JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXIX · NUMBER 1
November 2024

ARTICLES

DISCUSSION

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

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

JESP aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well argued, current, and of sufficiently general interest. *JESP* also endorses and abides by the Barcelona Principles for a Global 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. 29, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.3805

THE MORAL VIRTUE OF SOCIAL CONSCIOUSNESS

Anna Brinkerhoff

VERY BROADLY, social consciousness is a cognitive sensitivity to surrounding social injustices. We see it on display in public calls for climate action and protests against police brutality, but it is also present in the private recognition of the dire straits of the single mom next door who, despite working multiple jobs, still struggles to keep food on the table. Social consciousness is primarily a cognitive phenomenon: it is about how we *think* about social injustices. However, it has a distinct moral cast too: it is morally good, perhaps even morally required, to think in the ways constitutive of social consciousness. The goal of this paper is to develop an account of social consciousness that pays due respect to both its cognitive and moral dimensions.

To begin theorizing about social consciousness, it is helpful to note its similarity to the nearby concept of "wokeness"—that is, of being alert to racial injustices. Although wokeness may be a more familiar concept and is addressed in the relevant literature, I have chosen to focus on social consciousness for a few different reasons. First, there remain worries about epistemic appropriation and misuse whenever terms and concepts originating in marginalized communities are detached from communities—worries that were discussed as early as 1962 by William Melvin Kelley and have been developed more recently by Emmalon Davis.[1] Second, given its etymology, wokeness typically regards cognitive sensitivity to social injustices that are suffered specifically by Black people.[2] This paper aims to account for cognitive sensitivity to social injustices suffered not only by Black people but also by members of other historically marginalized social groups (including women, First Nations, the LGBTQ+ community) as

1 Kelley, "If You're Woke You Dig It"; and Davis, "On Epistemic Appropriation." For a helpful analysis of epistemic detachment, see Pollock, "Political Action, Epistemic Detachment, and the Problem of White-Mindedness."

2 The term 'woke' traces back to Black thinkers and communities in the 1920s but entered contemporary mainstream discourse after the 2014 killing of Michael Brown at the hands of police in Ferguson, Missouri. The term continued to gain prominence throughout the Black Lives Matter movement. For more on the etymology of 'woke' and 'wokeness', see Romano, "A History of Wokeness."

well as individuals at the intersection of multiple marginalized groups. Finally, the term 'woke' has become increasingly politicized, which threatens to cloud efforts to account for the related concept clearly and accurately.

That said, given the similarities between wokeness and social consciousness, any philosophical account of the former can be modified to apply to the latter. After all, both wokeness and social consciousness are forms of cognitive sensitivity to social injustices. Accounting for either of them amounts to spelling out what exactly this cognitive sensitivity amounts to—what cognitive states constitute the relevant sensitivity—and how exactly moral values govern or otherwise relate to those states.

With that in mind, consider Rima Basu's recent suggestion that wokeness can be understood through the lens of moral encroachment.[3] Moral encroachment is an epistemological view according to which moral considerations get a say in what is epistemically rational to believe. On Basu's view, the cognitive sensitivity at the center of wokeness amounts to believing in accordance with the dictates of moral encroachment, which is a moral duty.

After extending this view to social consciousness—I call it the *encroachment account of social consciousness*—I raise a few worries about it: not only does it involve controversial theoretical commitments, but it also entails unintuitive verdicts in relevant cases and implies that social consciousness is very (maybe even excessively) demanding. In light of these worries, I develop an alternative account of social consciousness: the *virtue account*. Taking cues from Nomy Arpaly's discussion on open-mindedness as a moral virtue, I suggest that the cognitive sensitivity at the center of social consciousness is better understood as a morally virtuous cognitive disposition that manifests itself primarily in certain doxastic states.[4] I argue that the virtue account not only weathers the worries that trouble the encroachment account but also captures several important features of social consciousness better than the encroachment account.

Before moving ahead, I want to pause on the starting assumption that social consciousness is primarily cognitive. Some may object that social consciousness is about acting, not just believing, in certain ways. To be socially conscious, we must fight against and redress social injustices. Belief without action is hollow. In response, it is worth emphasizing that what we believe heavily shapes how we act: thinking in the ways constitutive of social consciousness, on any account, will tend to lead to certain actions. So if someone fails to act in ways that we would expect a person with the relevant beliefs to act—or if they act in ways that seem to conflict with those beliefs—this gives us reason

3 Basu, "Radical Moral Encroachment," 17.

4 Arpaly, "Open-Mindedness as a Moral Virtue."

to doubt that they have the relevant beliefs or else doubt that those beliefs are rightly related to morality. For example, if an employer professes antiracism yet consistently hires and promotes white folks over equally qualified Black folks, then we have reason to doubt that the employer actually believes and values what she professes.

That said, for those who maintain that action is more central to social consciousness than the assumption allows, I invite you to take the following project as an attempt to account for the cognitive side of social consciousness rather than its whole.

1. The Encroachment Account

Rima Basu has recently suggested that we can understand "moral encroachment as a systematic treatment of the imperative to stay woke."[5] This suggestion points us toward the encroachment account of social consciousness. To get a good grasp on this account, we must get a good grasp on moral encroachment.

1.1. Moral Encroachment

Very broadly, moral encroachment is the view that morality gets a say in what is epistemically rational to believe.[6] Specifically, moral encroachment says that moral considerations help set the evidential threshold that a belief must pass in order to be epistemically rational. In cases where the moral stakes for a belief that *p* are high, the believer tends to need stronger evidence in support of *p* for the belief to be epistemically rational.

It is helpful here to consider the cases of high moral stakes that advocates of moral encroachment point to in order to motivate it. These cases tend to share similar features: a believer *S* infers something about an individual *J* based on statistical information about *J*'s social group *G*; *G* has been historically marginalized; and *S*'s socio-epistemic environment has been shaped by prejudiced attitudes and practices that negatively affect *G*. Consider the following cases.

5 Basu, "Radical Moral Encroachment," 19. It is worth noting that Basu does not set out to develop an account of wokeness. Her remarks about wokeness are made mainly in passing as she develops and defends moral encroachment.

6 For recent defenses of moral encroachment, see Basu, "What We Epistemically Owe Each Other," "The Wrongs of Racist Beliefs," "Radical Moral Encroachment," and "Can Beliefs Wrong?" See also Basu and Schroeder, "Doxastic Wronging"; Bolinger, "The Rational Impermissibility of Accepting (Some) Racial Generalizations"; Fritz, "Pragmatic Encroachment and Moral Encroachment"; Moss, "Moral Encroachment"; and Schroeder, "When Beliefs Wrong." For a helpful taxonomy of moral encroachment view, see Bolinger, "Varieties of Moral Encroachment." In this paper, I focus mostly on the Basu and Schroeder version of moral encroachment.

> *Server*: Spencer works as a server at a restaurant. He senses that white diners tip more than Black diners. Doing a bit of research online, Spencer finds a well-documented social trend that Black diners tip substantially below average. Spencer weighs the evidence before reaching his belief that Black diners tip substantially below average. A Black diner, Jamal, enters Spencer's restaurant. Spencer believes that Jamal will probably tip below average.[7]

> *Teacher*: Stacy is a fifth-grade teacher at a public elementary school. It is the first day of school, and she is meeting her students for the year for the first time. Two new students, Jenna and Joel, walk in. Stacy knows that on average, girls consistently score lower than boys on standardized math exams. In light of this, Stacy comes to believe that Jenna probably scored lower than Joel on last year's statewide standardized math exam.

Advocates of moral encroachment focus on the inferential belief in these cases: Spencer's belief that Jamal probably tips less than average and Stacy's belief that Jenna probably scored lower than Joel on last year's statewide math exam.[8] They then point to numerous moral features of these cases that explain why the moral stakes for these beliefs are high. Basu divides these moral features into three categories: upstream features, downstream features, and features of the belief itself.[9]

Upstream moral features regard the way the beliefs are formed. In these cases, the relevant statistical facts are true in part because the social group in question has been historically oppressed. With regard to Server, there are two main sociological explanations of racial disparity in tipping, both of which trace back to anti-Black racism: (1) because of low levels of income that result from being subject to a long legacy of anti-Black racism, Black people do not dine out at full-service restaurants as much and so are not as familiar with

7 This is an adapted version of a case introduced by Basu, "The Wrongs of Racist Beliefs." The adaption is from Gardiner, "Evidentialism and Moral Encroachment."

8 Throughout the paper, I will qualify the relevant inferential beliefs with *probably* or *likely*. Moral encroachment is motivated largely by cases of seemingly problematic beliefs that are supported by the believer's evidence and thus rational according to traditional theories of epistemic rationality. But *un*qualified inferential beliefs—"Jamal tips less than average" or "Jenna scored lower than Joel on last year's math exam"—may automatically go beyond the evidence. Evidence suggesting that most members of a set have some property does not firmly suggest that a randomly selected member has that property, but it does firmly suggest that a randomly selected member likely or probably has that property. So to ensure that the beliefs in question are supported by the believers' evidence, it is important to qualify the relevant inferential beliefs.

9 Basu, "The Wrongs of Racist Beliefs."

percentage-based tipping norms; or (2) Black diners are systemically discriminated against by servers.[10] With regard to Teacher, empirical studies suggest that stereotypes related to gender and mathematical ability negatively affect girls' performance in competitive testing environments.[11] Some advocates of moral encroachment suggest that the moral stakes of inferential beliefs like Spencer's and Stacy's are raised because the evidence on which they are based is ultimately a result of racism (or sexism or some other form of prejudice).[12]

The moral stakes are even higher when this upstream moral feature is combined with other moral considerations about harmful risks and costs posed by the beliefs in question. Some of these harms are downstream and regard potential actions that the beliefs may lead to. For example, Stacy's belief about Jenna might lead her (perhaps subconsciously) to overlook or fail to foster Jenna's mathematical talent. What's more, these beliefs contribute to collective harms from which the targeted individual and social group suffer. For example, many servers believing that Black diners tip substantially below average leads to systematically poor service to Black diners, which in turn discourages Black patronage and exacerbates the ills of social segregation; such beliefs being prevalent also makes it harder for restaurants to retain servers in areas with a large percentage of Black patrons, which makes owners averse to opening restaurants in Black communities. Both the risk of harmful actions and the risk of collective harm posed by the beliefs in question raise the moral stakes for the beliefs in Server and Teacher.

Other harms arise because of features of the beliefs themselves. The properties ascribed to Jamal and Jenna—likely tipping or scoring below average—"bring them down."[13] They are also potentially demeaning and offensive: if Jamal or Jenna found out about these beliefs, they would probably feel hurt, or, at least, feeling hurt would be an apt response for them to have. These features of the belief itself are also thought to raise the moral stakes in cases like Server and Teacher.[14]

In sum, the moral stakes are high for the inferential beliefs in question in cases like Server and Teacher because of various moral features. The high moral stakes in these cases raise the evidential threshold that the beliefs in question

10 For a helpful overview and critical discussion of the relevant sociological literature, see Brewster and Mallinson, "Racial Differences in Restaurant Tipping."

11 For a helpful overview of the role that stereotypes play in explaining the gender gap in math test scores, see Niederle and Vesterlund, "Explaining the Gender Gap in Math Test Scores."

12 Basu, "Radical Moral Encroachment," 14–15.

13 Schroeder, "When Beliefs Wrong," 124.

14 Basu, "What We Epistemically Owe Each Other," 920.

must pass in order to be rational. Group-level information about Black people and girls is not enough to push the relevant beliefs over this high evidential threshold. Beliefs that fail to pass the relevant evidential threshold are epistemically irrational. So according to moral encroachment, Spencer's and Stacy's respective beliefs are epistemically irrational, despite the group-level evidence they have in support of them.

1.2. The Encroachment Account of Social Consciousness

With a grasp on moral encroachment, we can better understand the thought that the cognitive sensitivity at the center of social consciousness amounts to something like abiding by the dictates of moral encroachment. It is worth quoting Basu here at length:

> [We can] understand moral encroachment … as the demand to stay woke. To be woke is to be aware of the moral demands of one's environment. With regard to our epistemic practices, it is the demand to be aware of the moral stakes of our beliefs about one another. [The demand to be woke] is the demand to be aware of the background against which our epistemic practices exist, i.e., the unjust world we inhabit, and to ensure that our epistemic practices are not only responsive to unjust features of our environment but that they also do not themselves contribute to those unjust features of our environment.[15]

In combination with Basu's view of moral encroachment, this passage points us toward an encroachment account of social consciousness that goes something like this:

Our socio-epistemic environment has been shaped by a long history of social injustices in ways that "stack the evidence" in favor of prejudiced beliefs.[16] As Basu puts it, "Facts may not be racist, but they may be products of racism," and so when reasoning and forming beliefs on the basis of such facts, "we must not ignore their provenance."[17] Given this, there is a moral duty to approach evidence and beliefs about marginalized social groups and the individuals that belong to them with extra care and sensitivity.[18] This is especially the case when the beliefs that may result pose harm to the individual and social group in question, thereby compounding the social injustices they already suffer.

15 Basu, "Radical Moral Encroachment," 17.

16 Basu, "The Wrongs of Racist Beliefs," 2497.

17 Basu, "Radical Moral Encroachment," 14.

18 Basu, "Radical Moral Encroachment," 15.

According to the encroachment account of social consciousness, this moral duty of extra care and sensitivity just is a moral duty to be socially conscious. When it comes to our "epistemic practices"—and how to specify the cognitive sensitivity at the center of social consciousness—being socially conscious requires us to believe in accordance with the dictates of moral encroachment. So in order to satisfy the moral duty to be socially conscious, we must not form or maintain beliefs that fail to pass an evidential threshold raised high by the sort of moral considerations found in Server and Teacher.

On the encroachment account, beliefs like Spencer's and Stacy's that fail to pass an evidential threshold set high by the relevant sort of moral considerations are not only epistemically irrational; they are also morally impermissible. After all, such beliefs constitute a violation of the moral duty to be socially conscious. So by virtue of believing the way they do about Jamal and Jenna, Spencer and Stacy are condemnable from both an epistemic standpoint (for being epistemically irrational) and a moral standpoint (for violating a moral duty).

2. WORRIES ABOUT THE ENCROACHMENT ACCOUNT

Despite its initial appeal, the encroachment account of social consciousness faces some worries.

2.1. Worry One: Controversial Theoretical Commitments

The first worry concerns the theoretical commitments entailed by the encroachment account. Moral encroachment itself is controversial.[19] For one, it goes against the traditional thought that epistemic rationality is determined alone by evidential and other truth-related considerations.[20] It may also be worried that morality (which is complex and multifaceted) cannot map cleanly onto epistemic rationality (which is rather cut and dry) as moral encroachment implies.[21] People who reject moral encroachment as a theory of rationality for these or other reasons will also have to reject it as a basis for an account of social consciousness.

The encroachment account also has an unsavory moral commitment: namely, it renders believers morally condemnable for believing something on the basis of good reason for thinking it is true—at least, reason that is good enough to rationalize beliefs in many if not most contexts.

19 For critical discussions of moral encroachment, see Begby, "Doxastic Morality"; Gardiner, "Evidentialism and Moral Encroachment" and "Against the New Ethics of Belief"; and Brinkerhoff, "Prejudiced Beliefs Based on the Evidence."

20 Conee and Feldman, *Evidentialism*.

21 Gardiner, "Against the New Ethics of Belief."

2.2. Worry Two: Troubling Verdicts in Relevant Cases

What is more, the encroachment account of social consciousness is committed to troubling or at least unintuitive verdicts in a set of important cases. These are variations on cases like Server and Teacher in which the inferential beliefs in question are couched within a robust understanding of the social injustices that have shaped the believer's socio-epistemic environment—specifically, the social injustices that led to the relevant statistical facts. Consider the following variations on Server and Teacher.

> *Informed Server*: Spencer knows that statistically, Black diners tip substantially below average and inferentially comes to believe that Jamal will likely tip below average. Spencer has recently read a lot about the historic and continued oppression of Black Americans. So in addition to the relevant statistical information, Spencer knows that Black Americans have been disadvantaged by structural racism for centuries in a multitude of ways that have negatively impacted Black communities. For one, it has led to systematic income inequality between Black Americans and white Americans. From his research, Spencer knows that it is this income inequality, not any vice or lack of virtue, that ultimately explains the tipping patterns of Black Americans.

> *Informed Teacher*: Stacy knows that statistically, girls tend to score lower on standardized math exams than boys and infers that Jenna likely scored lower than Joel on last year's statewide standardized math exam. Stacy has recently done a lot of research about the gender gap in mathematical achievement. In addition to the relevant statistical information, Stacy knows that women and girls have been historically characterized in ways that impugn their mathematical abilities. From her research, Stacy knows that the gender disparity in math performance is explained not by a lack of rationality or analytic prowess in girls and women but rather by the ways that negative stereotypes about women and math negatively affect girls' math performance.

According to the encroachment account of social consciousness, Informed Server and Informed Teacher are paradigmatic cases of a failure to be socially conscious. After all, many of the moral features that raise the moral stakes in the original cases carry over to these variations. This means that Informed Spencer's belief about Jamal and Informed Stacy's belief about Jenna must pass a high evidential threshold. According to moral encroachment, the informed believers' group-level evidence is not sufficient to push their respective inferential beliefs over this threshold, and so they are epistemically irrational. The

beliefs are also morally impermissible, rendering Informed Spencer and Informed Stacy morally condemnable. By virtue of believing in the ways they do, Informed Spencer and Informed Stacy violate the moral duty to be socially conscious. Their respective understandings of the socio-epistemic landscape do not exonerate them.

The worry is that these verdicts about Informed Server and Informed Teacher seem mistaken. At least to me, Informed Spencer and Informed Stacy and their inferential beliefs seem to be both epistemically and morally in the clear.[22] In fact, it might be thought that Informed Spencer's and Informed Stacy's respective beliefs are characteristic of—not contrary to—social consciousness. The encroachment account, then, is committed to troubling or at least unintuitive verdicts in cases like these.

2.3. Worry Three: Demandingness

The final worry is about how demanding social consciousness is on the encroachment account: abiding by the dictates of moral encroachment is *excessively* demanding; more precisely, it is too demanding to be something that morality requires.

It is widely thought that we lack voluntary control over what we believe: what we believe is largely an involuntary response to our evidence.[23] If that is right, then it may often be nearly psychologically impossible for us to believe in the way that is required by moral encroachment, especially in cases like Server and Teacher in which moral encroachment requires us not to believe something that is seemingly supported by the evidence.

Setting aside the difficulties that arise from doxastic involuntarism—and even supposing that beliefs are not merely at the mercy of the evidence—abiding by the dictates of moral encroachment would still be very difficult: it involves a fair amount of intellectual sophistication to discern when the moral stakes for a belief are high and, by extension, when the sort of evidence that is

22 By saying that these beliefs seem morally in the clear, all I mean is that there seems to be nothing inherently morally wrong with the beliefs themselves. I do not mean to imply that it is morally permissible for either Informed Spencer or Informed Stacy to act on them in ways that might disadvantage Jamal or Jenna. In fact, it seems clear to me that doing so would be morally impermissible. For example, it would be morally impermissible for Spencer to give Jamal poorer service in light of his inferential belief. See section 4 of Brinkerhoff, "Prejudiced Beliefs Based on the Evidence" for a discussion about how believers in cases like The (Informed) Server and The (Informed) Teacher may be morally condemnable even if relevant beliefs themselves are in the clear.

23 For canonical contemporary discussions of doxastic involuntarism, see Alston, "The Deontological Conception of Epistemic Justification"; and Audi, "Doxastic Voluntarism and the Ethics of Belief."

normally sufficient to rationalize a belief does not cut it. This sort of sophistication may be out of reach for many people in many contexts. More generally, it will be difficult for anybody—no matter their socio-epistemic environment, cognitive skills, or education level—to believe differently in different contexts based on similar evidence.

The fact that social consciousness is very demanding on the encroachment account does not alone give rise to the worry that it is *excessively* demanding. After all, as Basu emphasizes, being morally good is difficult, and so we should expect social consciousness to be difficult too.[24] The worry arises because on the encroachment account, being socially conscious is a moral duty; we are thereby morally in the wrong and blameworthy for having beliefs proscribed by moral encroachment. Social consciousness, the worry goes, may be morally good and important, and we morally should promote, pursue, and praise it. But perhaps it is too demanding to be something morality *requires*.

Appeals to doxastic involuntarism have often been used to argue against views that imply that there are (moral) duties or obligations on belief.[25] But even if these arguments can be successfully countered, there remain worries about moral duties to believe in accordance with the dictates of moral encroachment given difficulties arising from the required intellectual sophistication as well as the limits of our socio-epistemic environments.

To sharpen these worries, it is helpful to consider cases in which the socio-epistemic environment is even more impoverished than our own. Consider this case adapted from Arpaly.[26]

> *Farmboy*: Solomon believes that most women are bad at abstract thinking or, at least, not half as good as men. He was born and raised in a small, isolated farming community in a poor country, where this belief is not only assumed by everyone around him but also confirmed by his everyday interactions. Women in his community talk exclusively about family matters and gossip, even when Solomon tries to talk with them about morality and religion; the few people in his community who engage in abstract thinking are men; and works of abstract thought in the community's outdated library are authored solely by men. When Solomon meets Joyce, his new neighbor, he comes to believe that she is likely bad at abstract thinking.

24 Basu, "Radical Moral Encroachment," 19.

25 See the discussion of the problem of control in Basu and Schroeder, "Doxastic Wronging."

26 Arpaly, *Unprincipled Virtue*, 105–6.

According to the encroachment account, Solomon's belief about Joyce is not only epistemically irrational but also morally impermissible since it violates Solomon's moral duty to be socially conscious. By virtue of having it, Solomon is morally blameworthy. But this verdict does not seem quite right. As Endre Begby writes in a discussion of similar cases, "We will want to make room for the idea that people who grow up in deeply prejudiced settings with no rational access to contrary evidence should in some sense be counted as victims too."[27] Solomon is in an unfortunate evidential situation, through no fault of his own. He seems misinformed, not blameworthy or even prejudiced. Perhaps Solomon has room for moral improvement when it comes to his beliefs, but given the poverty of his socio-epistemic environment, it is doubtful whether his current doxastic states render him morally condemnable. After all, his belief about Joyce is based on good reason for thinking it is true.

By holding Solomon morally accountable for his belief about Joyce, the encroachment account leaves no room to count believers like Solomon as being hindered or excused by his socio-epistemic environment. This points us back to the demandingness worry: being socially conscious (when understood as requiring us to abide by the dictates of moral encroachment) is too demanding to be something that morality requires, especially in light of the constraints of our socio-epistemic environments.

3. The Virtue Account of Social Consciousness

Although none of these three worries is decisive against the encroachment account, they mount a significant case against it when taken together. It is worth looking for another account of social consciousness, then, that does not face similar worries. The account I have in mind gets off the ground with two ideas. The first idea is that social consciousness is a moral virtue (instead of, more narrowly, a moral requirement). The second idea is that the doxastic states described in Informed Server and Informed Teacher are characteristic of social consciousness, not inimical to it. This gives us a good place to start. Perhaps the cognitive sensitivity central to social consciousness amounts to something like a morally virtuous cognitive disposition—and the corresponding doxastic states—to recognize and remain alert to surrounding social injustices.

3.1. Another Moral-Doxastic Virtue: Open-Mindedness

To flesh out these ideas, it is helpful to look at another account of a moral-epistemic virtue: open-mindedness.

27 Begby, "Doxastic Morality," 168.

Open-mindedness is a cognitive trait that disposes us to "gain, lose, and revise beliefs in a particular reasonable way."[28] It is expressed or manifested in various doxastic states. For example, we see open-mindedness in a parent who, against his religious convictions, changes his mind about gay marriage when he observes his child thriving in a same-sex relationship; we also see it in a scientist who, after years of defending her pet hypothesis, rejects it upon encountering strong new evidence against it.

