if failing to defend certain political goods would likely lead to the sort of harm to life and limb that is normally required by revisionist justifications for defensive killing. These arguments thus explicitly tie the justification of wars of national defense to the value of preventing the killing or injuring of other innocents.[51]

This is why Estlund's argument is unsuccessful. Once we bear in mind the grounds on which the permissibility of wars of national defense is plausibly justified by revisionists, it becomes clear that this is not an instance of subtle justification. But if so, the comparison with national defense cannot put any pressure on them to acknowledge that genuine forms of subtle justification, such as arguments that appeal to the value of obedience, can successfully overcome the prohibition against killing.

There is an important lesson to be learned here about the different ways in which the value of protecting political goods and the value of obedience can plausibly contribute to the justification of lethal defensive killing. It is not hard to see why the value of protecting goods such as collective self-determination or political independence is the sort of thing that should figure in our proportionality calculation for the permissibility of resorting to defensive lethal force. Once we accept that we have vitally important interests in enjoying such goods, failing to protect them compromises our well-being, and this is clearly the sort of harm that can be weighed against the harm of killing innocents. It is implausible, however, to think about the value of obeying orders along the same lines.

Estlund writes that a given society "should get to decide whether a violent defense is to be mounted or not" ("The Weight of Authority in War," 787). I agree, and the argument I have just sketched explains why. A self-determining community might decide that it does not want to resist political aggression, and in that case, engaging in lethal defensive violence on its behalf would violate its right to decide for itself how to respond to the aggression—a right that, once again, is grounded in the interests of the members in making decisions together. Crucially, however, authority—understood as the power to issue binding orders—is not what is doing the work here. The right to political self-determination primarily belongs to political communities, not to states or political institutions, and thus is not inherently connected to authority. (See Renzo, "Revolution and Intervention," 237–38.) One way to see this important point is to consider situations in which a state is so unjust that it loses its right to exercise political authority, yet the people retain the right to make decisions together. Likewise, a political community might be colonized or annexed by another state and thereby lack political institutions capable of issuing binding orders; yet the people retain the right to make decisions together.

51 Fabre, "Cosmopolitanism and Wars of Self-Defence"; and Stilz, "Territorial Rights and National Defence."

Even if we accept that legitimate institutions are extremely valuable and that this generates a duty to obey the law, it is hard to see how the value of obedience could play a significant role—indeed, any role—in our proportionality calculation. Since failing to obey the law in itself does not undermine anyone's well-being or set back anyone's interests, it is implausible to think of obedience as the sort of good that can be weighed against the value of not contributing to the killing of innocents.[52] And this is the deeper reason to resist the sort of political argument for the duty to obey offered by Estlund. Given the difference in nature between the value underlying the justification of wars of national defense and the one underlying the justification of political arguments, it is a mistake to rely on the former to justify the latter.

Ultimately, this is the motivation for adopting the more formal approach to the question of obedience that I develop in my account. The reason we can have a duty to obey an order to fight an unjust war is not that the value of obedience can ever outweigh the value of not contributing to the killing of innocents. It is that we are under a moral duty to obey the law of just institutions that perform morally necessary tasks, and complying with this duty requires excluding from our deliberation certain competing moral considerations—even when these competing considerations are extremely weighty—unless stringent epistemic conditions are met.[53]

King's College London
massimo.renzo@kcl.ac.uk

REFERENCES

Buchanan, Allen. *Institutionalizing the Just War*. Oxford University Press, 2018.

Christiano, Thomas. "War, Legitimacy, and Democracy: Comments on Renzo." *Journal of Ethics and Social Philosophy* 30, no. 5 (2025): 752–59.

Estlund, David. "On Following Orders in an Unjust War." *Journal of Political*

52 Of course, disobedience can lead to the sort of harmful consequences that we can weigh against the value of not contributing to the killing of innocents. When these consequences are sufficiently serious, obedience might be morally required. But in these cases, the justification for obeying is grounded in the value of avoiding such consequences, not in the value of obedience. (As we have seen, what is required in such cases is, strictly speaking, not obedience but conformity to the directive.)

53 I am grateful to audiences at the University of Fribourg and King's College London for excellent feedback, as well as to Helen Frowe, David Owens, and Nikhil Venkatesh for helpful written comments.