In her account of open-mindedness, Arpaly assumes along with Aristotle that moral virtues and expressions of moral virtue necessarily stem from moral concern.[29] Moral concern amounts to desiring or caring about the right or the good and so boils down to various morally good affective states—for example, caring about the well-being of others or desiring to see them flourish. Consider the moral virtue of charity. Charity is a trait that disposes us to act in ways that benefit people faring poorly out of concern for their well-being. In order to be genuine expressions of charity, actions that benefit others who are faring poorly—for example, donating a large sum of money to a nonprofit organization—must be done with an intention of helping those in need. Donating a large sum of money with an intention of boosting one's own reputation is neither morally virtuous nor a genuine expression of charity.

Expressions of moral virtues are typically thought to be actions. The interesting thing about open-mindedness—and the thing that makes it relevant to our discussion of social consciousness—is that it is a cognitive disposition the manifestations of which are primarily doxastic states rather than actions. It might be wondered how expressions of moral virtue that are *doxastic* can stem from moral concern. After all, what makes it the case that a particular action expresses moral concern is that it is done with a morally good intention—for example, an intention to help those in need. But assuming that believing is largely involuntary, believing is not intentional in the relevant sense. Given this, it may seem puzzling how doxastic states can be expressions of moral concern or, by extension, how a cognitive disposition can be a moral virtue.

Even so, Arpaly argues, we can make sense of open-mindedness as a moral virtue and its doxastic states as genuine expressions of moral virtue. That is because beliefs can also stem from moral concern, although in a more indirect way than actions can: our concerns—including moral ones—affect what we come to believe indirectly by affecting our emotions, attention, ability to learn, and the conclusions we draw. Let us consider each in turn below.

28 Arpaly, "Open-Mindedness as a Moral Virtue," 75.

29 Arpaly, "Open-Mindedness as a Moral Virtue," 75.

3.2. The Effect of (Moral) Concerns on Belief

First, our concerns influence our emotions, which in turn affect what we believe. To borrow examples from Arpaly, "If you are infatuated with a woman, you might be blind to her faults, and if you are angry at a man, you might be blind to his virtues. . . . If you are afraid of your teacher, you might overestimate his height."[30]

Moral concerns also affect our emotions and thereby our beliefs. Out of moral concern, we may feel guilt for having broken a promise, anger at an injustice, or joy when another succeeds despite great hardship. This guilt, anger, and joy can in turn affect our beliefs. Joy at the success of another, for example, can influence our view about the good things in life.

Second, our concerns affect how we direct our attention, and this in turn affects what we believe. If I care about music, I will tend to notice what songs are playing in the background of the coffee shop and form corresponding beliefs ("This is 'Come Together' from *Abbey Road*") that I would otherwise lack. Similarly, if you are a gastronome, you will tend to pick up on subtle flavors and ingredients and form corresponding beliefs ("This soup has rosemary and a hint of sage") that people with less discriminating tastes lack. What is more, our concerns affect what we turn our attention *away* from in ways that affect our beliefs. If I am worried about my generation's obsession with celebrity culture, I may turn my attention away from the tabloids in the checkout line and thus lack beliefs that I would otherwise have if I flipped through their pages.

Moral concerns also affect our attention and thereby our beliefs. Out of moral concern, we might be more attentive to the needs and interests of others and to various moral features of our environment, and what we notice affects our beliefs. Upon noticing that a student is extremely shy, for example, his teacher may come to believe that there are better ways of encouraging him to participate than cold-calling on him in class. And a manager of a nursing home may come to believe that investing in therapy dogs is better than investing in new chairs for the dining room after he notices how much interacting with pets lifts the spirits of his residents. What is more, out of moral concern, we may turn our attention *away* from some things in ways that affect our beliefs. If I care about my friend's privacy, I will turn my attention away from the text she is furiously typing on her iPhone, and I will thus lack beliefs that I would otherwise have about its contents.

Third, our concerns affect our ability to learn and what we retain. If I care more about American politics than military history, for example, I will more readily learn and retain information about American politics than military

30 Arpaly, "Open-Mindedness as a Moral Virtue," 77.

history. Arpaly explains this in two ways.[31] First, I will be more likely to "do my homework" when it comes to American politics—studying, researching, and keeping up with political news—and this will in turn lead to a body of relevant beliefs that I otherwise would lack if I slack off instead. Second, given my interest, I will be more likely to remember the information—and retain corresponding beliefs—that I learn about American politics than about military history.

Moral concerns also affect our ability to learn and thereby affect our beliefs. For example, if you care about helping those in need, you will likely "do your homework" about which charities maximize the utility of your donations and thereby form corresponding beliefs. And if a boss cares about his employees' comfort and interests, he is more likely to remember that most of them prefer the office thermostat to be set higher than what he personally prefers.

Fourth, our concerns affect what conclusions we draw and how much confidence we have in them. If I care about getting something right (or about not getting something wrong), then I will tend to be more careful when reasoning about the relevant evidence—I might be more cautious in drawing conclusions and more skeptical about the conclusions I do draw. For example, if I care deeply about making a good impression during a big presentation at work, I will be less likely to conclude that the presentation is at noon based on a fuzzy memory of my boss saying so in an email last week.

Moral concerns also affect our conclusions. If an airplane mechanic cares about the well-being of the passengers on the planes under her care, she will be less likely to conclude that the plane is ready to fly without first double-checking the relevant evidence. And if you care about the well-being of a colleague with a peanut allergy, you will be extra careful before concluding that the cookies you are about to serve her are peanut-free.

3.3. Social Consciousness as a Moral Virtue

We have just seen how moral concerns can affect our beliefs indirectly by affecting our emotions, attention, ability to learn, and conclusions. With that in mind, we can now return to the suggestion that social consciousness is a moral virtue.[32]

Social consciousness, like open-mindedness, is a cognitive disposition. Roughly, it is the cognitive disposition to recognize and remain alert to surrounding social injustices. As a moral virtue, social consciousness is necessarily

31 Arpaly, "Open-Mindedness as a Moral Virtue," 77–78.

32 I do not intend for the virtue account of social consciousness to entail a commitment to virtue ethics as the correct normative ethical theory. The existence of moral virtues, I take it, is consistent with a variety of normative ethical theories.

rooted in moral concern—specifically, concerns for the well-being and interests of those who suffer the social injustice in question.[33] If I have a disposition to recognize and remain alert to gender injustices because I am a sociologist collecting data for my new book on sexism in the workplace, then it is not clear that this disposition is a moral virtue.[34]

The virtue of social consciousness is expressed through various doxastic states related to social injustices. Socially conscious people will tend to have true beliefs about the existence of social injustices—about their history, legacy, and continuing impact—that others lack. They will also tend to have beliefs about instantiations and effects of social injustices presently occurring and the mechanisms through which they are perpetuated. In addition, socially conscious people will tend to lack false beliefs that stereotype marginalized social groups in ways that prop up or reinforce social injustices. They will also tend away from defaulting to readily available but mistaken explanations of statistical or group-level information about such groups—for example, that Black diners tip less than average because they are less generous.

In order to be genuine expressions of social consciousness, these doxastic states must stem indirectly from moral concerns. That is, the doxastic states of socially conscious people are genuine expressions of moral virtue to the extent that these states result indirectly from the ways that their moral concerns have affected their emotions, attention, ability to learn, and conclusions. In socio-epistemic environments riddled with social injustices, the socially conscious person's moral concerns may involve caring generally about the

33 This part of the virtue account can help explain why "the performativity of wokeness" (or, more relevantly, "the performativity of social consciousness") is, at the very least, morally hollow. To the extent that public professions of beliefs and values characteristic of social consciousness are motivated by something other than moral concern for those who suffer the relevant injustices—whether it be an individual's desire to garner a reputation as someone who cares about "the right things" or to avoid criticism for failing to be sufficiently "woke," or a company's desire to attract the business of a demographic who tends to care about social injustices—these professions are, at the very least, not morally virtuous.

34 Details matter here. Perhaps I chose to research gender injustice because I was socially conscious in the first place and wanted to study something morally worthy. In this case, it is plausible that my disposition to notice gender injustices is ultimately rooted in moral concern for those suffering the injustices. So to make this example work, we need to imagine that this cognitive disposition is ultimately rooted not in moral concern but in something like personal career ambition. Perhaps I chose to research gender injustice in the workplace not because I was already socially conscious but mainly because it was a "hot topic" garnering lots of attention in my field while I was in grad school, and focusing on it made my prospects on the job market more promising. In this case, it is plausible that my cognitive disposition to notice gender injustices is not a moral virtue. Thanks to an anonymous reviewer for pointing this out.

flourishing of marginalized social groups and particularly about individuals who have been harmed by social injustices, wanting them to see justice and equal opportunity, and desiring the eradication of social injustices.

Out of moral concern, the socially conscious person will have emotional responses to the social injustices in her environment that go on to influence her beliefs. For example, she may feel anger at the unjust killing of George Floyd, and this anger may inform her beliefs about the urgency of police reform. Or she may feel admiration during the confirmation of Ketanji Brown Jackson to the Supreme Court, and this admiration may inform her beliefs about the importance of better representation in institutions of power.

Out of moral concern, the socially conscious person will be more likely to notice social injustices around her, which will lead her to form beliefs that she otherwise would not have about the existence of social injustices and their myriad instantiations. For example, a socially conscious admissions counselor may notice how legacy preferences disproportionately disadvantage applicants from low-income families, who are less likely to have a parent with a university degree. And a socially conscious Black teenager may notice that several of her Black neighbors but none of her numerous drug-using white peers are imprisoned on nonviolent drug charges. What is more, a socially conscious person will be more likely to turn her attention away from things that promote negative stereotypes about marginalized social groups, which will make her less likely to form corresponding beliefs. For example, a socially conscious person may choose to unfollow a friend who regularly tweets stigmatizing messages about Muslim immigrants or to ignore sitcoms that habitually portray women as ditzy and incompetent.

Out of moral concern, the socially conscious person will be more interested in issues involving social injustices. This in turn will lead her to "do her homework" about such matters, listen to the victims of the injustice, and remember what she has learned. For example, caring about the flourishing of Black communities, a socially conscious person may look into and remember information about the ways that historical redlining practices in the housing sector combine with current practices for funding local schools to systemically disadvantage Black students.[35] For another example, caring for the well-being of the global poor, a socially conscious person may research and remember how climate

35 "Research" will look different depending on the social position of the socially conscious person. If the socially conscious person belongs to the marginalized community in question, "research" may involve observing and reflecting on her surroundings and conversing with friends and family instead of reading books and doing Google searches. This is discussed further in section 5.3 below.

change will disproportionately harm those who are already the most economically disadvantaged.

Finally, out of moral concern, a socially conscious person may be more cautious when it comes to reasoning about marginalized social groups and their members. Wanting to get it right, a socially conscious person will be unlikely to draw hasty generalizations about social groups or individuals based on information about their apparent social groups. For example, a socially conscious person will be unlikely to conclude that girls are inherently worse at math upon encountering statistical information about gender disparities in mathematical achievement. A socially conscious person will also be less likely to commit other fallacies when reasoning about marginalized social groups, such as overestimating base rates of felonies among Black men. And despite knowing that most women in a particular office building are employed as administrative assistants, a socially conscious person may be wary of concluding that the woman he sees walking down the office hallway is an administrative assistant.

In sum: according to the virtue account, the cognitive sensitivity at the center of social consciousness amounts to a morally virtuous cognitive disposition to recognize and remain alert to surrounding social injustices. Like all moral virtues, this disposition is rooted in moral concern. Unlike most moral virtues, social consciousness is expressed primarily through doxastic states rather than through actions. In order to be morally virtuous, the doxastic states characteristic of social consciousness must flow from moral concern. Beliefs flow from moral concerns—not from morally good intentions—but through being indirectly affected by moral concerns. A socially conscious person's moral concerns affect her emotions, attention, ability to learn, and conclusions in ways that ultimately result in the doxastic states characteristic of social consciousness.

4. Weathering the Worries

We now have two competing accounts of social consciousness on the table: the encroachment account and the virtue account. In this section, I want to consider how the virtue account fares in light of the three worries that trouble the encroachment account. I argue that these worries simply do not arise for the virtue account when it is spelled out.

4.1. Weathering Worry One: Controversial Theoretical Commitments

The first worry for the encroachment account is that it entails a number of controversial theoretical commitments involved in affirming moral encroachment about epistemic rationality. The virtue account does not share these

commitments because it does not hinge on any specific view about epistemic rationality. This means that the virtue account can be accepted by both advocates of moral encroachment and those who reject it. The virtue account also does not have the troubling implication that people can be morally blameworthy for believing something based on good reasons for thinking it is true. On the virtue account, believers like Spencer and Stacy in the original versions of their cases may not be socially conscious—as far as we know, they do not have the doxastic states characteristic of social consciousness—but they are also not morally (or epistemically) condemnable by virtue of having the inferential beliefs in question.

4.2. Weathering Worry Two: Troubling Verdicts

The second worry for the encroachment account is that it renders unintuitive verdicts in relevant variations on the cases that motivated moral encroachment—cases like Informed Server and Informed Teacher. The encroachment account entails that Informed Spencer's and Informed Stacy's respective inferential beliefs are epistemically irrational, morally impermissible, and manifestations of a moral failure to be socially conscious. Intuitively, though, these beliefs seem to be both epistemically and morally in the clear, and their broader set of doxastic states seems characteristic of social consciousness rather than contrary to it.

The virtue account recommends a different set of verdicts about Informed Server and Informed Teacher and their respective beliefs. First, the virtue account does not entail that Informed Spencer's and Informed Stacy's inferential beliefs are morally or epistemically bad, or contrary to social consciousness. Rather, on the virtue account, the doxastic states in these cases are indeed characteristic expressions of social consciousness: Informed Spencer and Informed Stacy have true beliefs about the history, legacy, and current instantiations of surrounding social injustices, and they lack false beliefs involving negative stereotypes or explanations of group-level information. That said, we do not have quite enough information about Informed Spencer and Informed Stacy to render a firm judgment about whether they are socially conscious. That is because we do not know if their doxastic states flow from moral concern or something else.

Consider the two following ways that Informed Teacher could be fleshed out. In both variations, assume that Stacy has the same doxastic states detailed in Informed Teacher.

> *Morally Virtuous Informed Teacher*: Stacy cares deeply about the well-being of her students and desires to tailor her teaching to each student's

unique needs, interests, and skills in order to help them each to flourish. She was troubled upon learning about the gender gap in math achievement and started researching it. Stacy relies on the beliefs that result from her research to better serve her students. Stacy is now especially intentional about combating negative stereotypes regarding girls and math and about encouraging mathematical prowess in individual girl students when she sees it. She is saddened that negative stereotypes and gender norms have steered many women away from educational and career pursuits in STEM and becomes heartened whenever she hears of a woman flourishing in STEM. When Stacy notices one of her girl students struggling in math, she does not automatically conclude that that student is better suited for another academic subject or that she has below-average math skills.

Nonvirtuous Informed Teacher: Stacy is working toward her master's degree in education and is currently enrolled in a mandatory sociology course about gender and education. The professor has recently covered sociological studies about the gender gap in mathematical achievement and has notified the class that this material will feature prominently in the upcoming midterm. Stacy's beliefs about the gender gap in math performance ultimately flow from a desire to ace the midterm rather than from a desire to help her students flourish. She does not use this knowledge to make changes to her teaching or to better serve her students. Soon after the semester ends, Stacy forgets most of what she learned in the course.

In both variations, Stacy understands that it is ultimately sexist stereotypes that make it so that girls tend to score lower on average than boys on standardized math exams. So in both cases, Stacy has the doxastic states that are characteristic of social consciousness. But these doxastic states are morally virtuous expressions of social consciousness only in the first variation since they flow from moral concern. In the second variation, her doxastic states are not morally commendable because they do not flow from moral concern; even so, they are not necessarily morally condemnable either.

The upshot is that the virtue account renders more plausible verdicts about the moral and epistemic status of the doxastic states featured in cases like Informed Server and Informed Teacher that involve a robust understanding of the socio-epistemic landscape. The verdicts are that the informed believers' inferential beliefs are morally and epistemically in the clear, and their broader doxastic states may or may not be genuine expressions of social consciousness depending on how details about the believers' psychology are fleshed out.

4.3. Weathering Worry Three: Demandingness

Two features of the encroachment account give rise to the third worry: being socially conscious is, first, very difficult and, second, a moral duty. Social consciousness, the worry goes, is too difficult to be something that morality demands. This worry is brought out clearly in cases like Farmboy in which the believers' socio-epistemic environment is especially impoverished, making it exceedingly difficult to satisfy the posited moral obligation to believe in accordance with the dictates of moral encroachment.

The virtue account fares better here. Let us start with the first feature about the sheer difficulty of being socially conscious. There is no doubt that being socially conscious on the virtue account is fairly demanding—it involves certain forms of knowledge and an intellectual receptivity to available evidence that some believers may lack through no fault of their own. That said, unlike on the encroachment account, it does not involve refraining from having beliefs that are supported by the evidence. Given this, it may be plausible that the knowledge and receptivity needed for the moral virtue of social consciousness are more easily attainable for more people than the doxastic control and sophistication needed to abide by the dictates of moral encroachment.

But even if social consciousness on the virtue account is just as difficult as it is on the encroachment account, the virtue account is not troubled by the demandingness worry since it does not share the second feature. So long as having moral virtues is not morally obligatory, being socially conscious is not a moral duty on the virtue account.[36] Thus, the virtue account is not threatened by arguments against moral duties on belief that appeal to doxastic involuntarism. Also, it does not imply that those who are not socially conscious are thereby automatically blameworthy for their doxastic state.

This allows us to say that social consciousness is morally good and desirable: it is something that should be pursued, promoted, and praised. But because it is not a moral requirement, the virtue account does not automatically condemn believers who lack the relevant doxastic states through no fault of their own. In other words, on the virtue account, people who are innocently ignorant—either because of an impoverished socio-epistemic environment or because of a lack of intellectual sophistication or educational resources—are not morally

36 Some may contend that having moral virtues is morally obligatory, and so if social consciousness is a moral virtue, being socially conscious is a moral obligation. In this case, it is less clear that the virtue account fares better than the encroachment account when it comes to the demandingness worry. The stance that it does fare better would depend on the claim that social consciousness on the virtue account is less difficult than on the encroachment account. I think this claim is plausible, but I am hesitant to rest my case on it.

blameworthy for their beliefs.[37] So on the virtue account, there is room to count innocently ignorant believers like Solomon the farm boy as being hindered by the poverty of their socio-epistemic environments. That said, the virtue account also implies that these believers have plenty of room for moral growth, since, at the very least, they lack an important moral virtue.

5. CONCLUDING THOUGHTS

We have been considering two competing accounts of social consciousness: the encroachment account and the virtue account. I have been arguing that the virtue account weathers the worries that trouble the encroachment account. To conclude, I want to shore up further support for the virtue account by considering a few things about social consciousness that it is best positioned to capture.

5.1. Social Consciousness and Social Reform

Social consciousness is morally important not only because it compels believers to recognize surrounding social injustices but also because it enables believers to organize and implement social reforms that are needed to remedy those social injustices. A reform initiative is unlikely to be effective if it is not clear who is harmed by the relevant social injustice. Because of this, the doxastic states characteristic of social consciousness must include beliefs about who is harmed by the relevant social injustices—both general beliefs about the affected social groups and, importantly, inferential beliefs about affected individuals *qua* members of affected social groups.

For example, it is important for believers to have not only group-level beliefs about incarceration rates among Black men but also inferential beliefs about particular individuals *qua* Black men. It is important for a socially conscious person to recognize that just by virtue of his race, John, a Black man, is more likely than Jake, a white man, to be incarcerated, and that the race of Jadyn, an incarcerated Black man, may help explain why his incarceration was more likely. As Gardiner explains, "Central to [social] injustice is the effect on individuals' life chances. . . . When a particular person is incarcerated, underemployed, participating in a crime, and so on, one potential source of injustice is their race, gender, or other social category means the outcome was more likely. And these

37 These believers may be morally blameworthy by virtue of something else—for example, they may be blameworthy if they do not have a sufficient amount of moral concern. The point here is that they are not automatically blameworthy for their doxastic states. In other words, it is possible for believers like Solomon the farm boy to lack the beliefs characteristic of social consciousness and not be blameworthy for this in part because—despite their doxastic state—they do have sufficient amounts of moral concern.

are social facts we ought to acknowledge."[38] The point here is that it is only by having the relevant inferential beliefs that believers can grasp the full extent of the consequences of these social injustices and the tangible impact they have on real-life people, not just on abstract demographic groups. So inferential beliefs are essential to helping believers grasp the full extent of what needs to be done to address social injustices, and thus, inferential beliefs are essential to effective social reform.

In light of the role that social consciousness plays in ushering in social reform, the relevant inferential beliefs are an important part of the doxastic states that are characteristic of social consciousness. It is important, then, that an account of social consciousness is able to accommodate the potential moral value of these inferential beliefs. The virtue account can; the encroachment account cannot.

On the virtue account, inferential beliefs can be morally *commendable* for two reasons: they are morally commendable to the extent that they are an integral part of a set of doxastic states that enables believers to enact morally important social reform and that constitutes the moral virtue of social consciousness. In contrast, the encroachment account entails that these inferential beliefs are morally *condemnable* since they are proscribed by moral encroachment and thus constitute a violation of the moral duty to be socially conscious.

5.2. Accounting for Social Insensitivity

Let us call the opposite of social consciousness—whatever it is—social insensitivity.[39] I think that the picture of social insensitivity suggested by the virtue account is more robust and plausible than the picture suggested by the encroachment account.

Consider first what the encroachment account suggests about social insensitivity. Remember that on this account, being socially conscious is a moral obligation that requires us to believe in accordance with the dictates of moral encroachment. This suggests that social insensitivity centrally involves violating this obligation. Those who do not believe in accordance with the dictates of moral encroachment are socially insensitive and blameworthy for being such. In other words, on the encroachment account, believers who are not socially conscious are thereby socially insensitive.

38 Gardiner, "Evidentialism and Moral Encroachment," 182.

39 The term 'social insensitivity' closely resembles 'racial insensitivity', a concept that has been theorized about at length by José Medina. On Medina's view, racial insensitivity can be a form of "active ignorance," which has both cognitive and affective dimensions. See Medina, "Ignorance and Racial Sensitivity" for further discussion.

Now let us consider what the virtue account suggests about social insensitivity. The opposite of moral virtues are moral vices, and so on the virtue account, social insensitivity is a moral vice. Being socially insensitive, then, amounts to more than just lacking the doxastic states characteristic of social consciousness. Like its morally virtuous counterpart, social insensitivity involves both cognitive and affective components: both components must be present in order for the particular vice to obtain. Social insensitivity is a morally vicious cognitive resistance to social injustices: it has its own distinct set of characteristic doxastic states, and in order to be expressions of the vice of social insensitivity, these doxastic states must flow from morally pernicious affective states or from a lack of good ones.[40]

The doxastic states characteristic of social insensitivity are counterparts to those characteristic of social consciousness. Socially insensitive people either lack beliefs or have misguided ones about social injustices, their history, and their legacy. Their beliefs tend to stereotype marginalized social groups in negative ways, and these beliefs prop up the relevant social injustices. Socially insensitive people also tend to default to readily available but mistaken explanations of group-level information—for example, that girls score lower than boys on math tests because they are inherently less intelligent.

But having the doxastic states characteristic of social insensitivity is not sufficient for social insensitivity. In order to be a moral vice, these doxastic states must flow from morally pernicious affective states or from a lack of good ones—either a desire that a certain social group and its members fare poorly or a lack of care about their well-being. Just as moral concerns can indirectly cause the beliefs characteristic of social consciousness, morally bad concerns can indirectly cause the doxastic states characteristic of social insensitivity.

For an example of a believer with morally pernicious affective states, consider an anti-Black racist who wishes the worst for Black people. The racist may seek out uncharitable interpretations of statistical information about Black people and become angry when she hears of efforts to redress past racial injustices. For an example of someone with a lack of sufficient moral concern, consider a math teacher who simply does not care much about the well-being of her students, much less the educational flourishing of her girl students. She may remain ignorant of the sexist stereotypes that pervade STEM despite plenty of accessible evidence of their existence and impact. In these examples,

40 To clarify: it may be that having morally bad affective states is sufficient for being generally vicious on some level, but it is not sufficient for having the particular vice of social insensitivity. Similarly, it may be that having morally good affective states is sufficient for being generally virtuous on some level, but it is not sufficient for having the particular virtue of social consciousness.

the doxastic states characteristic of social insensitivity ultimately flow from morally bad affective states. This means that these believers have the moral vice of social insensitivity: their ignorance is not innocent, and it renders them morally condemnable.

The picture of social insensitivity painted by the virtue account makes room for innocently ignorant believers like Solomon. More generally, it makes room for believers who are neither socially conscious nor insensitive. Such believers may (1) have a sufficient amount of moral concern but nevertheless lack the cognitive states characteristic of social consciousness (and perhaps even have some of the cognitive states characteristic of social insensitivity) or (2) have the cognitive states characteristic of social consciousness but not ones that are rooted in moral concern.[41] When it comes to social consciousness/social insensitivity, Solomon is nonvirtuous and nonvicious in way 1 so long as his ignorance stems from features of his impoverished socio-epistemic environment rather than from morally bad affective states; the sociologist from section 3.3 may be nonvirtuous and nonvicious in way 2.[42]

In contrast, the picture of social insensitivity painted by the encroachment account does not leave room for innocently ignorant believers. After all, innocently ignorant believers and believers with morally bad affective states both violate the moral obligation to believe in accordance with the dictates of moral encroachment, and thus, both are equally socially insensitive.

5.3. *Social Consciousness in Marginalized Communities*

The concept of wokeness originated in Black communities. If social consciousness amounts to something close to wokeness, then we might expect social consciousness to be especially prominent in Black (and other marginalized) communities. The virtue account clearly predicts just this, more clearly than the encroachment account.

Remember that the virtue account says that social consciousness has both an affective component and a cognitive component. Both components are likely to be especially prominent in marginalized communities in part because

41 The type 2 nonvirtuous nonvicious person may have sufficient moral concern, but it is not what grounds the relevant cognitive disposition or its expressions, much like the person who has a disposition to tell the truth because of, say, autism rather than moral concern. For a discussion of this example, see Arpaly, "Open-Mindedness as a Moral Virtue," 75–76. Alternatively, this person may have morally bad affective states, in which case they may be generally vicious on some level even though they are not socially insensitive and are not morally blameworthy *for their beliefs*.

42 See note 34 above for discussion of relevant details about the sociologist example.

members of marginalized communities are more likely to have a personal connection with social injustices.

Start with the affective component, which involves moral concerns—that is, care about the well-being of people who are suffering social injustices and the flourishing of marginalized social groups. It is a fact about humans that we tend to care about things with which we have a personal connection, and so it would not be surprising if members of marginalized communities have the sorts of moral concerns at the heart of social consciousness. For example, consider Jasmine, a Black teenager whose brother has been a victim of police brutality and whose uncle and father are in prison on nonviolent drug offenses—a "crime" she sees her white peers get away with all the time. Given her close personal connection to the social injustices within the criminal justice system, Jasmine is especially likely to have the moral concerns at the heart of social consciousness. Jasmine probably cares deeply about the well-being of the victims of these injustices and the communities they harm, and she probably strongly desires the end of racism in the criminal justice system.

Now consider the cognitive component, which involves a cognitive disposition that results in characteristic doxastic states—that is, knowledge about injustices and a corresponding lack of ignorance. Standpoint theorists have long argued that members of marginalized communities have an epistemic advantage when it comes to knowledge about the inner workings of their social marginalization.[43] Many defend the *inversion thesis*: "Socially marginalized people, by virtue of their social location, have a superior epistemic position than non-oppressed people when it comes to knowing things about the workings of social marginalization that concern them."[44] Standpoint theorists defend the inversion thesis by arguing that socially marginalized people tend to have more informative experiences as well as greater motivation to understand their marginalization.

The epistemic advantage defended by standpoint theorists clearly involves beliefs (knowledge) about social injustices that are characteristic of social consciousness on the virtue account: socially marginalized people are more likely to know about the history, legacy, current instantiations, and inner workings

43 For recent development and defense of standpoint epistemology, see Toole, "From Standpoint Epistemology to Epistemic Oppression," "Recent Work in Standpoint Epistemology," and "Demarginalizing Standpoint Epistemology." For a defense of the claim that socially marginalized people have a mostly contingent (rather than in principle) epistemic advantage when it comes to the inner workings of their social marginalization, see Dror, "Is There an Epistemic Advantage to Being Oppressed?" For more related discussions, see Mills, *Blackness Visible*.

44 Dror, "Is There an Epistemic Advantage to Being Oppressed?" 619.

of the social injustices they suffer. It is less clear that this epistemic advantage involves or could explain a tendency to believe in accordance with the dictates of moral encroachment.

What is more, the fact that socially marginalized people tend to have the beliefs characteristic of social consciousness on the virtue account can be explained in the same way that standpoint theorists explain the inversion thesis. People in marginalized communities are too often intimately familiar with social injustices—in fact, sometimes their very safety hinges on knowing about them.[45] Because of this, members of marginalized communities are exposed to lots of evidence and information about social injustices in their everyday lives and have greater motivation to understand them. This exposure, in combination with the relevant motivations and moral concerns, naturally gives rise to a cognitive disposition to recognize and remain alert to the relevant injustices as well as to the corresponding beliefs. Jasmine, for example, does not need to read *The New Jim Crow* in order to know about social injustices riddling the criminal justice system and how they harm Black people; she gains this knowledge just through living in her community.

In light of this, the virtue account seems to clearly predict that social consciousness is a moral marker of marginalized communities: members of marginalized groups are more likely to have close personal connections to social injustices, and it is plausible that these close personal connections naturally give rise to both the affective and cognitive components of social consciousness.

In sum: social consciousness is at its root a cognitive sensitivity to surrounding social injustices. We have considered two competing ways to account for this cognitive sensitivity and its relationship to morality: the encroachment account and the virtue account. I have argued that the virtue account is better. Not only does it weather the worries that trouble the encroachment account, but it can also accommodate the role that social consciousness plays in social reform, supports a more robust picture of social insensitivity, and predicts and explains the prominence of social consciousness in marginalized communities.[46]

Concordia University
annakbrinkerhoff@gmail.com

45 It is notable that one of the earliest recorded uses of 'woke' is in the 1938 song "Scottsboro Boys" by Blues musician Huddie Ledbetter (Lead Billy). He urged his fellow Black Americans to "stay woke, keep their eyes open" to race-based risks of danger, especially when passing through parts of the American South.

46 For tremendously helpful feedback on this paper, I would like to thank Arianna Falbo, Olivia Sultanescu, Maria Waggoner, and two anonymous reviewers from the *Journal of Ethics and Social Philosophy*.

REFERENCES

Alston, William. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2 (1988): 257–99.

Arpaly, Nomy. "Open-Mindedness as a Moral Virtue." *American Philosophical Quarterly* 48, no. 1 (2011): 75–85.

———. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press, 2003.

Audi, Robert. "Doxastic Voluntarism and the Ethics of Belief." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup. Oxford University Press, 2001.

Basu, Rima. "Can Beliefs Wrong?" *Philosophical Topics* 46, no. 1 (2018): 1–17.

———. "Radical Moral Encroachment: The Moral Stakes of Racist Beliefs." *Philosophical Issues* 29, no. 1 (2019): 9–23.

———. "What We Epistemically Owe Each Other." *Philosophical Studies* 19, no. 4 (2019): 915–31.

———. "The Wrongs of Racist Beliefs." *Philosophical Studies* 176 (2019): 2497–515.

Basu, Rima, and Mark Schroeder. "Doxastic Wronging." In *Pragmatic Encroachment in Epistemology*, edited by Brian Kim and Matthew McGrath. Routledge, 2019.

Begby, Endre. "Doxastic Morality: A Modestly Skeptical Perspective." *Philosophical Topics* 46, no. 1 (2018): 155–72.

Bolinger, Renee Jorgensen. "The Rational Impermissibility of Accepting (Some) Racial Generalizations." *Synthese* 197 (2018): 2415–31.

———. "Varieties of Moral Encroachment." *Philosophical Perspectives* 32, no. 1 (2020): 5–26.

Brewster, Zachary, and Christine Mallinson. "Racial Differences in Restaurant Tipping: A Labour Process Perspective." *Services Industries Journal* 29, no. 8 (2009): 1053–75.

Brinkerhoff, Anna. "Prejudiced Beliefs Based on the Evidence." *Synthese* 199 (2021): 14317–331.

Conee, Earl, and Richard Feldman. *Evidentialism: Essays in Epistemology*. Oxford University Press, 2004.

Davis, Emmalon. "On Epistemic Appropriation." *Ethics* 128, no. 4 (2018): 702–27.

Dror, Lidal. "Is There an Epistemic Advantage to Being Oppressed?" *Noûs* 57, no. 3 (2023): 618–40.

Fritz, James. "Pragmatic Encroachment and Moral Encroachment." *Pacific Philosophical Quarterly* 98, no. 1 (2017): 643–61.

Gardiner, Georgi. "Against the New Ethics of Belief." Unpublished manuscript.
———. "Evidentialism and Moral Encroachment." In *Believing in Accordance with the Evidence: New Essays on Evidentialism*, edited by Kevin McCain. Springer, 2018.
Kelley, William Melvin. "If You're Woke You Dig It." *New York Times*, May 20, 1962.
Medina, José. "Ignorance and Racial Insensitivity." In *The Epistemic Dimension of Ignorance*, edited by Rik Peels and Martijn Blaauw. Cambridge University Press, 2016.
Mills, Charles. *Blackness Visible: Essays on Philosophy and Race*. Cornell University Press, 1998.
Moss, Sarah. "Moral Encroachment." *Proceedings of the Aristotelian Society* 118, no. 2 (2018): 177–205.
Niederle, Muriel, and Lise Vesterlund. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24, no. 2 (2010): 129–44.
Pollock, Darien. "Political Action, Epistemic Detachment, and the Problem of White-Mindedness." *Philosophical Issues* 31, no. 3 (2021): 299–314.
Romano, Aja. "A History of 'Wokeness.'" *Vox*, October 9, 2020. https://www.vox.com/culture/21437879/stay-woke-wokeness-history-origin-evolution-controversy.
Schroeder, Mark. "When Beliefs Wrong." *Philosophical Topics* 46, no. 1 (2018): 115–27.
Toole, Briana. "Demarginalizing Standpoint Epistemology." *Episteme* 19, no. 1 (2020): 47–65.
———. "From Standpoint Epistemology to Epistemic Oppression." *Hypatia* 34, no. 4 (2019): 598–618.
———. "Recent Work in Standpoint Epistemology." *Analysis* 81, no. 2 (2021): 338–50.

Journal of Ethics and Social Philosophy
VOL. 29, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.1746

BEYOND OUGHT-IMPLIES-CAN

IMPERSONAL OBLIGATORINESS IMPLIES HISTORICAL CONTINGENCY

Peter B. M. Vranas

You are the principal accountant of a company, and you are responsible for the filing of the company's tax return, which is due by 5 PM today. You do not need to file the return yourself: your assistant is also authorized to file it. What is obligatory for you is that the return *be filed* by 5 PM, not that *you* file it by 5 PM. At 4:55 PM, however, as your assistant is about to file the return, the computer network of your company crashes; as a result, you can no longer make it the case that the return is filed by 5 PM. If 'ought' implies 'can', then (1) it is no longer obligatory for you that the return be filed by 5 PM. But the tax regulation which requires that the return be filed by 5 PM is still in force, so one might argue that (2) it is still obligatory that the return be filed by 5 PM. Regardless of whether 2 is true, I maintain that 1 is compatible with 2: possibly, although it is not obligatory *for you* (or for anyone else) that the return be filed by 5 PM, it is *obligatory* that the return be filed by 5 PM. I defend this kind of view in section 1: I argue that some propositions are *impersonally* obligatory—namely, obligatory but not obligatory for anyone. But if it is impersonally obligatory that the return be filed by 5 PM although neither you nor anyone else can make it the case that the return is filed by 5 PM, then—as I argue in section 2—impersonal 'ought' does *not* imply 'can'. Is there a principle that holds for impersonal obligatoriness in lieu of ought-implies-can? I defend such a principle in section 3. I conclude in section 4.

1. OBLIGATORINESS: SIMPLICITER, PERSONAL, AND IMPERSONAL

What exactly is *impersonal* obligatoriness? To explain what it is, I start with some remarks about obligatoriness *simpliciter* and about *personal* obligatoriness. Unless I specify otherwise, I use 'obligatory' as shorthand for '*pro tanto* morally obligatory at the present time'.[1] Obligatoriness (i.e., obligatoriness

1 The label '*pro tanto*' is, strictly speaking, redundant: everything that is obligatory is *pro tanto* obligatory (because everything that is obligatory is either *pro tanto* obligatory or all-things-considered obligatory or both, and everything that is all-things-considered

simpliciter) is a familiar concept: to say that something is obligatory is to say that it is morally required. For example, it is obligatory (i.e., morally required) that people keep their promises. (This is not to say that keeping promises is *all-things-considered* obligatory.) *Personal* obligatoriness is also familiar: it is obligatoriness *for someone* (i.e., for some agent or for some group or plurality of agents).[2] For example, it is obligatory *for me* (but not *for you*) that I keep my promises: it is morally required *of me* (but not *of you*) that I keep my promises. It is convenient to take obligatoriness *simpliciter* and personal obligatoriness to apply to *propositions*, and in this paper I do so: I use (for example) 'It is obligatory (for me) that I keep my promises' interchangeably with 'The proposition that I keep my promises is obligatory (for me)'. It is also convenient to talk interchangeably about personal *obligatoriness* and about personal *obligations*,

---

obligatory is also *pro tanto* obligatory). (By contrast, the label 'merely *pro tanto*'—i.e., '*pro tanto* but not all-things-considered'—is not redundant.) Nevertheless, saying that something is *pro tanto* obligatory serves the useful function of emphasizing that it *need not* be—although it *may* be—all-things-considered obligatory (see Vranas, "'Ought' Implies 'Can' but Does Not Imply 'Must'," 495n15).

Although I consider only *moral* obligatoriness in this paper, my points also apply to other kinds of obligatoriness (legal, prudential, epistemic, etc.). Also, obligatoriness is relative to times: even if it is not obligatory *in the morning* that I meet you tonight, it may be obligatory *in the afternoon* that I meet you tonight (because at noon I promise to meet you tonight). Finally, given the qualifications 'morally' and 'at the present time', and given that '*simpliciter*' means 'without qualification', 'obligatoriness *simpliciter*' is something of a misnomer. Nevertheless, saying that something is obligatory *simpliciter* serves the useful function of emphasizing that it *need not* be—although it *may* be—personally (alternatively, impersonally) obligatory.

2 Why not say that personal obligatoriness is obligatoriness for some *person* (instead of *agent*) or group or plurality of persons? Because, by an ought-implies-can principle, persons who cannot act and thus are not agents (e.g., persons who are totally paralyzed) have no obligations (i.e., nothing is obligatory for them). But then why not use 'agential obligatoriness' instead of 'personal obligatoriness'? Because obligatoriness for someone is typically referred to as "personal" obligatoriness in the literature (see Broome, *Rationality Through Reasoning*, 13; Hintikka, "Some Main Problems of Deontic Logic," 60; Krogh and Herrestad, "Getting Personal," 135; McNamara, "Agential Obligation as Non-Agential Personal Obligation Plus Agency," 121, and "Deontic Logic"; Rönnedal, *An Introduction to Deontic Logic*, 58; cf. Ross, "The Irreducibility of Personal Obligation," 307), whereas "agential" obligatoriness is sometimes understood in the literature as the obligatoriness of *actions* or of propositions related to actions (see McNamara, "Agential Obligation as Non-Agential Personal Obligation Plus Agency," 121, and "Deontic Logic"; Vranas, "I Ought, Therefore I Can Obey," 6–7n15; contrast Chrisman, "'Ought' and Control," 436; Estlund, *Utopophobia*, 171; Price, *Contextuality in Practical Reason*, 46–47, 50). (Arguably, obligatoriness does not always apply to actions: see Broome, "Williams on *Ought*," 252–54, and *Rationality through Reasoning*, 16–18; McNamara, "Agential Obligation as Non-Agential Personal Obligation Plus Agency," 121–23; Vranas, "I Ought, Therefore I Can Obey," 6; Wedgwood, "The Meaning of 'Ought'," 131–37; contrast Schroeder, "*Ought*, Agents, and Actions," 24–33.)

and in this paper I do so: I make no distinction between (for example) the claims (i.e., propositions) that (1) it is obligatory for me *that I keep*—elliptically: *to keep*—my promises and (2) I have an unconditional obligation whose *satisfaction proposition* is the proposition that I keep my promises (or, as I say for simplicity: I have an obligation *satisfied exactly if I keep*—elliptically: an obligation *to keep*—my promises). Those who claim that there are no such entities as obligations can expunge my talk of personal obligations from this paper and replace it with talk of personal obligatoriness.[3]

How are obligatoriness *simpliciter* and personal obligatoriness related? First, *whatever is personally obligatory is also obligatory simpliciter*. For example, if it is obligatory for me that I join the army, then it is obligatory that I join the army: it is morally required that my obligation (to join the army) be satisfied.[4] Second, however, I will argue that the converse fails: *it is false that whatever is obligatory*

3 See, e.g., Liberman and Schroeder, "Commitment," 107. A *personal* obligation can be defined either as an obligation whose satisfaction proposition is *personally* obligatory or, equivalently, as an *owned* obligation (cf. Broome, "Williams on *Ought*," 256–58, and *Rationality Through Reasoning*, 12–25)—i.e., an obligation that has an *owner* (defined as someone who *has* the obligation; i.e., someone for whom the satisfaction proposition of the obligation is obligatory). Similarly, an *impersonal* obligation can be defined either as an obligation whose satisfaction proposition is *impersonally* obligatory or, equivalently, as an *unowned* obligation—i.e., an obligation that has no owner. (I am talking only about *unconditional* obligations.) I will argue that (1) some propositions are impersonally obligatory, but my arguments do not establish that (2) some obligations are impersonal (i.e., unowned), so I avoid talk of impersonal obligations in this paper. Those who deny 2 (see Wringe, "Needs, Rights, and Collective Obligations," 197, and "Global Obligations and the Agency Objection," 219; cf. Wedgwood, "The Meaning of 'Ought'," 128) might also want to deny 1, but then they would need to rebut my arguments for 1.

4 See Goble, "Normative Conflicts and the Logic of 'Ought'," 457; cf. Williams, "*Ought* and Moral Obligation," 118. (See Horty, *Agency and Deontic Logic*, 57–58 for a possible objection; for replies, see Broersen and Van der Torre, review of *Agency and Deontic Logic*, 55; Danielsson, review of *Agency and Deontic Logic*, 410; McNamara, review of *Agency and Deontic Logic*, 184.) It does not follow, and in fact it is *false*, that whatever is *all-things-considered* personally obligatory is also *all-things-considered* obligatory *simpliciter*. To see that this is false, suppose that it is obligatory *for you* that you win a certain prize (because you have promised to win, you can win, and you need the prize money to feed your child), it is also obligatory *for me* that I win that prize (because I have promised to win, I can win, and I need the prize money to feed my child), and it is impossible that we both win. Suppose also that it is morally more important that you win than that I win (because, without the prize money, your child is somewhat more likely than mine to die of starvation, and other things are equal), but it is morally more important *for me* that I win than that you win (because I have a special moral responsibility to my child but not to yours—you and your child are strangers in a distant country—and this morally outweighs for me the slightly greater need of your child), and there are no further normatively relevant considerations. Then the proposition that I win the prize is all-things-considered *personally* obligatory (it is all-things-considered obligatory *for me*) but is not all-things-considered obligatory *simpliciter* (it is instead

simpliciter is also personally obligatory. In other words, some things (i.e., propositions) are *impersonally* obligatory: they are obligatory but not obligatory for anyone.[5] One might find this claim puzzling: In the example I just gave, if it is obligatory that I join the army, does it not follow that it is obligatory *for me* that I join the army? (How could it be obligatory without being obligatory for me?) I argue in the next note that, no, it does not follow.[6] But even if it does follow, and thus the proposition that I join the army is (personally, hence) *not* impersonally obligatory, *other* propositions may be impersonally obligatory. In what follows, I provide three examples of such propositions.[7]

The Dog Example

For a first example of an impersonally obligatory proposition, suppose that, because the only judge in a certain town is severely allergic to dogs, a statutory regulation is enacted that prohibits dogs in the courthouse ("There shall be no dogs in the courthouse at any time") and that instructs the mayor of the

---

all-things-considered obligatory that *you* win the prize). (For a related example, see Broome, "Williams on *Ought*," 260–63, and *Rationality Through Reasoning*, 19–20.)

5 In the literature, "impersonal" obligatoriness is sometimes understood as what I call "obligatoriness *simpliciter*" (see McNamara, "Agential Obligation as Non-Agential Personal Obligation Plus Agency," 120) and other times understood as nonagential (see note 2 above) obligatoriness *simpliciter* (see Krogh and Herrestad, "Getting Personal," 135). I think that those uses of the term 'impersonal' can lead to confusion because, on those uses, whatever is personally (and nonagentially) obligatory is also impersonally obligatory. By contrast, on my use of 'impersonal', whatever is personally obligatory is *not* impersonally obligatory.

6 Suppose that you are an army recruiter, you are so persuasive that you can make it the case that I join the army, and you have promised your boss that I will join the army. Then it is obligatory *for you* that I join the army (see note 14 below for some objections), and thus it is obligatory that I join the army, but it does not follow that it is obligatory *for me* that I join the army: the fact that *you* have promised that I will join the army need not render it morally required *of me* that I join the army. (This example is inspired by Krogh and Herrestad, "Getting Personal," 138–39; cf. Broome, *Rationality Through Reasoning*, 20–21; McNamara, "Deontic Logic.") One can similarly argue, against Chisholm's suggestion that '*S* ought to bring it about that *p*' can be defined as 'It ought to be that *S* bring it about that *p*', that the latter does not entail the former ("The Ethics of Requirement," 150). On Chisholm's suggestion and similar ones, see Almotahari and Rabern, "The Onus in 'Ought'"; Feldman, *Doing the Best We Can*, 192–96; Forrester, *Being Good and Being Logical*, 68–73; García, "The *Tunsollen*, the *Seinsollen*, and the *Soseinsollen*"; Geach, "Whatever Happened to Deontic Logic?" 3–4; Harman, *Change in View*, 131–32; Hilpinen, "On the Semantics of Personal Directives," 148–49; Horty, "Agency and Obligation," 285–90, *Agency and Deontic Logic*, 44–58, and *Reasons as Defaults*, 68–69n4; Horty and Belnap, "The Deliberative Stit," 619–28; Kordig, "Relativized Deontic Modalities," 225–27; Krogh and Herrestad, "Getting Personal," 136–45; McNamara, "Deontic Logic"; Schroeder, "*Ought*, Agents, and Actions," 8–11; cf. Anderson, "Logic, Norms, and Roles," 43; Hartmann, *Ethics*, 259–60; Meinong, *On Emotional Presentation*, 141–42.