Philosophy 15, no. 2 (2007): 213–34.
———. "The Weight of Authority in War: Comments on Renzo." *Journal of Ethics and Social Philosophy* 30, no. 5 (2025): 776–94.
Fabre, Cécile. *Cosmopolitan War*. Oxford University Press, 2012.
———. "Cosmopolitanism and Wars of Self-Defence." In Fabre and Lazar, *The Morality of Defensive War*.
Fabre, Cécile, and Seth Lazar, eds. *The Morality of Defensive War*. Oxford University Press, 2014.
Feinberg, Joel. *Rights, Justice, and the Bounds of Liberty*. Princeton University Press, 1980.
Frowe, Helen. *Defensive Killing*. Oxford University Press, 2014.
Hurka, Thomas. "Proportionality in the Morality of War." In *Drawing Morals: Essays in Ethical Theory*. Oxford University Press, 2011.
Kiesewetter, Benjamin. "*Pro Tanto* Rights and the Duty to Save the Greater Number." In *Oxford Studies in Normative Ethics*, vol. 13, edited by Mark Timmons. Oxford University Press, 2023.
Kutz, Christopher. *On War and Democracy*. Princeton University Press, 2016.
———. "Soldiers and Moral Tragedy: Comments on Renzo." *Journal of Ethics and Social Philosophy* 30, no. 5 (2025): 760–75.
McMahan, Jeff. "Just War." In *A Companion to Contemporary Political Philosophy*, edited by Robert E. Goodin, Philip Pettit, and Thomas Pogge. 2nd ed. Blackwell, 2007.
———. *Killing in War*. Oxford University Press, 2009.
———. "What Rights May Be Defended by Means of War?" In Fabre and Lazar, *The Morality of Defensive War*.
Moore, Margaret. "Collective Self-Determination, Institutions of Justice, and Wars of National Defence." In Fabre and Lazar, *The Morality of Defensive War*.
Owens, David. "Rationalism About Obligation." *European Journal of Philosophy* 16, no. 3 (2008): 403–31.
Raz, Joseph. *The Authority of Law*. 2nd ed. Oxford University Press, 2009.
———. *Ethics in the Public Domain*. Clarendon Press, 1994.
———. *The Morality of Freedom*. Clarendon Press, 1986.
———. *Practical Reason and Norms*. 2nd ed. Oxford University Press, 1990.
Renzo, Massimo. "Associative Responsibilities and Political Obligation." *Philosophical Quarterly* 62, no. 246 (2012): 106–27.
———. "Democratic Authority and the Duty to Fight Unjust Wars." *Analysis* 73, no. 4 (2013): 668–76.
———. "Fairness, Self-Deception and Political Obligation." *Philosophical Studies* 169, no. 3 (2014): 467–88.

———. "A Pluralist Theory of Political Obligation." In *The Oxford Handbook of Political Obligation*, edited by George Klosko. Oxford University Press, 2025.
———. "Political Authority and Unjust Wars." *Philosophy and Phenomenological Research* 99, no. 2 (2019): 336–57.
———. "Political Self-Determination and Wars of National Defense." *Journal of Moral Philosophy* 15, no. 6 (2018): 706–30.
———. "Revolution and Intervention." *Noûs* 54, no. 1 (2020): 233–53.
———. "State Legitimacy and Self-Defence." *Law and Philosophy* 30, no. 5 (2011): 575–601.
Rodin, David. *War and Self-Defense*. Oxford University Press, 2002.
Schauer, Frederick. *Playing by the Rules*. Clarendon Press, 1991.
———. "Rules and the Rule of Law." *Harvard Journal of Law and Public Policy* 14 (1991): 645–94.
Shue, Henry. *Fighting Hurt*. Oxford University Press, 2016.
Simmons, A. John. *Justification and Legitimacy*. Cambridge University Press, 2001.
Stilz, Anna. "Territorial Rights and National Defence." In Fabre and Lazar, *The Morality of Defensive War*.
Tadros, Victor. *To Do, To Die, To Reason Why*. Oxford University Press, 2020.
———. *Wrongs and Crimes*. Oxford University Press, 2016.
Thomson, Judith Jarvis. *The Realm of Rights*. Harvard University Press, 1990.
Venezia, Luciano. "Mistaken Authority and Obligation." *Legal Theory* 26, no. 4 (2020): 338–51.
Venezia, Luciano, and Rodrigo E. Sánchez Brígido. "Between Traditionalism and Revisionism: Estlund and Renzo on the Obligation to Obey Orders to Fight in Unjust Wars." *Ratio Juris* 36, no. 4 (2023): 350–65.
Walzer, Michael. *Just and Unjust Wars*. 4th ed. Basic Books, 2006.
Williams, Bernard. *Moral Luck*. Cambridge University Press, 1981.
Wolf, Susan. "Above and Below the Line of Duty." In *The Variety of Values: Essays on Morality, Meaning, and Love*. Oxford University Press, 2015.
Wolff, Robert Paul. *In Defense of Anarchism*. Harper and Row, 1970.

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org
ISSN 1559-3061

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

Funding for the journal has been made possible through the generous commitment of the Division of Arts and Humanities at New York University Abu Dhabi.

جامعة نيويورك أبوظبي
NYU ABU DHABI