7 See Krogh and Herrestad, "Getting Personal," 145–46, for another example.

town to appoint a person solely responsible for enforcing the prohibition. Then, assuming that the regulation is not only legally but also morally binding, (1) it is obligatory that there be no dogs in the courthouse (at any time). Suppose further that the person who was solely responsible for enforcing the prohibition has died and the mayor has not yet appointed a replacement, so no one is currently responsible for enforcing the prohibition (although the regulation, and thus the prohibition, remains in force: it has not been repealed). Then, assuming that there are no further normatively relevant considerations, it is not obligatory *for anyone*—and thus it is *impersonally* obligatory—that there be no dogs in the courthouse. One might suggest that the regulation imposes obligations on everyone: (2) it is obligatory *for everyone* not to bring or keep dogs in the courthouse. One might even suggest that 1 is equivalent to 2 and, more generally, that every claim of obligatoriness *simpliciter* is equivalent to some claim (or other) of personal obligatoriness. I reply that, regardless of whether 1 entails 2, 2 does not entail 1: 2 is compatible with the claim—which is incompatible with 1—that dogs which no one brings or keeps in the courthouse (e.g., dogs that stray into the courthouse) are allowed to be in the courthouse. One might alternatively suggest that 1 is equivalent to the claim that (3) it is obligatory for everyone not to bring or keep dogs in the courthouse *and to remove* any dogs that stray into the courthouse. I reply that since (as I explained) no one is currently responsible for enforcing the prohibition against dogs in the courthouse, it is not obligatory for anyone to remove any dogs that stray into the courthouse, so 1 does not entail 3.[8] My replies support the conclusion that some claims of obligatoriness *simpliciter* are *not* equivalent to any claims of personal obligatoriness.

Even if one is unable to find any specific fault with the dog example, one might argue that the example is somehow faulty because it is conceptually impossible (for morality, or for anything else) to require something without requiring it *of* anyone. To see that this is conceptually possible, I reply, suppose that a fire code contains a provision formulated as follows: "Every building shall have an emergency exit." Then the fire code requires that every building have an emergency exit. But the fire code need not require *of* any particular agent (or group of agents) that *every* building have an emergency exit: maybe, through some other provision, the fire code requires of each agent only that any building owned by that agent have an emergency exit (and no agent owns every building). Or maybe the fire code does not require anything of anyone: maybe

8 Given that the regulation was enacted because the town judge is severely allergic to dogs, the regulation applies also to stray dogs (not just to pet dogs). One might ask: How could a regulation require that stray dogs behave in a certain way? I reply that the regulation does not require that: it requires that there *be* no dogs in the courthouse, not that dogs *refrain* from entering the courthouse.

the legislators who enacted the code could not agree on whether it should be required of the owners or of the builders of any given building that the building have an emergency exit and left the matter open for future legislators to decide, so the code is silent on the matter (it contains no relevant provision). In that case, the code sets a standard (and buildings that lack an emergency exit are in violation of the standard) but does not assign anyone responsibility for complying with the standard. I conclude that it is conceptually possible to require something without requiring it *of* anyone (and even without requiring *anything* of anyone). One might respond that, even if (1) this is conceptually possible for a fire code, it does not follow that (2) it is conceptually possible for morality. I agree, but the point of the fire code example is not to support 2 by using 1: I have already supported 2 by using the dog example. The point is instead to refute the general claim that it is conceptually impossible to require something without requiring it of anyone, and the fire code example does refute this general claim.

The Voting Example

For a second example of an impersonally obligatory proposition, suppose that you have an obligation (because you have promised) to vote, and I also have an obligation (because I have promised) to vote; it is possible that we both vote, and there are no further normatively relevant considerations. Then it is obligatory that we both vote (since it is morally required that we both keep our promises).[9] But it is not obligatory *for you* that we both vote: what is obligatory for you is instead that *you* vote. Similarly, it is not obligatory *for me* that we both vote: what is obligatory for me is instead that *I* vote. And it is not obligatory *for anyone else* either that we both vote: for whom could it be obligatory, given that there are no further normatively relevant considerations? It follows that it is not obligatory *for anyone* that we both vote. In sum, it is *impersonally* obligatory (i.e., obligatory but not obligatory for anyone) that we both vote.

Objecting to my claim that it is not obligatory for anyone that we both vote, one might claim that it is obligatory *for our group* (namely, the group that consists of you and me) that we both vote. For this objection to get off the ground, it must be assumed that *any* two agents form a group; otherwise (i.e., if there are two agents who do not form a group), I can avoid the objection by assuming that you and I do not form a group. If not every group is an agent, I can assume

9 In this example, it is both obligatory that you vote (because it is obligatory for you that you vote) and obligatory that I vote (because it is obligatory for me that I vote) and it is possible that we both vote, so it is reasonable to infer that it is obligatory that we both vote (although, for reasons I will not go into, I do not accept the *general* principle that, if it is both obligatory that p and obligatory that q and it is possible that both p and q, then it is obligatory that both p and q).

that our group is not an agent, and I can reply to the objection by appealing to the claim that *an entity has an obligation only if the entity is an agent*.[10] One might respond that even some groups that are not agents have obligations: an unstructured group of pedestrians who happen to witness a mugging has an obligation to stop the mugging. I have two replies. First, even if the group of pedestrians is not a *full-fledged* agent (due to its lack of structure), the group is still an agent in the relevant sense (namely, an entity that can act) *if* it can act to stop the mugging (and if it cannot do so, then it has no obligation to do so).[11] Second, the group of pedestrians has an obligation to stop the mugging only if (1) it is blameworthy (in the absence of any justification or excuse) if it fails to

10 For (at least tentative) endorsements of (versions of) this claim, see Aas, "Distributing Collective Obligation," 14; Björnsson, "Essentially Shared Obligations," 111, 117; Collins, "Collectives' Duties and Collectivization Duties," 231, 239–40, and *Group Duties*, 35, 60–95; Isaacs, *Moral Responsibility in Collective Contexts*, 148–49 (cf. "Collective Responsibility and Collective Obligation," 44–45); Lawford-Smith, "The Feasibility of Collectives' Actions," 458; Pinkert, "What We Together Can (Be Required to) Do," 188–89; Schwenkenbecher, "Joint Duties and Global Moral Obligations," 315, 317–18, and "Joint Moral Duties," 61. For (at least implicit) rejections (based on an objection that I go on to examine in the text), see Cripps, "Climate Change, Collective Harm and Legitimate Coercion"; May, "Collective Inaction and Shared Responsibility," *Sharing Responsibility*, and "Collective Inaction and Responsibility," 218; Wringe, "Global Obligations and the Agency Objection," 220–24, "From Global Collective Obligations to Institutional Obligations," 174–77, "Collective Obligations," 484–85, and "Global Obligations, Collective Capacities, and 'Ought Implies Can'," 1530–32. For discussions, see Björnsson, "Collective Responsibility and Collective Obligations Without Collective Moral Agents," 130–34, and Schwenkenbecher, *Getting Our Act Together*, 31–36.

11 One might object by contesting my understanding of an agent as an entity that can act (cf. Helm, "Plural Agents," 19; List and Pettit, "Group Agency and Supervenience," 87–88; Pettit, "Responsibility Incorporated," 178; Wringe, "Global Obligations, Collective Capacities, and 'Ought Implies Can'," 1529; contrast Aas, "Distributing Collective Obligation," 14; Bratman, *Shared Agency*, 125–26; Estlund, *Utopophobia*, 218): one might claim that, although the group of pedestrians can act, it is not an agent (but is instead a *potential* or *putative* agent: see Isaacs, *Moral Responsibility in Collective Contexts*, 144–45, and "Collective Responsibility and Collective Obligation," 45; Wringe, "Global Obligations and the Agency Objection," 221–24, "From Global Collective Obligations to Institutional Obligations," 176–77, "Collective Obligations," 484–85, and "Global Obligations, Collective Capacities, and 'Ought Implies Can'," 1531n28; cf. Björnsson, "Essentially Shared Obligations," 109; Cripps, "Climate Change, Collective Harm and Legitimate Coercion," 176–78; May, "Collective Inaction and Shared Responsibility," *Sharing Responsibility*, 109, 122, and "Collective Inaction and Responsibility," 216–18). I reply that if it is granted that *an entity has an obligation only if the entity is either an agent or a potential agent*, then I can assume that our group is not even a potential agent: we are unrelated (we are supposed to vote at different elections in different countries), and we have no way to communicate or even to become aware of each other's existence.

stop the mugging.[12] Similarly, our group has an obligation satisfied exactly if we both vote only if (2) it is blameworthy (in the absence of any justification or excuse) if we fail to both vote. But although I can grant that 1 is plausible, 2 is implausible: if we fail to both vote (i.e., you fail to vote or I fail to vote), our *group* is not blameworthy (for that failure), since our group has not promised that we will both vote—instead, *you* have promised that *you* will vote, and *I* have promised that *I* will vote. (Of course, blameworthiness can also arise from factors other than breaking promises, but I supposed that there are no further normatively relevant considerations.)[13]

In the voting example, no claim of personal obligatoriness is equivalent to the claim that (1) it is obligatory that we both vote. One might object that 1 is equivalent to the claim that (2) it is obligatory for you to vote *and* it is obligatory for me to vote. I reply that 1 does not entail 2: possibly (though not actually), 2 is false, but 1 is true because (3) it is obligatory for *you* that *I* vote and it is obligatory for *me* that *you* vote. (To see how 3 can be true, suppose that you have promised that I will vote and you can make it the case that I vote, and I have promised that you will vote and I can make it the case that you vote.)[14]

12 Cf. Blomberg and Petersson, "Team Reasoning and Collective Moral Obligation," 491n14, 505–6; Darwall, "Why Obligations Can't Be Bipolar (Directed) All the Way Down."

13 A third possible reply is to deny that the group of pedestrians has an obligation to stop the mugging and claim instead that each pedestrian has a *collectivization* obligation: an obligation to take steps towards forming a collective agent that can stop the mugging (Collins, "Collectives' Duties and Collectivization Duties" and *Group Duties*; cf. Held, "Can a Random Collection of Individuals Be Morally Responsibile?" 480; Jansen, "A Plural Subject Approach to the Responsibilities of Groups and Institutions," 98; Lawford-Smith, "The Feasibility of Collectives' Actions," 458; Schwenkenbecher, "Joint Duties and Global Moral Obligations," 317, 321–22, "Joint Moral Duties," 62n3, and *Getting Our Act Together*, 117–18; contrast Estlund, *Utopophobia*, 356–57n30). Note that the objection I examined in the text relies on the claim that (1) it is obligatory for *our group* that we both vote, but one might alternatively propose an objection based on the claim that (2) it is *jointly* obligatory for *you and me* that we both vote (in other words, you and I *jointly have* an obligation satisfied exactly if we both vote). The contrast between 1 and 2 relies on the distinction between *individual obligatoriness* (which relates a *single* entity—in the case of 1, a *group*—to a proposition) and *joint obligatoriness* (which relates *multiple* entities—in the case of 2, the *members* of a group—to a proposition); on this distinction, see Pinkert, "What We Together Can (Be Required to) Do," 187–90 (see also Björnsson, "Essentially Shared Obligations"; Schwenkenbecher, "Joint Duties and Global Moral Obligations," "Joint Moral Duties," and *Getting Our Act Together*). I reply that 2 is false because, if we fail to both vote, we are not jointly blameworthy (for that failure), since we have not jointly promised that we will both vote.

14 I realize that promises result in obligations only under certain conditions (e.g., when the promises are not obtained by coercion or deception), but I assume throughout this paper that those conditions are met. One might object that when I promise that you will vote, I typically do not acquire an obligation satisfied exactly if you vote: I acquire instead an obligation satisfied exactly if *I make it the case* that you vote (see Broome, *Rationality*

One might alternatively object that 1 is equivalent to the claim that (4) it is obligatory *for someone* that you vote and it is obligatory *for someone* that I vote. I reply that 1 does not entail 4: possibly (though not actually), 4 is false, but 1 is true because (5) it is obligatory for you that (a) *you vote exactly if I vote* and it is obligatory for me that (b) *either I vote or you vote (or both)*. (The point is that propositions a and b jointly entail that we both vote.)[15] Prompted by 5, one might suggest that a proposition *P* is impersonally obligatory only if some personally obligatory propositions jointly entail *P*. I reply that this suggestion is falsified by the dog example (in which the proposition that there are no dogs in the courthouse is impersonally obligatory but is not entailed by any personally obligatory propositions); see also the example that follows.

The Poisoning Example

For a third example of an impersonally obligatory proposition, suppose that your daughter has been given a deadly poison. There is only one antidote, available only at the National Antidote Center. You email the director of the center, and you receive in reply the following email, which contains only true claims:

> I am sorry to hear that your daughter has been poisoned. There is another person (to whom I am separately sending an identical email) whose daughter has been given the same poison, but there is only one dose of the antidote. I am asking you, and I am also asking that other person, to pay me a bribe by sending in the next hour ten thousand dollars to my bank account; my account details are attached. If only one of you pays, then I will give the antidote to the daughter of whoever pays;

Through Reasoning, 17). In reply, I can grant that this is typically so, but I assume that the specific wording of my promise makes it clear that my promise counts as kept exactly if you vote, even I do not make it the case that you vote. One might also object that I have no obligation satisfied exactly if *you* do something (e.g., you vote) because (1) my obligations are obligations for *me* to do (or to refrain from doing) something: they are satisfied exactly if *I* do (or I refrain from doing) something (see Schwenkenbecher, "Joint Duties and Global Moral Obligations," 320). I reply that 1 is false: if I promise my mother that my son will call her today (not that I will *make* him call her, although I *can* make him call her) and, a couple of seconds after I promise, my son calls my mother on his own (without any prompting from me, and being unaware of my promise), then the obligation that I acquire when I promise is satisfied although I do not do (and I do not refrain from doing) anything (see McNamara, "Agential Obligation as Non-Agential Personal Obligation Plus Agency," 121; cf. Broome, "Williams on *Ought*," 254, and *Rationality Through Reasoning*, 16–18; Krogh and Herrestad, "Getting Personal," 151; Vranas, "I Ought, Therefore I Can Obey," 6). One might object that there are things I *fail* to do, but I reply that it does not follow that I *refrain* from doing them: to refrain from doing something is to *make it the case* that one fails to do it (see Belnap et al., *Facing the Future*, 40–45).

15 See Goble, "Normative Conflicts and the Logic of 'Ought'," 481n13, for a similar example.

> but if both of you pay or neither of you pays, then I will randomly choose one of the two girls and give her the antidote. The antidote is perfectly safe and effective, but the girl who does not get it will be dead tomorrow. Don't try to change my mind: you will be unable to communicate with me in the next hour because I have taken a drug that in a few seconds will render me unconscious for a bit more than an hour.

Suppose that (unbeknownst to you) I am the other person to whom this email refers, but there cannot be any communication between you and me in the next hour. Suppose also that each of us can easily afford to pay ten thousand dollars in the next hour, and there are no further normatively relevant considerations. In this example, it is (*pro tanto*) obligatory that *we both fail to pay* (since it is morally required that people fail to bribe public officials). Moreover, the case in which we both fail to pay (and in which the girl who gets the antidote is chosen randomly) is overall morally better than the alternative cases: (1) it is better than the case in which we both pay because in that case two bribes are paid (and the girl who gets the antidote is again chosen randomly), and (2) it is better than the case in which only one of us pays because in that case a bribe is paid and (unfairly) determines which girl gets the antidote.[16] Since it is both *pro tanto* obligatory and overall morally best that we both fail to pay,

16 In all cases, exactly one girl gets the antidote, and (in the absence of further normatively relevant considerations) it does not matter morally *which* girl gets it; but it does matter morally *how* the girl who gets it is chosen. One might argue that the case in which we both pay is overall morally better than the case in which we both fail to pay: in both cases, the girl who gets the antidote is chosen randomly, but in the case in which we both pay, you satisfy your stronger obligation to promote your daughter's survival (because, as I explain shortly in the text, if you pay, then your daughter has a significantly higher chance of getting the antidote than if you do not pay) and you violate your weaker obligation not to bribe a public official (and I also do so), whereas in the case in which we both fail to pay, you satisfy your weaker obligation not to bribe a public official and you violate your stronger obligation to promote your daughter's survival (and I also do so). In reply, I submit that the poisoning example shows that *a case in which people satisfy their weaker obligations can be overall morally better than a case in which people satisfy their stronger obligations*; but if one disagrees, I can show this by modifying the example as follows. Suppose that there are exactly two doses of the antidote, and the director writes: "If only one of you pays, then I will keep one dose, and I will give the other dose to the daughter of whoever pays; if both of you pay, then I will keep one dose, and I will randomly choose one of the two girls and give her the other dose; and if neither of you pays, then I will randomly choose one of the two girls and give her one dose, and depending on the outcome of a coin toss I will either keep the other dose or give it to the other girl." In this modified example, the fact that if we both fail to pay there is a significant chance that an extra girl gets the antidote outweighs the fact that if we both pay we satisfy our stronger obligations, so the case in which we both fail to pay is overall morally better than the case in which we both pay. For simplicity, I stick to the unmodified poisoning example in the text.

it is *all-things-considered* obligatory that we both fail to pay. Nevertheless, it is not obligatory *for anyone* that we both fail to pay. To start with, it is not obligatory *for you* that we both fail to pay: it is instead (*pro tanto*) obligatory for you that *you* fail to pay. And it is also (all-things-considered) obligatory for you that you *pay*: if I pay, then your daughter has a 50 percent chance of getting the antidote if you pay but has no chance if you do not pay, and if I do not pay, then your daughter has a 100 percent chance of getting the antidote if you pay but has only a 50 percent chance if you do not pay. (If you pay, you violate your obligation not to bribe a public official, but this is outweighed by the fact that you increase your daughter's chance of surviving. Admittedly, you reduce *my* daughter's chance of surviving, but this is outweighed by the fact that you have a special moral responsibility to *your* daughter.) Similarly, it is not obligatory *for me* that we both fail to pay: it is instead (*pro tanto*) obligatory for me that *I* fail to pay, and it is also (all-things-considered) obligatory for me that I *pay*. Finally, it is not obligatory *for anyone else* that we both fail to pay: it is not obligatory *for our group* (see my discussion of the voting example), and—by an ought-implies-can principle—it is not obligatory *for the director*, since the director is unconscious and thus cannot make it the case that we both fail to pay.[17] In sum, it is impersonally obligatory that we both fail to pay. Moreover, the proposition *P* that we both fail to pay is *all-things-considered impersonally obligatory* (i.e., it is both all-things-considered obligatory and impersonally obligatory), but it is not the case that some all-things-considered personally obligatory propositions jointly entail *P*.[18]

17 My claim that *now* (shortly after the director became unconscious) it is not obligatory for the director that we both fail to pay is compatible with the claims that (1) *before* the director became unconscious, it *was* obligatory for the director that we both fail to pay, and that (2) *after* the director becomes conscious again, it *will* be obligatory for the director to return any bribes paid by you or me. (By assumption, which girl gets the antidote depends on who pays, regardless of whether any paid bribes are returned.)

18 I define an *all-things-considered personally obligatory* proposition as a proposition that is all-things-considered obligatory for someone (*Definition 1*). However, by analogy with my definition of an all-things-considered impersonally obligatory proposition as a proposition that is both all-things-considered obligatory and impersonally obligatory (*Definition 2*), one might propose defining an all-things-considered personally obligatory proposition as a proposition that is both all-things-considered obligatory and personally obligatory (*Definition 1**). Also, by analogy with my definition of an impersonally obligatory proposition as a proposition that is obligatory but not obligatory for anyone, one might propose defining an all-things-considered impersonally obligatory proposition as a proposition that is all-things-considered obligatory but not all-things-considered obligatory for anyone (*Definition 2**). To reject both Definition 1* and Definition 2*, I argue below that some proposition *P* is (1) all-things-considered obligatory, (2) personally obligatory, and (3) not all-things-considered obligatory for anyone. Then *P* is (by 1 and

The poisoning example is a moral analog of the Prisoner's Dilemma. A common view is that the Prisoner's Dilemma "illustrates a conflict between individual and group rationality."[19] I suggest instead that the Prisoner's Dilemma illustrates a conflict between individual (or *personal*) and *impersonal* rationality: it is sometimes (impersonally) rationally required that people fail to do what is rationally required of them. Similarly, the poisoning example illustrates a conflict between personal and impersonal obligatoriness: it is sometimes all-things-considered impersonally obligatory (and thus also all-things-considered obligatory *simpliciter*) that people fail to do what is all-things-considered obligatory for them. If so, then impersonal obligatoriness is irreducible to personal obligatoriness (and so is also obligatoriness *simpliciter*). One might object that in the poisoning example it is overall morally best but it is not obligatory (and thus it is not impersonally obligatory) that we both fail to pay, so the example illustrates only a conflict between what is all-things-considered personally obligatory and what is overall morally best.[20] In reply, compare the poisoning example with

---

2) all-things-considered *personally* obligatory according to Definition 1* but is (3) *not* all-things-considered obligatory *for anyone* (and this is undesirable—and precluded by Definition 1). Moreover, *P* is (by 1 and 3) all-things-considered *impersonally* obligatory according to Definition 2* but is (2) *personally* obligatory (and this is undesirable). To argue that there is such a proposition *P*, modify the poisoning example by supposing that a teenage hacker who reads the director's emails can make it the case that we both fail to pay (by remotely shutting down our internet-connected devices) and promises a bystander that we will both fail to pay, but then the hacker's father orders the hacker to disregard that promise. Then the proposition *P* that we both fail to pay is (1) all-things-considered obligatory and (2) personally obligatory (it is obligatory for the hacker, given the hacker's promise to the bystander), but is (3) not all-things-considered obligatory for anyone (it is not all-things-considered obligatory for the hacker, assuming that the order given by the hacker's father outweighs the hacker's promise to the bystander).

19 Kuhn, "Prisoner's Dilemma"; cf. Blomberg and Petersson, "Team Reasoning and Collective Moral Obligation." See also Campbell, "Background for the Uninitiated."

20 One might argue that, although it is natural to say "It ought to be the case that we both fail to pay," this sentence does not express a deontic claim (of obligatoriness): it expresses instead the evaluative claim that it is overall morally best that we both fail to pay. In a similar vein, James Forrester argues that "'There should be no more war' places no obvious obligations on anyone to act in any way; it says little more than that a world without war would be a better world than a world with war" (*Being Good and Being Logical*, 56–57; cf. Smith, "Moral Realism, Moral Conflict, and Compound Acts," 342; Tomalty, "The Force of the Claimability Objection to the Human Right to Subsistence," 5), and many other authors make similar points about "ought to be" sentences (see Castañeda, "On the Semantics of the Ought-to-Do," 450; Finlay and Snedegar, "One Ought Too Many," 104; Guendling, "Modal Verbs and the Grading of Obligations," 122–23; Haji, *Deontic Morality and Control*, 15; Hansson, "The Varieties of Permission," 197; Harman, "Relativistic Ethics," 113, 118; Humberstone, "Two Sorts of 'Ought's'," 10; Mason, "Consequentialism and the 'Ought Implies Can' Principle," 319; McConnell, "'"Ought" Implies "Can"' and the Scope

the following modification of it: instead of asking each of us to pay a bribe, the director asks each of us to donate in the next hour ten thousand dollars to what we both know (but the director does not know) to be a wasteful charity that squanders most donations. There is a significant difference between the unmodified poisoning example and the modified one: bribing a public official violates an obligation, but donating to (what one knows to be) a wasteful charity violates no obligation (although it is not morally best). As a result, satisfying our obligations not to bribe a public official requires that we both fail to pay (i.e., fail to bribe) in the unmodified example, but nothing similarly requires that we both fail to pay (i.e., fail to donate) in the modified example. I capture this difference by saying that in the unmodified example it is obligatory (i.e., morally required) that we both fail to pay whereas in the modified example it is not; but the objection fails to capture the difference because it leads to saying instead that in both examples it is overall morally best but not obligatory that we both fail to pay.[21]

Does the concept of impersonal obligatoriness play any significant roles in moral reasoning and in moral theorizing? To see that it does, consider again the dog, voting, and poisoning examples. In the dog example, moral reasoners who know that it is impersonally obligatory that there be no dogs in the courthouse

of Moral Requirements," 438; Robinson, "Ought and Ought Not," 195; Sidgwick, *The Methods of Ethics*, 33; van Fraassen, "Values and the Heart's Command," 6). I reply that I am not claiming that every "ought to be" sentence expresses a deontic claim (of obligatorinesss) rather than an evaluative claim. But some "ought to be" sentences do so: the sentence "It ought to be the case that we both vote" can express the proposition that (1) it is obligatory that we both vote. I agree with Forrester that 1 "places no obvious obligations on anyone to act in any way": as I argued, 1 does not entail that it is obligatory for me to vote or that it is obligatory for you to vote. But it does not follow that 1 is not a deontic claim: the reason why it is obligatory that we both vote is not that this would make the world a better place (in fact, the opposite may be the case) but is instead that each of us has promised (and for this reason has an obligation) to vote, and this suggests that 1 is a deontic claim.

21 If one accepts the consequentialist view that something is all-things-considered obligatory exactly if it is overall morally best, then one should say (contrary to what I said) that, even in the modified example, it is all-things-considered obligatory (since it is overall morally best) that we both fail to pay. I reply first that the objection I addressed in the text does not even get off the ground if one accepts the consequentialist view because then one may not say that in the unmodified example it is overall morally best *but not obligatory* that we both fail to pay. Moreover, the fact that the consequentialist view fails to capture the difference I noted in the text is a reason to reject the consequentialist view. I propose instead that, in the modified example, it is all-things-considered obligatory that either we both fail to pay or we both pay because these are the only two cases in which the girl who gets the antidote is randomly and thus fairly chosen (although the case in which we both fail to pay is morally better than the case in which we both pay). (In my discussion of the unmodified example, I implicitly appealed to the claim that if something is *both pro tanto* obligatory and overall morally best, then it is all-things-considered obligatory; but this claim does not entail the consequentialist view.)

may infer that they have a *reason* (though not an obligation) to remove any dogs that stray into the courthouse (assuming that they can do so) and may also infer that they have a reason (and arguably even an *obligation*) not to bring or keep dogs in the courthouse. More generally, moral reasoners who know that it is impersonally obligatory that *p* may infer that they have a reason (and in some cases even an obligation) to contribute to its becoming the case that *p* (assuming that they can do so)—but they may infer this only under certain conditions, as I argue next, and one task for moral theorizers is to identify those conditions. To see that some conditions are needed, suppose that in the voting example your father, who knows that it is impersonally obligatory that we both vote, can contribute to its becoming the case that we both vote by convincing you to vote, but also knows that, if he does so, then you will vote for a racist candidate that he opposes. Then your father need not have any reason (and may not infer that he has a reason) to convince you to vote. Finally, in the poisoning example, moral reasoners who realize that there is a conflict between personal and impersonal obligatoriness may infer that they have a reason to avoid (to the extent that they can) situations that lead to such conflicts. Moral theorizers, on the other hand, have the task of figuring out whether such conflicts are problematic for morality. These issues deserve further investigation, but it is not a goal of this paper to provide a complete theory of impersonal obligatoriness.[22]

22 One might think that the distinction between personal obligatoriness and obligatoriness *simpliciter* amounts to a *de re*/*de dicto* distinction: according to Forrester (*Being Good and Being Logical*, 65–66), "the 'ought to be' is a *de dicto* operator, while the 'ought to do' is *de re*," because (1) "the 'ought to be' operator … operates on entire propositions" but "the 'ought to do' operator … operates on predicates only," and (2) "It ought to be that George takes out the garbage" might be true even if there is no such person as George, but "George ought to take out the garbage" "cannot possibly be true unless there is such a person as George." I reply first that the distinction between personal obligatoriness and obligatoriness *simpliciter* does not correspond exactly to the distinction between 'ought to do' and 'ought to be' (cf. Humberstone, "Two Kinds of Agent-Relativity," 146): only some claims of personal obligatoriness (namely, those that are also claims of agential obligatoriness: see note 2 above) are "ought to do" claims, and only some "ought to be" claims (namely, those that are deontic rather than evaluative: see note 20 above) are claims of obligatoriness *simpliciter*. In what follows, I address analogs of 1 and 2 that are about obligatoriness *simpliciter* instead of 'ought to be' and about personal obligatoriness instead of 'ought to do'. Let '*Tg*' stand for "George takes out the garbage," and introduce the operators '*O*' ("it is obligatory that") and 'O_g' ("it is obligatory for George that"). (1′) Both operators can prefix either closed formulas (*OTg*: It is obligatory that George take out the garbage (*de dicto simpliciter*); O_gTg: It is obligatory for George that he take out the garbage (personal *de dicto*)) or open formulas ($\lambda x(OTx)g$: George is such that it is obligatory that he take out the garbage (*de re simpliciter*); $\lambda x(O_gTx)g$: George is such that it is obligatory for him that he take out the garbage (personal *de re*)). ('λ' is the predicate abstraction quantifier.) (2′) Even if the *de dicto simpliciter* claim above does not entail that George exists but the

2. IMPERSONAL 'OUGHT' DOES NOT IMPLY 'CAN'

Consider the following ought-implies-can principle:

> OIC: If at a given time it is obligatory for an agent that *p*, then at that time the agent can (i.e., has both the ability and the opportunity to) make it the case that *p*.

This principle is formulated in terms of obligatoriness *for an agent*.[23] In this section, I argue that no version of this principle holds for *impersonal* obligatoriness. Note first that the following sentence does *not* express a version of the above principle: "If at a given time it is impersonally obligatory that *p*, then at that time the agent can make it the case that *p*." This sentence expresses no principle at all: Who is "the agent"? To avoid this problem, one might propose replacing 'the agent' with 'some agents' (understood as referring to a single agent, a group of agents, or a plurality of agents). This proposal yields the following principle:

> IOIC_1: If at a given time it is impersonally obligatory that *p*, then at that time some agents can (i.e., have both the ability and the opportunity to) make it the case that *p*.

---

personal *de re* claim does entail that George exists, the fact that there are also personal *de dicto* and *de re simpliciter* claims shows that the distinction between personal obligatoriness and obligatoriness *simpliciter* cuts across the *de re*/*de dicto* distinction. (Strictly speaking, if *g* is a constant that denotes George at every world, then the personal *de re* and *de re simpliciter* claims are logically equivalent to the corresponding *de dicto* claims; to avoid this, I could use a *descriptor* instead of *g* (Priest, *An Introduction to Non-Classical Logic*, 355): a descriptor need not denote the same object at every world.)

23 Several remarks are in order. First, strictly speaking, my formulation of OIC should be prefixed with 'By virtue of conceptual necessity' (and similarly for the other principles that I consider later). Second, like (personal) obligatoriness (cf. note 1 above), ability (plus opportunity) is relative to times: even if *in the morning* you can run in tomorrow's marathon, maybe *in the afternoon* you cannot (because at noon you break your leg). Third, many ought-implies-can principles have been formulated in the literature (see Vranas, "I Ought, Therefore I Can Obey," 3n3, for references), but here I consider OIC because I take something like OIC to be the most plausible ought-implies-can principle for (unconditional) personal obligatoriness and thus the best starting point in the quest for an ought-implies-can principle for (unconditional) impersonal obligatoriness. Fourth, in previous work (see Vranas, "I Ought, Therefore I Can Obey," 30), I formulated (and I argued that it is better to formulate) ought-implies-can principles in terms of personal *obligations* instead of personal *obligatoriness*. Nevertheless, here I formulate OIC in terms of personal *obligatoriness* because I plan to distinguish OIC from *impersonal* versions of it: I formulate those versions in terms of impersonal *obligatoriness* because (as I said in note 3 above) I avoid talk of impersonal *obligations*.

This principle might seem plausible: if it is impersonally obligatory that there be no dogs in the courthouse, then arguably some agents can make it the case that there are no dogs in the courthouse. I argue next, however, that IOIC$_1$ is false. Suppose that (1) it is obligatory *for you* that you win a gold medal in a given race (because you have promised to win and you can win), (2) it is also obligatory *for me* that I win a gold medal in that race (because I have promised to win and I can win), (3) it is possible that we *both* win a gold medal (because it is possible that we tie for first place), (4) no one can make it the case that we tie for first place (in particular, we cannot coordinate our actions before or during the race so as to finish at the same time), and (5) there are no further normatively relevant considerations. Then (by 1, 2, 3, and 5) it is impersonally obligatory that we both win a gold medal (as one can see by reasoning as in the voting example of section 1), but (by 4) no agents can make it the case that we both win a gold medal; so IOIC$_1$ is false. One might respond that, although we cannot *make it the case* that we both win a gold medal, in a sense we *can* both win—or it is *feasible* for us that we both win—a gold medal: we can make it the case that we both try to win, and if we both tried to win it *might* be the case that we tie for first place. More generally, say that at a given time it is *feasible* for some given agents that *p* exactly if there is something that at that time those agents can make the case such that, if they were to make it the case, then it *might* be the case that *p*. (It follows that, if at a given time some given agents can make it the case that *p*, then at that time it is feasible for those agents that *p*.) One might then propose the following principle, which is *not* refuted by the race example:

> IOIC$_2$: If at a given time it is impersonally obligatory that *p*, then at that time it is feasible for some agents that *p*.

This is a very weak principle because the above concept of feasibility is very weak (and is weaker than most feasibility concepts in the literature).[24] For example, it is feasible for you that you win ten lotteries because, if you bought tickets for ten lotteries (which I assume you can do), it *might* be the case that you win all ten lotteries. Nevertheless, I argue next that even this very weak principle is false. Modify the race example by supposing that if we both entered the race, then either I would kill you or you would kill me (we would fight a duel to the death, and each of us can win such a duel). In this modified example, it is again impersonally obligatory that we both win a gold medal. But it is not feasible for any agents that we both win a gold medal because, no matter

24 Cf. Estlund, *Utopophobia*, 243–48; Southwood, "Does 'Ought' Imply 'Feasible'?" 11–17, and "The Feasibility Issue"; Wiens, "Political Ideals and the Feasibility Frontier."

what any agents were to make the case (among the things that they can make the case), it *would not* be the case (and thus it is false that it *might* be the case) that we both win a gold medal: either we would not both enter the race and then we would not both win (assuming that entering the race is necessary for winning), or we would both enter the race and then again we would not both win (since either I would kill you or you would kill me). One might object that if we both entered the race but neither of us killed the other, then it might be the case that we both win. I agree, but I reply that, given that if we both entered the race either I would kill you or you would kill me, I assume that no agents can make it the case that we both enter the race but neither of us kills the other. I conclude that IOIC_2 is false.[25]

The above counterexample to IOIC_2 relies on (the impersonal obligatoriness of) the proposition that we both win a gold medal. This proposition does not entail that we *make it the case* that we both win a gold medal (since this proposition does not preclude that we both win by coincidence, that we just happen to tie for first place), and thus is not an *agential* proposition, defined as a proposition to the effect that some agents make something the case.[26] To avoid the counterexample, one might propose restricting IOIC_2 to agential propositions. This proposal yields the following principle:

> IOIC_3: If at a given time it is impersonally obligatory that some given agents *make it the case* that *p*, then at that time it is feasible for some agents that *p*.

I argue next, however, that this principle is also false. Suppose that you have decided to compete in two marathons that are scheduled on the same day, one in the morning and one in the afternoon. Suppose also that (1) it is obligatory *for my uncle*—and thus it is obligatory—that you finish the *first* marathon (because my uncle has promised that you will do so, and he can make it the case that

25 Given how I defined feasibility, it is feasible for us that we both win a gold medal exactly if there is something *we* can make the case such that, if we were to make it the case, then it might be the case that we both win a gold medal. To avoid my counterexample to IOIC_2, one might propose to define instead feasibility so that it is feasible for us that we both win a gold medal exactly if there is something *you* can make the case and there is something *I* can make the case such that, if you were to make the former the case and I were to make the latter the case, then it *might* be the case that we both win a gold medal. On the alternative definition of feasibility, it *is* feasible for us that we both win a gold medal: you can win and I can win, and if you were to win and I were to win, then (it would, and thus) it might be the case that we both win. In reply, I reject the alternative definition because it has the undesirable consequence that even if (1) you *would not* win if I were to win and (2) I *would not* win if you were to win, it is feasible for us that we both win.

26 Cf. note 2 above and the "stit paraphrase thesis" in Belnap et al., *Facing the Future*, 7–8.

you do so: he can give you a performance-enhancing pill), (2) it is (similarly) obligatory *for your aunt*—and thus it is obligatory—that you finish the *second* marathon, (3) it is possible that you finish both marathons, (4) no one can make it the case that you finish both marathons, because if you finished the first marathon then you would be so exhausted that (even if you took a pill) you *would not* finish the second marathon, and (5) there are no further normatively relevant considerations. Then (by 1, 2, 3, and 5) it is impersonally obligatory that (you make it the case that) you finish both marathons (as one can see by reasoning as in the voting example of section 1), but (by 4) it is not feasible for any agents that you finish both marathons: no matter what any agents were to make the case (among the things that they can make the case), it *would not* be the case (and thus it is false that it *might* be the case) that you finish both marathons. I conclude that IOIC_3 is false. (In this counterexample to IOIC_3, I can assume that it is due to "human nature"—whatever this means—that you would not finish the second marathon if you finished the first, so one cannot avoid the counterexample by redefining feasibility as compatibility with human nature.)[27]

To avoid my counterexamples to IOIC_2 and IOIC_3, one might retreat to a concept of feasibility even weaker than the very weak concept I used above: one might define feasibility as *historical possibility*—namely, as compatibility with all historical facts (and maybe also the laws of nature: see note 29 below). In fact, in the next section I defend the principle that impersonal 'ought' implies 'historically possible'. (Although whatever is historically possible in a sense *can* happen, this 'can' is defined without reference to the abilities of any agents; so I take the above principle to be a replacement for OIC rather than a version of OIC, and I stand by my claim that no version of OIC holds for impersonal obligatoriness.) One might claim that the above principle is too weak to be worth defending: Is it not obvious that whatever is (impersonally) obligatory is compatible with the historical facts? For example, how could the (historically impossible) proposition that the sun did *not* rise yesterday be obligatory (today)? I have three points in reply. First, it is not so obvious that *other* historically impossible propositions—for example, the proposition that people always keep their promises (which is historically impossible because some promises have been broken)—fail to be obligatory (today). Second, even if one finds a claim obvious, it is good to have an *argument* for the claim: after all, many apparently obvious claims (e.g., the claim that simultaneity is nonrelative) have turned out to be false. Third, my arguments in this section suggest that no significantly stronger replacement for OIC is defensible. Moreover, in the next section I also defend the following

27 See Estlund, "Human Nature and the Limits (If Any) of Political Philosophy"; and Southwood, "The Relevance of Human Nature."

replacement for the principle that 'ought' implies '*can avoid*': if a proposition is obligatory at a given time, then its *negation* is historically possible at that time. This principle contradicts the view of several authors that every logically necessary proposition is obligatory (see note 37 below).

3. IMPERSONAL 'OUGHT' IMPLIES 'HISTORICALLY CONTINGENT'

To define *historical contingency*, define first the *history* of the world up to and including a given time as the conjunction of all true propositions that are not about any later time. Say that a proposition is *historically necessary* (in other words, is *settled*) at a given time exactly if it is logically entailed by the history of the world up to and including that time.[28] For example, the proposition that the sun rose yesterday is historically necessary today. Say also that a proposition is *historically impossible* at a given time exactly if its negation is historically necessary at that time. For example, the proposition that the sun did not rise yesterday is historically impossible today. Finally, say that a proposition is *historically contingent* at a given time exactly if it is neither historically necessary nor historically impossible at that time (equivalently, exactly if both it and its negation are *historically possible*—i.e., not historically impossible—at that time). For example, the proposition that the sun will rise tomorrow is historically contingent today (in other words, today it is historically contingent that the sun will rise tomorrow).[29]

28 See Vranas, "I Ought, Therefore I Can Obey," 5n8. Cf. Lewis, *On the Plurality of Worlds*, 7; Thomason, "Indeterminist Time and Truth-Value Gaps"; van Fraassen, "A Temporal Framework for Conditionals and Chance," 94.

29 This is so even if the history of the world up to and including today *in conjunction with the laws of nature* logically entails that the sun will rise tomorrow. The laws of nature are not part of (more precisely, are not logically entailed by) the history of the world (up to and including today) because they are about *all* times, including *future* ones. But then, one might object, the principle that (1) impersonal 'ought' implies 'historically contingent' is too weak because it does not preclude nomologically impossible propositions that are historically contingent (e.g., the proposition that you will run faster than light) from being obligatory. I reply that the argument I will give for 1 can be easily modified (by replacing the history of the world with the conjunction of the laws of nature) to defend the principle that (2) impersonal 'ought' implies 'nomologically contingent'. So I can defend (and I accept) the conjunction of 1 with 2—namely, the principle that (3) impersonal 'ought' implies 'both historically contingent and nomologically contingent'—although for simplicity I consider only 1 in the text. (An alternative possible reply to the above objection is to define the *history** of the world up to and including a given time as the conjunction of *the laws of nature* with all true propositions that are not about any later time (cf. Lange, *Laws and Lawmakers*, 211n48) and to defend the principle—which is stronger than 3—that (4) impersonal 'ought' implies 'historically* contingent'. I do not adopt this reply because my argument for 1—in particular, its premise P1 (see below in the text)—cannot be modified to defend 4. I take this to be a good thing because 4 has, for example, the controversial

Given this terminology, I can formulate a principle that (I submit) holds for impersonal obligatoriness in lieu of ought-implies-can (although it *also* holds for personal obligatoriness, so I formulate it in terms of obligatoriness *simpliciter*)—namely, the principle that *obligatoriness implies historical contingency*:

> OIHC: If at a given time it is obligatory that *p*, then at that time it is historically contingent that *p*. More precisely: *by virtue of conceptual necessity, every proposition that is obligatory at a given time is historically contingent (i.e., neither historically necessary nor historically impossible) at that time*. Equivalently: *by virtue of conceptual necessity, no proposition that is either historically necessary or historically impossible at a given time is obligatory at that time*.

This principle is the conjunction of two principles:

> OIHC+: If at a given time it is obligatory that *p*, then at that time it is historically possible that *p* (i.e., it is *not* historically impossible that *p*).[30]

> OIHC–: If at a given time it is obligatory that *p*, then at that time it is historically possible that it is not the case that *p* (i.e., it is *not* historically necessary that *p*).[31]

---

consequence that if the history of the world up to and including today, in conjunction with the laws of nature, logically entails that I will kill you tomorrow, then it is not obligatory that I fail to kill you tomorrow.)

30 One might object to OIHC+ by claiming that, in the dog example (see section 1), if a dog strays into the courthouse at noon, then the proposition that there are never any dogs in the courthouse is historically impossible at noon but is still obligatory at noon (since the regulation that prohibits dogs in the courthouse is still in force at noon). I reply first that, assuming that the regulation is not retroactive, it does not prohibit the presence of dogs in the courthouse at times *prior* to its enactment, and thus it does not render obligatory at any time the proposition that there are *never* any dogs in the courthouse. Instead, for any time *t* starting at the time at which it takes effect, the regulation renders obligatory at *t* (and maybe also at some later times, although this is irrelevant for present purposes) the proposition P_t that there are no dogs in the courthouse at any time after *t* (and also, for any time interval that starts after *t*, the proposition that there are no dogs in the courthouse at any time in that interval, although this is again irrelevant for present purposes). If a dog strays into the courthouse at noon, then, for any time *t* prior to noon, P_t is historically impossible at noon, and thus (by OIHC+) is not obligatory at noon. But the proposition P_{noon} (that there are no dogs in the courthouse at any time *after* noon) is still historically contingent at noon, and is (compatibly with OIHC) still obligatory at noon. So the claim that it is still obligatory at noon that there be no dogs in the courthouse, understood as the claim that P_{noon} is still obligatory at noon, is compatible with OIHC (and thus with OIHC+). (Cf. Vranas, "I Ought, Therefore I Can Obey," 16.)

31 One might object to OIHC– by claiming that, since backwards causation is conceptually possible, the following scenario is also conceptually possible: in 2030, as I am about to

OIHC+ is structurally similar to OIC, and OIHC– is structurally similar to a principle that captures the idea that 'ought' implies 'can avoid'. To defend OIHC, I will defend first OIHC+ and then OIHC–. Let $O_t(P)$ be the claim that proposition P is obligatory at time t, and let $O_t(P|H_t)$ be the claim that P is (conditionally) obligatory at t given the history H_t of the world up to and including t. Given this notation, here is my argument for OIHC+:

P1. $O_t(P)$ conceptually entails $O_t(P|H_t)$.
P2. $O_t(P|H_t)$ conceptually entails $O_t(P \,\&\, H_t|H_t)$.
P3. $O_t(P \,\&H_t|H_t)$ conceptually entails that $P \,\&\, H_t$ is logically possible.

Therefore,

OIHC+: $O_t(P)$ conceptually entails that $P \,\&\, H_t$ is logically possible (i.e., P is historically possible at t: P is logically compatible with H_t).

Assuming that conceptual entailment is transitive, the argument is deductively valid.

P1 follows from the principle that $O_t(P)$ conceptually entails $O_t(P|Q)$ for any proposition Q that is historically necessary at t (as H_t is). To see that this principle is true, suppose that you are a soldier and your base is on alert, so your commander orders you at 5 PM to stand watch during the night (from midnight to 4 AM). Additionally, your commander decrees at 6 PM (without revoking the 5 PM order to you) that no one has to stand watch during the night if the alert is lifted before midnight. Then, assuming that there are no further normatively relevant considerations, it is obligatory (starting at 5 PM) that you stand watch during the night, but it is not (conditionally) obligatory (at any time starting at 6 PM) that (P) you stand watch during the night given that (Q) the alert is lifted before midnight. Suppose next that the alert is lifted at 8 PM, so the proposition that the alert is lifted before midnight is historically necessary starting at 8 PM. Then, starting at 8 PM, it is no longer obligatory that you stand watch during the night. More generally, if $O_t(P|Q)$ is false but Q is historically necessary at t, then $O_t(P)$ is also false—and this is equivalent to the above principle. Note

enter a time machine in my garage and travel back to 1930, you promise me that (P) the light in my garage will turn on shortly after I arrive in 1930, and in 2031 you push a button that causes the light in my garage to turn on shortly after I arrive in 1930. In this scenario (the objection continues), P is obligatory for you—and thus is obligatory—in 2030 but is historically necessary in 2030, contrary to OIHC–. (The objection assumes that there is only a single timeline, and so do I throughout this paper.) In reply, I submit that what your promise in the above scenario renders obligatory for you in 2030 is not P, but is instead the proposition R that you *make it the case* that P is true (e.g., by pushing the button in 2031): R is *not* historically necessary in 2030, so the claim that R is obligatory in 2030 is compatible with OIHC–.

that this reasoning does not carry over to propositions that are *not* historically necessary at *t*. In the above example, at 7 PM it is still obligatory that you stand watch during the night, although (a) it is not (conditionally) obligatory at 7 PM that you stand watch during the night given that the alert is lifted before midnight and (b) it is *true* (though at 7 PM not yet *historically necessary*) that the alert will be lifted before midnight.

P2 follows from the principle that $O_t(P|Q)$ is conceptually equivalent to $O_t(P \,\&\, Q|Q)$ for any proposition Q. To see that this principle is true, note that conditionalizing on Q amounts to "shrinking" the logical space (and all propositions) by considering only worlds at which Q is true; so, given (i.e., conditional on) Q, the obligatoriness (more generally, the deontic) status of P is the same as the status of the proposition that one gets by "shrinking" P—namely, $P \,\&\, Q$ (this is the proposition that one gets from P by considering only worlds at which Q is true).

Finally, P3 follows from the principle that (conditional) impersonal 'ought' implies 'logically possible': by virtue of conceptual necessity, every proposition that is (conditionally) obligatory at some time (or other) is logically possible. I take this principle to be relatively uncontroversial.

To defend next OIHC–, I note first that an argument parallel to my argument for OIHC+ can be given by replacing my talk of a proposition being *obligatory* with talk of a proposition being *forbidden* (i.e., *impermissible*). Letting $F_t(P)$ be the claim that proposition P is (*pro tanto*) forbidden at time t, and similarly for $F_t(P|H_t)$, here is the parallel argument: $F_t(P)$ conceptually entails $F_t(P|H_t)$, which in turn conceptually entails $F_t(P \,\&\, H_t|H_t)$, which in turn conceptually entails that $P \,\&\, H_t$ is logically possible, so $F_t(P)$ conceptually entails that P is historically possible at t (*impermissibility implies historical possibility*). Given this result, OIHC– quickly follows: $O_t(P)$ is conceptually equivalent to the claim that $\sim P$ (i.e., the negation of P) is forbidden at t, which by the above result conceptually entails that $\sim P$ is historically possible at t, so $O_t(P)$ conceptually entails that $\sim P$ is historically possible at t.[32] This concludes my argument for

32 The conceptual equivalence between the all-things-considered obligatoriness of P and the all-things-considered impermissibility of $\sim P$ is widely accepted in deontic logic (see, e.g., Belzer, "Deontic Logic"; Hilpinen and McNamara, "Deontic Logic," 43; McNamara, "Deontic Logic"; Rönnedal, *An Introduction to Deontic Logic*, 28–29), but it seems clear that there is also a conceptual equivalence (to which I appeal in the text) between the (*pro tanto*) obligatoriness of P and the (*pro tanto*) impermissibility of $\sim P$: for example, the claim that it is now obligatory that there be no dogs in the courthouse is conceptually equivalent to the claim that it is now forbidden that there be any dogs in the courthouse. Consequently, the principle that, by virtue of conceptual necessity, every proposition that is (conditionally) forbidden at some time (or other) is logically possible (which underlies my claim that $F_t(P \,\&\, H_t|H_t)$ conceptually entails that $P \,\&\, H_t$ is logically possible) is

OIHC–, and also my argument for OIHC. Note that since $O_t(P)$ is conceptually equivalent to $F_t(\sim P)$ and the claim that P is historically contingent at t is logically equivalent to the claim that $\sim P$ is historically contingent at t, OIHC is conceptually equivalent to the principle that, by virtue of conceptual necessity, every proposition that is *forbidden* at a given time is historically contingent at that time. It is not too hard to see that OIHC is, therefore, also conceptually equivalent to the principle that, *by virtue of conceptual necessity, every proposition that is either historically necessary or historically impossible at a given time is neither obligatory nor forbidden at that time*.[33]

One might argue that, even if OIHC is true, it is too weak because it only precludes propositions that are *wholly about the past* of a given time (e.g., the proposition that I skipped breakfast yesterday) from being obligatory at that time. I reply that this is not so: OIHC also precludes some propositions that are *not* wholly about the past of a given time from being obligatory at that time. For example, suppose that on Friday you take out a loan repayable in ten monthly installments; the first installment is due on Monday, but you fail to pay it on time. Then the proposition that *you pay all ten installments on time* is historically impossible on Tuesday but is *not* wholly about the past of Tuesday (i.e., *not* all times that the proposition is about are in the past of Tuesday) and is (by OIHC) *not* obligatory on Tuesday. To see that this result is correct, suppose for *reductio* that (1) on Tuesday it *is* obligatory that you pay all ten installments on time. Note that (2) the proposition that you pay all ten installments on time (which entails that you pay the first installment on Monday) is incompatible with the proposition (which is historically possible on Tuesday) that you pay the first installment on Tuesday or later: it is impossible to pay the first installment twice. But (3) if a proposition P is obligatory at time t and P is incompatible with a proposition

---

conceptually equivalent to the principle that, by virtue of conceptual necessity, no logically necessary proposition is (conditionally) obligatory at any time. See note 37 below for a possible objection to the latter principle.

33 It follows that every such proposition is not *all-things-considered* forbidden either at the given time (because being all-things-considered forbidden at a given time entails being (*pro tanto*) forbidden at that time). It does not follow, however, that every such proposition is all-things-considered *permissible* at the given time; defending this lies beyond the scope of the present paper, but for a defense, see Bedke, "Passing the Deontic Buck," 147–51; Olson, "In Defense of Moral Error Theory," 68–70, *Moral Error Theory*, 11–15, and "Error Theory in Metaethics," 60–62.

By ignoring P1 and considering only P2 and P3, my argument also supports the principle that $O_t(P|H_t)$ conceptually entails that P is historically contingent at t. Given that H_t is historically necessary at t, it follows that $O_t(H_t|H_t)$ is false.

Q which is historically possible at *t*, then *Q* is forbidden at *t*.[34] 1, 2, and 3 jointly entail that (4) on Tuesday it is forbidden that you pay the first installment on Tuesday or later. But since you do not pay the first installment on Monday, when it is due, (5) on Tuesday it is clearly *not* forbidden that you pay the first installment on Tuesday or later.[35] The contradiction between 4 and 5 completes the *reductio*, and I conclude that on Tuesday it is *not* obligatory that you pay all ten installments on time—and it counts in favor of OIHC that it explains why.

Barry Loewer and Marvin Belzer defend a principle that appears to contradict OIHC: "If the truth of *A* is settled at *t,* then at *t* it ought to be that *A*."[36] The contradiction may be only apparent: "it ought to be that *A*," as used by Loewer and Belzer, may not correspond to obligatoriness. Nevertheless, it may be worth noting some problems with the principle (which does contradict OIHC) that every proposition that is historically necessary (i.e., settled) at a given time is obligatory at that time.[37] First, suppose that you call me at 2 PM

34 Equivalently: if a proposition *P* is obligatory at time *t* and *P* entails a proposition *R* which is not historically necessary at *t*, then *R* is also obligatory at *t*. Some objections to this principle can be proposed (see Broome, *Rationality Through Reasoning*, 126; Heuer, "Reasons and Impossibility," 243–44; Kolodny, "Instrumental Reasons"; Raz, "The Myth of Instrumental Rationality"; White, "Transmission Failures"; contrast Kiesewetter, "Instrumental Normativity"), but they are irrelevant to the case at hand.

35 In fact, (6) on Tuesday it is *obligatory* that you pay the first installment on Tuesday or later: since you do not pay the first installment on Monday, when it is due, on Tuesday you must pay it as soon as possible. Although 6 does not entail 5 (since the impermissibility in 5 is *pro tanto*), 6 provides another argument against the claim that (1) on Tuesday it is obligatory that you pay all ten installments on time: if 6 and 1 are both true, then on Tuesday incompatible propositions are obligatory, which seems clearly false.

36 Loewer and Belzer, "Dyadic Deontic Detachment," 306; cf. "Help for the Good Samaritan Paradox," 125; Feldman, *Doing the Best We Can*, 43, 189.

37 Several authors accept the following weaker principle: every logically necessary proposition is obligatory (at every time). (See Anderson, "The Formal Analysis of Normative Systems," 181–83; Åqvist, "Deontic Logic," 616–17, and *Introduction to Deontic Logic and the Theory of Normative Systems*, 19–20; Bailhache, *Essai de Logique Déontique*, 17–19, 23–24; Hansson, "An Analysis of Some Deontic Logics," 380; Prior, "Escapism," 137–38; Segerberg, "Some Logics of Commitment and Obligation," 152; Stenius, "The Principles of a Logic of Normative Systems," 253; cf. Føllesdal and Hilpinen, "Deontic Logic," 13; van Fraassen, "The Logic of Conditional Obligation," 421. For rejections of the principle, see al-Hibri, *Deontic Logic*, 14–16; Carmo and Jones, "Deontic Logic and Contrary-to-Duties," 270, 294, 338; Dahl, "'Ought' Implies 'Can' and Deontic Logic," 501; Harrison, "More Deviant Logic," 23; Mares, "Andersonian Deontic Logic," 11–12; Pigden, "Logic and the Autonomy of Ethics," 139; Prior, *Formal Logic*, 221–22; van Rijen, review of *Doing the Best We Can*, 265; von Wright, "Deontic Logic," 10–11, "On the Logic of Norms and Actions," 8, and "Action Logic As a Basis for Deontic Logic," 60. On this debate, see also Humberstone, *Philosophical Applications of Modal Logic*, 246, and "Recent Thought on *Is* and *Ought*," 1429.) Defenders of the principle typically acknowledge that our intuitions (concerning,

and thus you satisfy your obligation (arising from your promise) to call me by 2 PM. Then, according to the above principle, after 2 PM it is (still) obligatory (since it is historically necessary) that you call me by 2 PM, and this remains obligatory forever (even after you die). Second, suppose that you kill me at 2 PM, and thus you violate your obligation to never kill me. Then, according to the above principle, after 2 PM it is obligatory (since it is historically necessary) that at some time (or other) you kill me. I take it that these consequences of the above principle are implausible enough to warrant rejecting the principle.[38]

---

for example, the claim that it is obligatory that either it is raining or it is not raining) are inconclusive or even go against the principle, but nevertheless accept the principle because it is "harmless" and it simplifies deontic logic. But given my rejection in the text of the stronger principle that every proposition that is historically necessary at a given time is obligatory at that time, it would, in fact, *complicate* deontic logic to accept that *some* historically necessary propositions (the logically necessary ones) are obligatory while other ones are not. Moreover, as al-Hibri (*Deontic Logic,* 15) notes, it seems false that it is *morally* (and also, I add, *legally,* etc.) obligatory that either it is raining or it is not raining.

38 Objecting to my argument for OIHC, one might claim that my argument commits me to the principle that $O_t(P)$ conceptually entails $O_t(P \,\&\, Q)$ for any proposition Q that is historically necessary at t (because $O_t(P)$ conceptually entails $O_t(P \,\&\, Q|Q)$ by the principles that I used to defend P1 and P2, and $O_t(P \,\&\, Q|Q)$ conceptually entails $O_t(P \,\&\, Q)$ by "unalterability detachment": see Nute and Yu, "Introduction," 9). But the above principle (and thus my argument for OIHC) should be rejected (the objection continues) because it has the consequence that (1) it is obligatory that both (P) I pay my taxes this year and (Q) Lincoln is assassinated at some time or other (assuming that P is obligatory, and given that Q is historically necessary), and this consequence is almost as implausible as the claim (which I reject) that (2) Q is obligatory—or so the objection goes. In reply, I grant that 1 *appears* false, but I submit that this is because there are three *apparently* sound but, in fact, unsound arguments against 1. *First argument*: 1 entails 2, and 2 is false, so 1 is false. This argument is unsound because 1 does not entail 2: $O_t(P \,\&\, Q)$ entails $O_t(Q)$ only if Q is *not* historically necessary at t. *Second argument*: $P \,\&\, Q$ is partly about the past (i.e., some of the times that $P \,\&\, Q$ is about are in the past), but (3) no proposition that is partly about the past of t is obligatory at t, so 1 is false. This argument is unsound because 3 is false (see Vranas, "I Ought, Therefore I Can Obey," 17): if at 8 AM you promise that you will never smoke again (starting immediately) and, for this reason, you acquire a corresponding obligation O, then at any time t shortly (e.g., one nanosecond) after 8 AM the proposition that you never smoke starting at 8 AM is partly about the past of t but is obligatory (because it is obligatory for you: otherwise, you would—implausibly—have obligation O for at most a single time instant—namely, at most at 8 AM). *Third argument*: Q is all-things-considered forbidden, and thus so is $P \,\&\, Q$ (since it entails Q), so 1 is false. This argument is invalid (because $P \,\&\, Q$ can be both *pro tanto* obligatory and all-things-considered forbidden), but is also unsound because, although Q was forbidden *before* Lincoln was assassinated, Q is no longer forbidden: it is implausible to claim that Q remains forbidden forever (cf. Vranas, "I Ought, Therefore I Can," 200-201n10, and "I Ought, Therefore I Can Obey," 9). In sum, I see no good reason to reject 1. Moreover, here is a scenario in which 1 is true: if I promise that $P \,\&\, Q$ will be true, then $P \,\&\, Q$ is obligatory *for me* (given that I can make it the case that $P \,\&\, Q$ is true: by paying my taxes, I can bring to completion a causal process—see

4. CONCLUSION

I argued that some propositions are impersonally obligatory—namely, obligatory *simpliciter* (i.e., morally required) but not personally obligatory (i.e., not morally required of anyone). This suggests, and some of my examples confirm, that obligatoriness *simpliciter* is irreducible to personal obligatoriness. I submit that claims of obligatoriness *simpliciter* tell us what is morally required from a standpoint that is distinct from—but takes into account and weighs against each other—the standpoints that correspond to particular agents.[39] In this respect, the distinction between obligatoriness for a given agent and obligatoriness *simpliciter* is analogous to the distinction between goodness for a given agent and goodness *simpliciter*.[40] I also argued that personal obligatoriness and obligatoriness *simpliciter* are subject to different constraints: personal obligatoriness is constrained by the abilities of agents (and also by historical contingency), whereas obligatoriness *simpliciter* is constrained by historical contingency but not by the abilities of agents. I conclude that personal obligatoriness and obligatoriness *simpliciter* are significantly different, and the distinction between them deserves further investigation.[41]

University of Wisconsin–Madison
vranas@wisc.edu

REFERENCES

Aas, Sean. "Distributing Collective Obligation." *Journal of Ethics and Social Philosophy* 9, no. 3 (2015): 1–22.

Vranas, "Therefore I Can Obey," 18n41—resulting in *P* & *Q* becoming historically necessary), and then (1) *P* & *Q* is obligatory. In no analogous scenario is 2 true (because, if I promise that *Q* will be true, *Q* is not obligatory for me since I cannot make it the case that *Q* is true), so I deny that 1 is almost as implausible as 2.

39 Cf. Nagel, *The Possibility of Altruism* and *The View from Nowhere*.

40 See Schroeder, "Value Theory"; cf. Hurka, "Against 'Good for'/'Well-Being', for 'Simply Good.'"

41 I am grateful to Matthew Chrisman, James Goodrich, Peter Graham, Alan Hájek, Ishtiyaque Haji, Lloyd Humberstone, Leon Leontyev, Russ Shafer-Landau, Alan Sidelle, Elliott Sober, Brian Talbot, Allard Tamminga, Michael Titelbaum, Peter Vallentyne, Bruno Whittle, Timothy Luke Williamson, Michael Zimmerman, and especially Aviv Hoffmann and two anonymous reviewers for comments. Thanks also to Joel Ballivian, Michael Bruckner, Lukas Myers, Alexander Pho, and an audience at the University of Missouri (April 2021, via Zoom) for interesting questions, and to my mother for typing the bulk of the paper.

Almotahari, Mahrad, and Brian Rabern. "The Onus in 'Ought'." *Analysis* 83, no. 1 (2023): 13–21.

Anderson, Alan Ross. "The Formal Analysis of Normative Systems." In *The Logic of Decision and Action,* edited by Nicholas Rescher. University of Pittsburgh Press, 1967.

———. "Logic, Norms, and Roles." *Ratio* 4, no. 1 (1962): 36–49.

Åqvist, Lennart. "Deontic Logic." In *Handbook of Philosophical Logic*. Vol. 2, *Extensions of Classical Logic,* edited by Dov M. Gabbay and Franz Guenthner. Reidel, 1984.

———. *Introduction to Deontic Logic and the Theory of Normative Systems.* Bibliopolis, 1987.

Bailhache, Patrice. *Essai de Logique Déontique.* Librairie Philosophique J. Vrin, 1991.

Bedke, Matt. "Passing the Deontic Buck." In *Oxford Studies in Metaethics*, vol. 6, edited by Russ Shafer-Landau. Oxford University Press, 2011.

Belnap, Nuel, Michael Perloff, and Ming Xu. *Facing the Future: Agents and Choices in Our Indeterminist World.* Oxford University Press, 2001.

Belzer, Marvin. "Deontic Logic." In *Routledge Encyclopedia of Philosophy*. Taylor and Francis, 1998. https://www.rep.routledge.com/articles/thematic/deontic-logic/v-1.

Björnsson, Gunnar. "Collective Responsibility and Collective Obligations Without Collective Moral Agents." In *The Routledge Handbook of Collective Responsibility,* edited by Saba Bazargan-Forward and Deborah Tollefsen. Routledge, 2021.

———. "Essentially Shared Obligations." *Midwest Studies in Philosophy* 38, no. 1 (2014): 103–20.

Blomberg, Olle, and Björn Petersson. "Team Reasoning and Collective Moral Obligation." *Social Theory and Practice* 50, no. 3 (2024): 483–516.

Bratman, Michael E. *Shared Agency: A Planning Theory of Acting Together.* Oxford University Press, 2014.

Broersen, Jan, and Leendert van der Torre. Review of *Agency and Deontic Logic*, by John F. Horty. *Artificial Intelligence and Law* 11, no. 1 (2003): 45–61.

Broome, John. *Rationality Through Reasoning.* Wiley-Blackwell, 2013.

———. "Williams on *Ought*." In *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams,* edited by Ulrike Heuer and Gerald Lang. Oxford University Press, 2012.

Campbell, Richmond. "Background for the Uninitiated." In *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem,* edited by Richmond Campbell and Lanning Sowden. University of British Columbia Press, 1985.

Carmo, José, and Andrew J. I. Jones. "Deontic Logic and Contrary-to-Duties." In *Handbook of Philosophical Logic*, vol. 8, 2nd ed., edited by Dov M. Gabbay and Franz Guenthner. Kluwer, 2002.

Castañeda, Hector-Neri. "On the Semantics of the Ought-to-Do." *Synthese* 21, nos. 3–4 (1970): 449–68.

Chisholm, Roderick M. "The Ethics of Requirement." *American Philosophical Quarterly* 1, no. 2 (1964): 147–53.

Chrisman, Matthew. "'Ought' and Control." *Australasian Journal of Philosophy* 90, no. 3 (2012): 433–51.

Collins, Stephanie. "Collectives' Duties and Collectivization Duties." *Australasian Journal of Philosophy* 91, no. 2 (2013): 231–48.

———. *Group Duties: Their Existence and Their Implications for Individuals.* Oxford University Press, 2019.

Cripps, Elizabeth. "Climate Change, Collective Harm and Legitimate Coercion." *Critical Review of International Social and Political Philosophy* 14, no. 2 (2011): 171–93.

Dahl, Norman O. "'"Ought" Implies "Can"' and Deontic Logic." *Philosophia* 4, no. 4 (1974): 485–511.

Danielsson, Sven. Review of *Agency and Deontic Logic*, by John F. Horty. *Philosophical Quarterly* 52, no. 208 (2002): 408–10.

Darwall, Stephen. "Why Obligations Can't Be Bipolar (Directed) All the Way Down." Unpublished manuscript, 2021.

Estlund, David M. "Human Nature and the Limits (If Any) of Political Philosophy." *Philosophy and Public Affairs* 39, no. 3 (2011): 207–37.

———. *Utopophobia: On the Limits (If Any) of Political Philosophy.* Princeton University Press, 2020.

Feldman, Fred. *Doing the Best We Can: An Essay in Informal Deontic Logic.* Reidel, 1986.

Finlay, Stephen, and Justin Snedegar. "One Ought Too Many." *Philosophy and Phenomenological Research* 89, no. 1 (2014): 102–24.

Føllesdal, Dagfinn, and Risto Hilpinen. "Deontic Logic: An Introduction." In *Deontic Logic: Introductory and Systematic Readings,* edited by Risto Hilpinen. Reidel, 1971.

Forrester, James William. *Being Good and Being Logical: Philosophical Groundwork for a New Deontic Logic.* M. E. Sharpe, 1996.

García, J. L. A. "The *Tunsollen*, the *Seinsollen*, and the *Soseinsollen*." *American Philosophical Quarterly* 23, no. 3 (1986): 267–76.

Geach, Peter T. "Whatever Happened to Deontic Logic?" *Philosophia* 11, nos. 1–2 (1982): 1–12.

Goble, Lou F. "Normative Conflicts and the Logic of 'Ought'." *Noûs* 43, no. 3

(2009): 450–89.
Guendling, John E. "Modal Verbs and the Grading of Obligations." *Modern Schoolman* 51, no. 2 (1974): 117–38.
Haji, Ishtiyaque. *Deontic Morality and Control.* Cambridge University Press, 2002.
Hansson, Bengt. "An Analysis of Some Deontic Logics." *Noûs* 3, no. 4 (1969): 373–98.
Hansson, Sven Ove. "The Varieties of Permission." In *Handbook of Deontic Logic and Normative Systems,* edited by Dov M. Gabbay, John F. Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre. College Publications, 2013.
Harman, Gilbert. *Change in View: Principles of Reasoning*. MIT Press, 1986.
———. "Relativistic Ethics: Morality as Politics." *Midwest Studies in Philosophy* 3, no. 1 (1978): 109–21.
Harrison, Jonathan. "More Deviant Logic." *Philosophy* 53, no. 203 (1978): 21–32.
Hartmann, Nicolai. *Ethics.* Vol. 1, *Moral Phenomena.* Translated by Stanton Coit. Macmillan and Allen and Unwin, 1932.
Held, Virginia. "Can a Random Collection of Individuals Be Morally Responsible?" *Journal of Philosophy* 67, no. 14 (1970): 471–81.
Helm, Bennett W. "Plural Agents." *Noûs* 42, no. 1 (2008): 17–49.
Heuer, Ulrike. "Reasons and Impossibility." *Philosophical Studies* 147, no. 2 (2010): 235–46.
Hibri, Azizah al-. *Deontic Logic: A Comprehensive Appraisal and a New Proposal.* University Press of America, 1978.
Hilpinen, Risto. "On the Semantics of Personal Directives." *Ajatus: Yearbook of the Philosophical Society of Finland* 35 (1973): 140–57.
Hilpinen, Risto, and Paul McNamara. "Deontic Logic: A Historical Survey and Introduction." In *Handbook of Deontic Logic and Normative Systems,* edited by Dov M. Gabbay, John F. Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre. College Publications, 2013.
Hintikka, Jaakko. "Some Main Problems of Deontic Logic." In *Deontic Logic: Introductory and Systematic Readings,* edited by Risto Hilpinen. Reidel, 1971.
Horty, John F. *Agency and Deontic Logic.* Oxford University Press, 2001.
———. "Agency and Obligation." *Synthese* 108, no. 2 (1996): 269–307.
———. *Reasons as Defaults*. Oxford University Press, 2012.
Horty, John F., and Nuel Belnap. "The Deliberative Stit: A Study of Action, Omission, Ability, and Obligation." *Journal of Philosophical Logic* 24, no. 6 (1995): 583–644.
Humberstone, I. Lloyd. *Philosophical Applications of Modal Logic*. College Publications, 2015.

———. "Recent Thought on *Is* and *Ought*: Connections, Confluences and Rediscoveries." *Journal of Applied Logics* 6, no. 7 (2019): 1373–1446.

———. "Two Kinds of Agent-Relativity." *Philosophical Quarterly* 41, no. 163 (1991): 144–66.

———. "Two Sorts of 'Ought's'." *Analysis* 32, no. 1 (1971): 8–11.

Hurka, Thomas. "Against 'Good for'/'Well-Being', for 'Simply Good'." *Philosophical Quarterly* 71, no. 4 (2021): 803–22.

Isaacs, Tracy. "Collective Responsibility and Collective Obligation." *Midwest Studies in Philosophy* 38, no. 1 (2014): 40–57.

———. *Moral Responsibility in Collective Contexts*. Oxford University Press, 2011.

Jansen, Ludger. "A Plural Subject Approach to the Responsibilities of Groups and Institutions." *Midwest Studies in Philosophy* 38, no. 1 (2014): 91–102.

Kiesewetter, Benjamin. "Instrumental Normativity: In Defense of the Transmission Principle." *Ethics* 125, no. 4 (2015): 921–46.

Kolodny, Niko. "Instrumental Reasons." In *The Oxford Handbook of Reasons and Normativity,* edited by Daniel Star. Oxford University Press, 2018.

Kordig, Carl R. "Relativized Deontic Modalities." In *The Logical Enterprise,* edited by Alan Ross Anderson, Ruth Barcan Marcus, and Richard M. Martin. Yale University Press, 1975.

Krogh, Christen, and Henning Herrestad. "Getting Personal: Some Notes on the Relationship Between Personal and Impersonal Obligation." In *Deontic Logic, Agency and Normative Systems*, edited by Mark A. Brown and José Carmo. Springer, 1996.

Kuhn, Steven. "Prisoner's Dilemma." In *Stanford Encyclopedia of Philosophy* (Winter 2019). http://plato.stanford.edu/archives/win2019/entries/prisoner-dilemma/.

Lange, Marc. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford University Press, 2009.

Lawford-Smith, Holly. "The Feasibility of Collectives' Actions." *Australasian Journal of Philosophy* 90, no. 3 (2012): 453–67.

Lewis, David K. *On the Plurality of Worlds*. Blackwell, 1986.

Liberman, Alida, and Mark Schroeder. "Commitment: Worth the Weight." In *Weighing Reasons,* edited by Errol Lord and Barry Maguire. Oxford University Press, 2016.

List, Christian, and Philip Pettit. "Group Agency and Supervenience." *Southern Journal of Philosophy* 44, no. S1 (2006): 85–105.

Loewer, Barry, and Marvin Belzer. "Dyadic Deontic Detachment." *Synthese* 54, no. 2 (1983): 295–318.

———. "Help for the Good Samaritan Paradox." *Philosophical Studies* 50, no. 1

(1986): 117–27.
Mares, Edwin D. "Andersonian Deontic Logic." *Theoria* 58, no. 1 (1992): 3–20.
Mason, Elinor. "Consequentialism and the 'Ought Implies Can' Principle." *American Philosophical Quarterly* 40, no. 4 (2003): 319–31.
May, Larry. "Collective Inaction and Responsibility." In *Individual and Collective Responsibility*, 2nd ed., edited by Peter A. French. Schenkman, 1998.
———. "Collective Inaction and Shared Responsibility." *Noûs* 24, no. 2 (1990): 269–77.
———. *Sharing Responsibility*. University of Chicago Press, 1992.
McConnell, Terrance. "'"Ought" Implies "Can"' and the Scope of Moral Requirements." *Philosophia* 19, no. 4 (1989): 437–54.
McNamara, Paul. "Agential Obligation as Non-Agential Personal Obligation Plus Agency." *Journal of Applied Logic* 2, no. 1 (2004): 117–52.
———. "Deontic Logic." In *Stanford Encyclopedia of Philosophy* (Summer 2019). http://plato.stanford.edu/archives/sum2019/entries/logic-deontic/.
———. Review of *Agency and Deontic Logic*, by John F. Horty. *Mind* 113, no. 449 (2004): 179–85.
Meinong, Alexius. *On Emotional Presentation.* Translated by Marie-Luise Schubert Kalsi. Northwestern University Press, 1972.
Nagel, Thomas. *The Possibility of Altruism*. Clarendon Press, 1970.
———. *The View from Nowhere*. Oxford University Press, 1986.
Nute, Donald, and Xiaochang Yu. "Introduction." In *Defeasible Deontic Logic*, edited by Donald Nute. Kluwer, 1997.
Olson, Jonas. "Error Theory in Metaethics." In *The Routledge Handbook of Metaethics,* edited by Tristram McPherson and David Plunkett. Routledge, 2017.
———. "In Defense of Moral Error Theory." In *New Waves in Metaethics,* edited by Michael Brady. Palgrave Macmillan, 2011.
———. *Moral Error Theory: History, Critique, Defence*. Oxford University Press, 2014.
Pettit, Philip. "Responsibility Incorporated." *Ethics* 117, no. 2 (2007): 171–201.
Pigden, Charles R. "Logic and the Autonomy of Ethics." *Autralasian Journal of Philosophy* 67, no. 2 (1989): 127–51.
Pinkert, Felix. "What We Together Can (Be Required to) Do." *Midwest Studies in Philosophy* 38, no. 1 (2014): 187–202.
Price, A. W. *Contextuality in Practical Reason.* Clarendon Press, 2008.
Priest, Graham. *An Introduction to Non-Classical Logic: From If to Is.* 2nd ed. Cambridge University Press, 2008.
Prior, Arthur N. "Escapism: The Logical Basic of Ethics." In *Essays in Moral Philosophy,* edited by Abraham I. Melden. University of Washington Press, 1958.
———. *Formal Logic*. Clarendon Press, 1955.

Raz, Joseph. "The Myth of Instrumental Rationality." *Journal of Ethics and Social Philosophy* 1, no. 1 (2005): 2–28.
Robinson, Richard. "Ought and Ought Not." *Philosophy* 46, no. 177 (1971): 193–202.
Rönnedal, Daniel. *An Introduction to Deontic Logic*. Published by the author, 2009.
Ross, Jacob. "The Irreducibility of Personal Obligation." *Journal of Philosophical Logic* 39, no. 3 (2010): 307–23.
Schroeder, Mark. "*Ought*, Agents, and Actions." *Philosophical Review* 120, no. 1 (2011): 1–41.
———. "Value Theory." In *Stanford Encyclopedia of Philosophy* (Fall 2016). https://plato.stanford.edu/archives/fall2016/entries/value-theory/.
Schwenkenbecher, Anne. *Getting Our Act Together: A Theory of Collective Moral Obligations.* Routledge, 2021.
———. "Joint Duties and Global Moral Obligations." *Ratio*, n.s., 26, no. 3 (2013): 310–28.
———. "Joint Moral Duties." *Midwest Studies in Philosophy* 38, no. 1 (2014): 58–74.
Segerberg, Krister. "Some Logics of Commitment and Obligation." In *Deontic Logic: Introductory and Systematic Readings,* edited by Risto Hilpinen. Reidel, 1971.
Sidgwick, Henry. *The Methods of Ethics.* 7th ed. Hackett, 1981.
Smith, Holly M. "Moral Realism, Moral Conflict, and Compound Acts." *Journal of Philosophy* 83, no. 6 (1986): 341–45.
Southwood, Nicholas. "Does 'Ought' Imply 'Feasible'?" *Philosophy and Public Affairs* 44, no. 1 (2016): 7–45.
———. "The Feasibility Issue." *Philosophy Compass* 13, no. 8 (2018): e12509.
———. "The Relevance of Human Nature." *Journal of Ethics and Social Philosophy* 9, no. 3 (2015): 1–8.
Stenius, Erik. "The Principles of a Logic of Normative Systems." *Acta Philosophica Fennica* 16 (1963): 247–60.
Thomason, Richmond H. "Indeterminist Time and Truth-Value Gaps." *Theoria* 36, no. 3 (1970): 264–81.
Tomalty, Jesse. "The Force of the Claimability Objection to the Human Right to Subsistence." *Canadian Journal of Philosophy* 44, no. 1 (2014): 1–17.
Van Fraassen, Bas. C. "The Logic of Conditional Obligation." *Journal of Philosophical Logic* 1, nos. 3–4 (1972): 417–38.
———. "A Temporal Framework for Conditionals and Chance." *Philosophical Review* 89, no. 1 (1980): 91–108.
———. "Values and the Heart's Command." *Journal of Philosophy* 70, no. 1

(1973): 5–19.
Van Rijen, Jeroen. Review of *Doing the Best We Can*, by Fred Feldman. *Noûs* 27, no. 2 (1993): 264–67.
Von Wright, Georg Henrik. "Action Logic as a Basis for Deontic Logic." In *Normative Structures of the Social World,* edited by Giuliano di Bernardo. Rodopi, 1988.
———. "Deontic Logic." *Mind* 60, no. 237 (1951): 1–15.
———. "On the Logic of Norms and Actions." In *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics,* edited by Risto Hilpinen. Reidel, 1981.
Vranas, Peter B. M. "I Ought, Therefore I Can." *Philosophical Studies* 136, no. 2 (2007): 167–216.
———. "I Ought, Therefore I Can Obey." *Philosophers' Imprint* 18, no. 1 (2018): 1–36.
———. "'Ought' Implies 'Can' but Does Not Imply 'Must': An Asymmetry Between Becoming Infeasible and Becoming Overridden." *Philosophical Review* 127, no. 4 (2018): 487–514.
Wedgwood, Ralph. "The Meaning of 'Ought'." In *Oxford Studies in Metaethics*, vol. 1, edited by Russ Shafer-Landau. Clarendon Press, 2006.
White, Stephen J. "Transmission Failures." *Ethics* 127, no. 3 (2017): 719–32.
Wiens, David. "Political Ideals and the Feasibility Frontier." *Economics and Philosophy* 31, no. 3 (2015): 447–77.
Williams, Bernard. "*Ought* and Moral Obligation." In *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press, 1981.
Wringe, Bill. "Collective Obligations: Their Existence, Their Explanatory Power, and Their Supervenience on the Obligations of Individuals." *European Journal of Philosophy* 24, no. 2 (2016): 472–97.
———. "From Global Collective Obligations to Institutional Obligations." *Midwest Studies in Philosophy* 38, no. 1 (2014): 171–86.
———. "Global Obligations and the Agency Objection." *Ratio*, n.s., 23, no. 2 (2010): 217–31.
———. "Global Obligations, Collective Capacities, and 'Ought Implies Can'." *Philosophical Studies* 177, no. 6 (2020): 1523–38.
———. "Needs, Rights, and Collective Obligations." In *The Philosophy of Need*, edited by Soran Reader. Cambridge University Press, 2005.

Journal of Ethics and Social Philosophy
VOL. 29, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.3449

KANTIAN FREE RIDING

Jan Willem Wieland

IMAGINE that pollution from boats fishing a local lake has become serious enough to affect the catch. One thousand fishers agree to do their part in cleaning the lake in order to save their livelihood.[1] As one of the fishers, I reason as follows:

> Either enough others do their part, or too many others do not. On the one hand, if enough others do their part, the lake will be sufficiently clean for the fish, and I do not need to contribute my bit. On the other hand, if too many others do not do their part, the fish population will die anyway, and doing my bit will again be a mere waste. Either way, it is better for me to do nothing.

The same reasoning, however, holds for any of the others, and if too many people act in this way, the fish will die, and we all lose our jobs. Importantly, I *do* care about collective success: a clean lake and a healthy fish population. After all, my livelihood depends on it. It is just that collective success depends on whether *enough* people are prepared to spend their time cleaning the lake, not on whether *I* do so.

This scenario relies on an assumption that the group is of such a size that one person's contribution will not make any relevant difference.[2] Of course, I can make *a* difference. If I remove some pieces of plastic from the lake, I might save one fish. Yet what we are concerned about here is the fish population as a whole and whether it is healthy enough to reproduce so that all fishers—including me—can make a living. The assumption is that no single person's contribution makes a difference to *that*.

But even though I care about collective success, I also want to avoid unnecessary costs. It is not that I really want to free ride on others and only want to defect in secret when enough others cooperate. I just do not want to waste my time and energy. In light of this, the question is: *Why cooperate*?

1 This case is adapted from Cullity, "Moral Free Riding," 11.

2 See, e.g., Parfit, *Reasons and Persons*, ch. 3.

Table 1. Social Dilemmas

	Enough others cooperate	Too many others defect
I cooperate	Collective success + cooperation costs	Collective failure + cooperation costs
I defect	Collective success	Collective failure

Dilemmas with the structure presented in table 1 are called *social dilemmas*. In the literature, we find at least two prominent types of replies to why one should cooperate in these dilemmas.[3] Consequentialists point out that this presentation of the decision situation is incomplete: there is still a chance, however small, that defecting will make a difference to collective success, and we should not run this risk.[4] Others have suggested that even when the chance of making a difference to collective success is too small or entirely absent, one may still help or play an instrumental or causal role in bringing it about.[5]

In this paper, I follow a diametrically opposed approach by assuming that there can be reasons to cooperate even in the absence of instrumental considerations. That is, even when one's cooperation has zero impact and fails to make any relevant instrumental contribution (e.g., in cases that involve simply too many parties or where enough parties are already cooperating and thereby guaranteeing collective success), defecting may still be problematic because it is unfair and not *universalizable*.[6] Simply put, defecting lets others do the work and "makes an exception of oneself."

This paper will contrast two opposite elaborations of this basic idea. On the one hand, there is an *others-based sense* of making an exception of oneself: others who cooperate prefer not to pay the cooperation costs but do not act on that preference; if *you* do act on it, you make an exception of yourself in this first sense. On the other hand, there is an *agent-based sense*: the agent who defects prefers others to cooperate and in this way makes an exception of herself. In the following, I explore this second interpretation, which is Kantian in nature and focused on the agent's mindset.

3 For an overview, see Nefsky, "Collective Harm and the Inefficacy Problem."

4 See, e.g., Kagan, "Do I Make a Difference?"; and Hedden, "Consequentialism and Collective Action."

5 See, e.g., Braham and van Hees, "An Anatomy of Moral Responsibility"; Nefsky, "How You Can Help"; and Gunnemyr and Touborg, "Reasons for Action."

6 Lomasky and Brennan suggest that what matters is unfairness, not universalizability ("Is There a Duty to Vote?" 77). In a sense, I agree. What matters is unfairness, though the Kantian test—as I take it—offers a particular interpretation of this notion.

The ambition is to defend this approach against one key problem—namely, how it can distinguish unfair free riding from innocent coordination.[7] In response to this problem, I will argue that what goes wrong in free riding cases is that agents fail to make their conduct conditional on other people's preferences. In innocent coordination cases, basically, people care about what others prefer (e.g., to play tennis now or later) and let their conduct depend on this. In unfair free riding cases, in contrast, people are indifferent to what others prefer. They defect (e.g., fail to do their part in the cleaning of the lake) regardless of whether others prefer to defect too. Only the latter, I propose, fails the Kantian test.

I start by explaining the Kantian test that I employ, as well as the problem with it (section 1). Next, I look critically at existing proposals to solve this problem (section 2). Finally, I set forth a new account (sections 3 to 5). Importantly, my aims are systematic rather than interpretative. That is, I propose a novel way to distinguish free riding from coordination, but the proposal is not intended to originate in any way in Kant's own writings (though some of the authors I discuss do have this different focus).

1. THE KANTIAN TEST

Why is it problematic to make a false promise—lie that one will pay one's debts later—as a means to get money? Korsgaard's classic analysis is this.[8] Imagine a world where making a false promise is the standard means to get money, and ask whether one can still achieve one's purpose (here, getting money) by taking the given means (here, lying) in that hypothetical world. This is not the case: in a world where everyone makes false promises, no one would believe them, and therefore one would not be able to obtain any money in this way. The Kantian test, schematically, is:

1. *Maxim*: "To achieve purpose *P*, I will do action *A*."
2. *Universalization*: Imagine that *A* is the standard procedure for achieving *P*, i.e., all who pursue *P* do *A* as a means to this end.
3. *Test*: In this world, can I still achieve *P* by doing *A*? If not, I run into a *practical contradiction*.[9]

7 In this paper, I use the term 'free riding' broadly: it covers not only not paying for public, nonexcludable goods but defecting more generally, including the case of enslaving others. Pettit labels the latter "foul dealing" ("Free Riding and Foul Dealing").

8 Korsgaard, "Kant's Formula of Universal Law."

9 This is Korsgaard's account of the formula of universal law, in particular of the "contradiction in conception" test. I do not have the space to address alternative interpretations. For the purposes of this paper, what matters is this overall agent-based diagnosis of making an exception of oneself, as contrasted to the others-based diagnosis that I will discuss later.

This test offers a powerful analysis of numerous cases. For example, slavery is morally problematic because in a hypothetical world where everyone keeps slaves, you are enslaved yourself and cannot avoid working. Why is this problematic? The deeper diagnosis is this: if you run into a practical contradiction, you are making an exception of yourself and assume that others will act differently—for example, that they will not keep *you* as a slave. Similarly, if you want to free ride in a social dilemma, you are assuming that enough others will *not* free ride (i.e., in order to be able to free ride on something in the first place), and in this way you are assuming that you are more important than they are. That is what is morally problematic.

The Kantian test offers a straightforward analysis of our social dilemma. I will not do my part in cleaning the lake because I want to enjoy my life and spend my time on things I like better than cleaning the lake. In a world where everyone slacks, there is no clean lake, there are no more fish, and I am jobless. In such a world, I have no money to survive and cannot enjoy my life. Practical contradiction. I assume that other people do clean the lake yet make an exception of myself.[10]

One may point out that much depends on the exact formulation of my purpose. What if we formulate it not as "to enjoy my life" but simply as "to avoid spending unnecessary energy"? In a world where everyone slacks, it seems that I can still avoid spending unnecessary energy.[11] Does this mean that defecting is universalizable after all and thus nonproblematic? That does not sound right. Defecting is universalizable only if *no relevant purpose* is frustrated. Which purposes matter in this context? In social dilemmas, we said, people are interested in two things:

i. collective success; and
ii. not wasting cooperation costs.[12]

10 To be sure, if I do not really care about a clean lake (e.g., if there are alternative ways for me to make money), then this practical contradiction does not arise.

11 One may wonder if this is actually true. After all, in such a world I must work even harder than I did before, e.g., travel to a different lake to fish or learn a new trade or grow my own food to survive, which all take energy.

12 We assume that *P* in the test ranges over all interests the agent has while pursuing *A*, i.e., all interests that may be explicitly or implicitly endorsed by her. See Korsgaard on implicit purposes that can be frustrated ("Kant's Formula of Universal Law," 41–42). We do not need to assume that people always have interest in both i and ii, just that people who face social dilemmas have these preferences. Note that ii need not be self-interested. Saying that one does not want to waste unnecessary efforts is not the same as saying that one wants to have as much as possible for oneself.

Sometimes i is implicit. A certain fisher might think he just wants to enjoy himself rather than spend unnecessary time cleaning. But as long as his enjoyment depends on the quality of the lake, he cares about collective success, at least implicitly. Moreover, fishers assign greater value to i than to ii. A clean lake is way more valuable to them than a day off. They would be happy to clean if that would mean that they can keep their jobs. But again, collective success depends on what many people do, not on what they individually do.

Now, even though ii may not be frustrated in a hypothetical world where everyone defects, the more important concern, i, still would be frustrated. Defecting in social dilemmas is not universalizable in exactly this sense: I make an exception of myself and assume that enough people act differently *to realize collective success*.[13]

So far, so good. As promising as the Kantian test appears, it suffers from a major objection—namely, that it fails to distinguish problematic free riding from innocent coordination. Consider Scanlon's well-known counterexample: "To avoid waiting for an empty court, I will play tennis on Sunday morning."[14] In a hypothetical world where everyone acts on this maxim and plays tennis on Sunday morning, it is super crowded, and I cannot avoid waiting for an empty court. Practical contradiction. Still, it would be absurd to think that it is wrong to act on this maxim. What goes for the tennis maxim, goes for coordination cases generally.[15] Thus Herman:

> I select my driving route to school by observing where others do not like to go. I go to the movies at six o'clock because there are crowds at eight o'clock. The intention is to do what others are not doing. The condition of success for such actions is that others not act the same way.[16]

In both free-riding and coordination cases, the agent is assuming that other people act differently. Yet in coordination cases, that is just fine. But what is the difference? Why is making an exception of oneself fine in some cases but not in others? Note that in a sense, one might not really be "making an exception

13 I borrow this Kantian analysis from Wieland, "Cooperation, Kantian-Style." Albertzart offers a different Kantian analysis ("A Kantian Solution"). According to Wieland, you should cooperate because you should not leave the work of solving your problems to others. According to Albertzart, in contrast, you should cooperate because you should not leave the work of solving *other people's* problems to others.

14 Herman, *The Practice of Moral Judgment*, 138.

15 As well as for further cases: "Some poor people get their food by searching through the rubbish that others throw away. That method must be exceptional, but is not wrong, or unfair" (Parfit, *On What Matters*, 1:284).

16 Herman, *The Practice of Moral Judgment*, 139.

of oneself" in coordination cases, but the question is whether the Kantian test can make the difference. One may suspect that the answer is negative. Wood has stated this clearly:

> [Korsgaard's test] is obviously mistaken because there are clearly any number of quite innocent actions that depend for their success on the fact that they will be exceptional, that others will choose not to do anything similar. . . . The principal kind of "exceptional" behavior which suits Korsgaard's remark is "free riding." But any plausible moral objection to free riding presupposes the existence of a determinate moral principle or duty with which everyone is supposed to comply.[17]

Wood's objection is this. Both free riding and coordination cases share the same structure, and the Kantian test will treat them alike. If we want to distinguish between them, we need some principle that is external to the Kantian test. For example, we may say that in contrast to the tennis player, the fishers are under some moral duty, but only because there is some principle in place—other than the Kantian test—that grounds a duty to do one's part in cleaning the lake but not to stay home on Sunday morning.

The Kantian test has generated a respectable track record of controversies. Philosophers have proposed numerous false positives ("the test is empty") and false negatives ("the test is too strong") and made various attempts to counter them. The current problem poses a special challenge. If the Kantian test is unable to distinguish unfair free riding from innocent coordination, the whole test is flawed.[18] Moreover, if it "presupposes the existence of a determinate moral principle," as Wood puts it, we might as well skip the test altogether.[19] It is this paper's ambition to answer this problem.

17 Wood, *Kant's Ethical Thought*, 108.

18 Contrast further objections such as the false positives (e.g., killing from despair) identified by Korsgaard, "Kant's Formula of Universal Law," 42–43. As Korsgaard suggests, the Kantian test is not really designed for such maxims, and they are also rarely acted upon. One cannot say the same thing about coordination maxims (a subset of false negatives). They are not rare, and the test should be able to handle such everyday maxims.

19 Given this problem, Herman suggests that we should not test such specific maxims as the coordination ones but test only "generic maxims" such as "making a false promise for self-interested purposes." Moreover, the idea is to take the outcome of the test not too strictly but only as input for our moral deliberation (*The Practice of Moral Judgment*, ch. 7). Wood suggests invoking the formula of humanity rather than universal law (Wood, *Kant's Ethical Thought*, 110). Coordination maxims do not seem to violate the other Kantian formula. The very ambition of this paper is to determine if there is also a solution for Korsgaard's initial account.

On the one hand, Herman is right that in both free riding and coordination cases, "the condition of success is that others not act the same way." On the other hand, these cases do not look *exactly* the same. In the tennis case, many people are not interested in playing on Sunday morning. They want to go to church, be in bed, have an extended breakfast, or do other sports. We could agree on coordinating our actions, and everyone would be fine. The same does not apply to problematic free riding cases. It is not as if I would take a slave (or make a false promise, refuse to clean, etc.), others would not, and we would all be fine.

In light of this, one may want to distinguish different kinds of norms.[20] On the one hand, there are moral norms whereby we expect that others *should* act in some way (e.g., do not enslave others). On the other hand, there are descriptive norms whereby we expect that others actually act in some way (e.g., do not play tennis on Sunday morning) but not that they also *should* do so. Just appealing to such a distinction, however, will not solve our problem.[21] What we want to know is whether the Kantian test can make the difference.

2. EXISTING SOLUTIONS

What can Kantians do? There are two broad strategies: either claim that the counterexamples (the coordination maxims) are ill formed and offer specific instructions for maxim reformulation; or leave the maxims as they are but tweak the universalization step. I will look at the most promising existing proposals, starting with an instance of the second strategy by Pogge.[22]

In both free-riding and coordination cases, I want to do *A* ("play tennis on Sunday morning" or "take a slave") but not that others do the same. Yet if we look more closely at everyone's preferences, there is also a difference. In coordination cases, as a matter of fact, other people do not prefer to do *A* ("play tennis on Sunday morning"), while this is less clear in free-riding cases. Others may also want to take a slave. Given this, we may tweak the universalization step "imagine that all who pursue *P*, do *A* as a means" to:

> 2*. Imagine that all who pursue *P and actually prefer A as a means to P*, do *A* as a means to achieve *P*.

20 E.g., Bicchieri, *Norms in the Wild*.

21 Contractualists à la Scanlon may suggest that what distinguishes the two cases is that others can reasonably object to my justification in the free riding cases ("I only want to enjoy my free time!") but not to my justification in the coordination cases ("I want to avoid crowded courts"). See Scanlon, *What We Owe to Each Other*.

22 Pogge, "The Categorical Imperative."

In the Kantian test, then, we do not imagine that everyone plays tennis on Sunday morning but rather that only those who actually want to play tennis on Sunday morning (as a means to avoid crowded tennis courts) play on Sunday morning. That is much like the actual world, and the alleged practical contradiction disappears. Surely it is still possible to avoid crowded tennis courts in a world where no one wants to play on Sunday morning. I do not make an exception of myself in that other people do not want the same.

However, 2* does not work in the slavery case. Many people might not want to do *A* (take a slave) as a means to *P* (avoid working) because they care about other people. If this is so, "To avoid working, I will take a slave" may well pass the test. In a world where hardly anyone is enslaved—because people consider that morally problematic—I might not be enslaved and thereby avoid a practical contradiction. In light of this and following Pogge, we may tweak the universalization one step further:

> 2**. Imagine that all who pursue *P* and would prefer *A* as a means to *P* had *A* not been morally problematic to them, do *A* as a means to achieve *P*.[23]

In a world where no one is obstructed by moral concerns, many would still prefer to stay in bed on Sunday morning. However, when people are no longer obstructed by moral concerns regarding enslaving others, many might well choose to take slaves to work for them. And if they do, I will likely be enslaved in that world and hence no longer be able to avoid working. Practical contradiction. In social dilemmas too, if it were morally okay to defect (e.g., not help out with cleaning the lake), people would simply defect in order to avoid wasting cooperation costs—and hence run into a practical contradiction since they would no longer benefit from collective success.

Even though Pogge's 2** seems promising, it has not received widespread acceptance. It is quite a complex step to apply, as we have to go to hypothetical worlds that are very different from our own. How do we know whether people would take a slave if doing so were morally fine, and how do we know that enough people would do it so that I would run into a practical contradiction?[24]

A further concern with 2**, which I will clarify later, is that it is too focused on what *other* people prefer. As I see it, the Kantian test should identify agents

23 Per Pogge, "an agent is permitted … to adopt some given maxim just in case he can will that everyone be permitted to adopt it.… Other things being unchanged, can he will our world to be such that everyone feels (morally) free to and those so inclined ('by nature') actually do adopt this maxim?" ("The Categorical Imperative," 190).

24 As Kerstein adds, people might still choose not to keep slaves for other, nonmoral reasons (*Kant's Search for the Supreme Principle of Morality*, 171–74).

who make an exception of themselves in exactly this sense: *they* prefer other people—but not themselves—to cooperate. Therefore, let us move to proposals that focus more on the agent's own maxims.

According to McCarty, actions like playing tennis on Sunday morning take place only under certain terms. In the tennis case, it is not the case that one can play whenever one wants. Sometimes the courts are full, and then, according to the rules of many clubs, one has to wait one's turn. Hence, McCarty claims, the act should be described not as "playing tennis" but rather as "playing tennis or waiting one's turn."[25] It is right that one cannot play tennis in a world where everyone plays at once. But in such a world, one can still play *or wait one's turn*. As McCarty concludes, "When the maxims . . . are formulated so as to include references to the background policies or agreements they presuppose, they easily convert from false negatives to true positives."[26]

We may accept that McCarty's suggestion blocks a *logical contradiction*. One runs into a logical contradiction basically when it is no longer possible to perform the action after universalization. As just seen, the action in the tennis case, if properly described, does not face this problem. However, even when there is no logical contradiction, there might still be a practical one. I want to find an empty court; in a world where everyone goes to the club on Sunday morning, I can still play or wait my turn, but I cannot avoid crowded courts. The practical contradiction remains.

Actions do not take place only under certain terms but are usually also intended only under certain conditions. McCarty offers this example: "If I turn one hundred, I will buy a red sports car."[27] To test this maxim, then, we should not imagine that the whole world population is buying red sports cars but only those who reach one hundred years of age. Another instructive example is offered by Kagan: "To get lunch, I will go to some local pizza restaurant in Naples but only if I want pizza, am nearby, and the restaurant has place for me."[28] To test this maxim, again, we should not imagine that billions are trying to crowd into a single restaurant but rather imagine that only those who want pizza, are nearby, and the restaurant has place for them are trying. That yields no practical contradiction.[29]

25 McCarty, "False Negatives of the Categorical Imperative," 185.

26 McCarty, "False Negatives of the Categorical Imperative," 186.

27 McCarty, "False Negatives of the Categorical Imperative," 183.

28 Kagan, "Kantianism for Consequentialists," 138.

29 Such more fully described maxims are more adequate descriptions of an agent's intentions. Thus Kagan states, "I do not have reason to go to Naples regardless of how crowded it is, how inconvenient it is to get to it, and so on" ("Kantianism for Consequentialists," 138).

According to Cholbi too, such *conditionalization* is the key to accounting for coordination cases. For example, according to him, the tennis maxim may be reformulated as "playing tennis *when the courts are available and my schedule permits it*."[30] In a world where everyone plays tennis when the courts are available—whether this be Sunday morning or some other time—it is still possible to avoid crowded courts. The practical contradiction disappears.

However, what details should be included in a maxim? To justify his reformulation, Cholbi invokes the following counterfactual test.

> Include in a maxim only those descriptions which, if altered, would lead the agent to act differently.[31]

In the tennis case, we may ask: What if the courts were free on Saturday rather than Sunday? Would I still want to play on Sunday? Presumably not. I simply want to avoid waiting for a court and want to play when I am free and there is space. If this is so, the detail of Sunday morning is irrelevant according to this test and can, as Cholbi proposes, indeed be omitted.[32]

The counterfactual test removes many irrelevant details in this way. Unfortunately, however, it also removes *relevant* details. Consider the maxim "To avoid working, I will take a slave." Applying the counterfactual test, we may ask: What if I could not get away with it? Or what if others would enslave me in turn? Would I then still take a slave? Arguably not. But then we should add all sorts of conditions: "I will take a slave, but only when I can get away with it, when others would not enslave me, etc." Such a maxim avoids a practical contradiction.[33]

Some have suggested that we should even go more general and describe the tennis maxim as "maintaining my physical well-being" or as "developing my

30 Cholbi, *Understanding Kant's Ethics*, 153.

31 Galvin, "Maxims and Practical Contradictions," 408–9. This test was suggested in O'Neill, *Acting on Principle*, 107. Suppose I drink a cup of coffee in the morning. Ask: What if there was just water in the cup, would I still drink it? If so, the detail of coffee is irrelevant and should not be included in the maxim. I am just drinking to quench my thirst. But if I would not drink it if it did not contain caffeine, the detail about coffee *is* relevant. Galvin does not accept this test since it removes relevant details as well.

32 What if you want to play *only* on Sunday, and so the detail is relevant? See section 3 below.

33 There is more. Would I still keep a slave if I could take different measures to avoid working? Arguably not. But then we should not mention that detail and just speak of "taking smart strategies." In that case, we do not imagine a world where everyone keeps slaves (and so where I am enslaved myself). Instead, people might be doing various things: some keep slaves, others take regular employees, some buy lottery tickets, others pray for a miracle, still others invest in lucrative businesses, and so on. Again, the practical contradiction disappears.

talents."[34] Such maxims pass the Kantian test, though the issue is that we would need some criterion (like the counterfactual test just discussed) of *when* maxims may be generalized in this way. Without such a criterion, there is no reason why we would not generalize "spending time on my own hobbies and taking a slave to work for me" or "spending time on my own hobbies and not doing my part in cleaning the lake" in a similar way—which then would become false positives.

Despite these worries, one takeaway message of these proposals is that maxims should not include irrelevant details like "Sunday morning."[35] Moreover, if we merely say that you go to the club *when many others do not*, the practical contradiction disappears: it is still possible to avoid waiting when people go to the club when many others do not. This would also help in other coordination cases. In a world where everyone tries to enter a particular building at nine in the morning, there is a long line, and I cannot enter. The same problem does not arise for "To enter the building, I will go through the door *when others do not*."

However, Herman reminds us that a similar move is available in the bad cases.[36] For example, to avoid a practical contradiction in the slavery case, I may say that I act on the maxim "To avoid working, I will take a slave *when others do not take me as a slave*." In a world where others act like me, I will still escape enslavement and pass the Kantian test. Hence, what we would need is a compelling story on *why* this latter maxim is ill formed and not to be tested.

Here would be such a story. In the tennis case, the addition "when others do not play" describes how the agent actually attempts to achieve *P* ("avoid waiting"). In coordination cases generally, details about the conduct of others are relevant in this way. If I want to meet people, I go where they go. If I want to avoid people, I go where they do not go. And so on. This does not carry over to the slavery case. There, the addition "when others do not take me as a slave" does not describe how I attempt to achieve *P* ("avoid working") but rather how I can avoid a practical contradiction. Of course, if we want to test for a practical contradiction, we should not add such information.[37] The general idea would be as follows:

> Include in a maxim information about the conduct of other agents whenever this informs us about how the agent attempts to achieve *P* rather than avoid a practical contradiction.

34 See Allison, *Kant's Groundwork for the Metaphysics of Morals*, 196–203; Sensen, "Universalizing as a Moral Demand," 171; and Nyholm, "Kant's Universal Law Formula Revisited," 290.

35 Glasgow suggests we may remove the detail of Sunday morning because temporal locations may generally be neglected ("Expanding the Limits of Universalization," 41–44).

36 Herman, *The Practice of Moral Judgment*, 139.

37 On this theme, see Sneddon, "A New Kantian Response to Maxim-Fiddling."

Once again, I think we should not be satisfied. In social dilemmas such as the fishers' case, it may actually be relevant to add information of the sort "I will avoid doing my part in cleaning the lake *but only when enough others do their part.*" For this information actually describes how I attempt to achieve *P*. I can enjoy my life exactly because others cooperate and I free ride on their efforts. Yet this maxim constitutes a false positive. If people defect but only when enough others do their part, the lake pollution is resolved—namely, by others—and I can still keep my job and enjoy my life. Yet I make an unfair exception of myself, and my maxim should not pass the test.

3. A NEW SOLUTION

At this point, one may think that Kantians should give up. We have seen various proposals to separate innocent coordination from unfair free riding, but none fully satisfy. Can we do better? In the following, I present a novel solution. I adopt the overall idea (entertained by McCarty, Kagan, Cholbi, and others) that maxims can be conditional, but I invoke only a very specific type of conditionalization. Namely, I am interested in maxims that are conditional *on other people's preferences*. As we saw, Pogge also refers to other people's preferences, but I appeal to them only in an indirect way. That is, what matters on my account is whether or not *the agent* cares about the preferences of others (whatever they actually may be).

In coordination cases, people have different preferences (some want to play tennis now, others want to play later) and let their conduct depend on the preferences that others have (I want to play whenever enough others do not want to). In free riding cases, in contrast, the same does not apply: people do not let their conduct depend on what others prefer, and that, I suggest, is what is morally problematic. Contrast:

a. "To avoid waiting, I will play tennis on Sunday morning unless too many others also prefer to play at that time."
b. "To enjoy my life, I will not do my part in cleaning the lake regardless of whether other people also prefer not to clean."

Maxim a is conditional on other people's preferences, while maxim b is not. The former passes the Kantian test. If people restrict their tennis playing, no problems will ensue. People go to the club when (enough) others prefer not to go. If too many others prefer to go, they will not go. In such a case, the courts will likely be available, and then it will still be possible for me to avoid waiting time (and no practical contradiction arises). Maxim b, in contrast, fails to pass the Kantian test. If everyone refuses to do their part regardless of what others

prefer, the lake will likely be polluted. For in such a case, people will defect and refuse to clean the lake *even when* too many others prefer to defect too. Defection will be all over the place, the lake will be polluted, and we will all lose our jobs. Practical contradiction.[38]

What does it take exactly to make your conduct conditional on other people's preferences? There are various ways to unpack this. One way is to take people's preferences into account in your explicit practical reasoning. You may think to yourself, "They do not want to play now, so in that case I will go." Yet making your conduct conditional may also proceed less explicitly. One intuitive account is counterfactual. Your conduct is conditional on other people's preferences when you would act differently in counterfactual situations where others have different preferences. For example, your playing tennis on Sunday morning is conditional on other people's preferences when you would not play at that time if too many others would also prefer to play at that time (which they actually do not). Additionally, making conduct conditional likely involves certain dispositions on behalf of the agent: paying attention to what other people actually prefer and—when this is relevant—even actively inquiring into this.

The proposed account is inspired by Kleingeld's account of Kant's other formula: the formula of humanity. Kleingeld suggests reading this formula in an *agent-focused* way, i.e., focused on the agent's mindset.[39] To illustrate this, she describes a case where a dictator subjects people to dangerous medical experiments, and one of them "happens genuinely to consent to the treatment—say, a radical act-utilitarian who is convinced of the experiment's overwhelming benefits for large numbers of humans in the long run and who believes that these benefits vastly outweigh his own agony."[40] As Kleingeld argues, the act-utilitarian's consent is not enough to permit the dictator's experiments. More generally, agent *A* avoids using other person *B* as a mere means not simply if *B* gives (genuine) consent to be used by *A* but if *A cares about that* and makes her use of *B* conditional on *B*'s consent. In this case, the dictator does not care a bit if anyone gives consent and would still have done the experiments without it, and in this way the dictator acts wrongly.[41]

38 Regarding a, the universalization step reads, "All who pursue *P*, do *A* but not if too many others prefer to do *A* too" while regarding b, we still have, "All who pursue *P*, do *A*, i.e., even when too many others prefer to do *A* as well."

39 Kleingeld, "How to Use Someone 'Merely as a Means,'" 404.

40 Kleingeld, "How to Use Someone 'Merely as a Means,'" 393.

41 According to Kant, these "ways of representing the principle of morality are at bottom only so many formulae of the very same law" (*Groundwork of the Metaphysics of Morals*, 4:436). And as an anonymous reviewer has pointed out, my proposal "renders the formula of universal law much closer to the formula of humanity than other proposals do: by acting

Making one's conduct conditional should not be conflated with the more familiar notion of a *conditional cooperator*.[42] The latter acts on something such as "I will cooperate (clean the lake) on the condition that others do so too." This agent is well intended and willing to do her part as long as she has assurance that others will join her. The kind of agent I am talking about is well intended too but does not make her conduct conditional on the actions of others (or on what she expects others to do); her conduct is conditional on their preferences. My tennis player cares not just about when others actually or likely play tennis but about when they prefer to do so. (More on this difference in due course.)

The question arises whether we can also coordinate in an *unconditional* way, or free ride only *conditionally*. Consider:

c. "To avoid waiting, I will play tennis on Sunday morning regardless of what other people prefer."
d. "To enjoy my life, I will avoid doing my part in cleaning the lake but only if enough others actually prefer to clean it."

This time, d is conditional on other people's preferences, while c is not. And as before, the conditional maxim passes the test, but the unconditional one does not. Let us consider them in turn. Acting on c means that you want to play *only* on Sunday morning and do not care if others want that too. Hence, c can also be read as "I will go to the club on Sunday morning even when others also want to play at that moment and it is super crowded." This yields a practical contradiction. If everyone goes to the club on Sunday morning regardless of what others prefer, the courts will be packed, and I will not be able to avoid waiting time.

Is this plausible? That is, is acting on c indeed morally problematic? I think it is. The core moral wrong here is not a failure of reciprocity.[43] It is not just that people do something for the unconditional tennis player (not play on Sunday) and that the latter fails to return the favor. The wrong is also not one of taking unfair advantage of others. It does not sound right to say that the unconditional tennis player exploits others. Neither is the mistake one of stubbornness. The unconditional tennis player may be stubborn and unwilling to change her

---

only on maxims that respect the preferences of others, we take their ends into account." Indeed, if we benefit from and rely on the cooperation of others, then in a way we are using them to realize collective success for us, and we should make our conduct conditional on their consent (according to Kleingeld) or preferences (according to my account). This is not the exact same, yet the parallel is interesting.

42 The notion of conditional cooperator comes from the social contract tradition, e.g., Gauthier, *Morals by Agreement*.

43 As examined by, e.g., Brown, "Reciprocity Without Compliance."

schedule, but she may also just be indifferent (i.e., care neither about her own schedule nor about those of others).

Is the core mistake a failure of cooperativeness?[44] It is true that the unconditional tennis player, as well as other agents who act on similar unconditional maxims, can rightly be characterized as uncooperative. They face all the social dilemmas we talked about and yet will never be able to solve them if they—and all others—act on similar maxims. Yet what about an unconditional slave holder (i.e., who acts on "To avoid working, I will take a slave regardless of what other people prefer")? To only say that this agent is uncooperative does not get to the bottom of what goes wrong.

Of course, we are inclined to say that where the unconditional tennis player goes wrong is in making an exception of herself. But the same—which is the whole problem from the outset—applies to the conditional tennis player. The latter too wants to play tennis when others do not and expects others to act differently. But then how should we describe what all and only unconditional agents do wrong? I think it is just this: *they are indifferent towards others*. Specifically, they fail to care enough about what other people prefer.[45]

Finally: maxim *d*. Acting on *d* means that you intend to defect but only when enough others do not mind paying the cooperation costs. Imagine (somewhat unrealistically) that all your fellow citizens actually like to clean the lake. Imagine that they hold a competition to see who can collect the most plastic, and it is actually an honour for them to do this. You let them and do not step in. Such a maxim would pass the test. If people defect, but only when enough others prefer to cooperate, the lake will be clean, and I can benefit from their cooperation. In such a case, it may not be clear that acting on *d* should indeed be morally fine. One might think that I am still making an exception of myself in such a case. But it is important to see that I am not making *more* of an exception of myself than the tennis player who acts on maxim *a* is making an exception of herself. We both expect that others will act differently, but innocently so, as we make our conduct conditional on what other people prefer.

44 Cholbi writes regarding the maxim "To improve my backhand, I will play tennis with Katrina on the public courts every Wednesday at 4 PM" that "in insisting that she play with Katrina at 4 PM, etc., our tennis player is being uncooperative, demanding that she be able to pursue her own ends in the way she desires, heedless of the ends that other rational agents have and the ways they desire to pursue them" (*Understanding Kant's Ethics*, 154). Timmermann makes a similar point in terms of dining with friends (*Groundwork of the Metaphysics of Morals*, 158).

45 Note that it can be permitted to ignore preferences of others that do not affect you in any way. E.g., you may well be indifferent to someone's preference to watch tennis over some documentary. This becomes problematic only if you want to watch something *together*.

It seems what drives our concern here is that in the actual world, people virtually never act on d. There are hardly any cases where others do not mind paying the cooperation costs (after all, that is why they are "costs"), and so there are hardly any cases where people defect but only in such special conditions. If people defect, they defect less conditionally—that is, they act on b rather than d—and that is then what is morally problematic.

Let us compare a variant of the case where enough of the fishers prefer to clean up the lake, though this time they do not enjoy the activity of cleaning but prefer to clean because they believe they have a duty to contribute to the public good that they enjoy. Imagine that one fisher refuses to contribute, and he defends himself by saying that he does make his conduct conditional on the preferences of the others (namely, to clean). There seems to be something wrong with this fisher, even when he makes his conduct conditional. How can the proposed view account for this? In response, let me note that people may prefer to cooperate and do their duty, but in addition to that, they may still *also* prefer not to pay the costs of doing so.[46] After all, cooperation is still costly for them. In this case, the fishers have to sacrifice their weekend. The proposal of this paper is that in social dilemmas, we should make our conduct conditional upon these latter preferences, and we should cooperate if others also prefer to avoid the cooperation costs.

Let me consider a further problem case in some more detail. A selfish husband exploits his wife's self-sacrificing devotion to their children but would devote more time to childcare and housework if his wife was less self-sacrificing. If he acts on "To enjoy my life, I will avoid doing my part at home but only if my wife prefers to do it on her own," we still think he is acting wrongly.

In light of such cases, it is important to highlight that people's preferences can be adapted to the circumstances in which they find themselves. To cope with injustices, people might no longer prefer to be free or to have more time for themselves.[47] Genuinely caring about people's preferences, then, involves not simply making one's conduct conditional on people's adaptive preferences but inquiring into what they *really* prefer. Presumably, the wife does not really prefer to do all the childcare and housework alone, and the husband still acts wrongly if he acts on "To enjoy my life, I will not do my part at home regardless of what my wife really prefers."

What do people really prefer?[48] In this case, we may check what the wife preferred *before* she got oppressed. Alternatively, if she grew up in oppressive

46 Trifan, "What Makes Free Riding Wrongful?" 171–72.

47 Nussbaum, "Adaptive Preferences and Women's Options."

48 See the debate on desire satisfactionism, i.e., the view that one's well-being consists of the satisfaction of one's desires, especially those that are "laundered" in some relevant way (e.g., Goodin, "Laundering Preferences").

circumstances and never had any other preferences, we could also ask, "What would she prefer if she were not oppressed?" Arguably, in that counterfactual situation, she would prefer to share the responsibilities at home. The husband would then need to make his conduct conditional on this counterfactual preference. Such counterfactuals can be instructive, though they are not without problems. For example, if the wife were not oppressed, she might not be married in the first place and not have *any* preference about the household she does not have, and then the husband would not be able to make his conduct conditional on that.[49]

Fortunately, we may not have to specify an exact account of people's real preferences. What matters, from a Kantian perspective, is that agents *themselves* make an effort to figure this out. They do not avoid making an exception of themselves if they simply refer to what others actually happen to prefer. They should go further and check if that is what the others really prefer. Note that this duty of inquiry can be more or less demanding. In the tennis case, one could just quickly check when others prefer to play. In the household case, the husband could start doing his share of the work and after some months, ask if that is what his partner prefers.

Why not make one's conduct conditional on other people's *conduct* rather than on their preferences? Consider: "To avoid working, I will take a slave but only if others do not."[50] In the actual world, many people do not enslave others. One should not pass the Kantian test if one merely makes one's conduct conditional on that fact. Instead, the proposal is that we should make our conduct conditional on other people's preferences. Given that people do not prefer to be enslaved, one can only enslave them in an unconditional way ("To avoid working, I will enslave others regardless of whether they prefer to be enslaved") and thus fail the Kantian test.[51]

49 We also cannot check what the wife would prefer if she were not wronged: this paper's aim is to offer an account of how the husband is acting wrongly—he makes an exception of himself and does not care about what his wife prefers—and we should therefore not import a separate account of how the wife is wronged.

50 This is an example of *conditional defection*: "I will defect but only if enough others cooperate." Another example: refusing a vaccine for some infectious disease but only when enough people already got vaccinated to secure herd immunity. See Giubilini, Douglas, and Savulescu, "The Moral Obligation to Be Vaccinated," 553. In such a case, we would still think that vaccine refusers can be unfair, and the Kantian test should be able to handle such cases.

51 Thus far, we have said that people can fail to make their conduct conditional on whether other people prefer to act similarly (e.g., play tennis on Sunday morning) or prefer not to pay the cooperation costs (e.g., refrain from cleaning the lake). But it seems that one may also fail to make one's conduct conditional on whether other people prefer to be *treated* similarly (here, not to be enslaved) or on whether they prefer that *others* act in some way (here, not to enslave others).

Taken together, according to this approach, coordination is permitted when the agent cares about other people's preferences, while free riding is not permitted when the agent is indifferent about this. In principle, as discussed, there *can* be coordination cases that are not so innocent (namely when an agent intends to act unconditionally), as well as free-riding cases that are not so problematic (namely, when an agent does intend to act conditionally). When taking a closer look at those rare cases, though, that is probably exactly what we should conclude.

Note, finally, that this is not to imply that *only* conditional maxims pass the test. There are certain things I can do regardless of other people's preferences that do not run into a practical contradiction. For one thing, insofar as the Kantian test is concerned, it is fine to cooperate in social dilemmas unconditionally and, for example, do one's part in cleaning the lake regardless of what other people prefer. My preference for collective success will not be frustrated in a world where everyone cooperates unrestrictedly. Next, I will consider one further instance of this type.

4. COMPETITION

Is it permitted to lead the life of a scholar? Well, is it a case of unfair free riding or innocent coordination? According to Pogge, it depends:

> If enough others are enjoying physical labor, then the maxim "to lead the life of a scholar" would seem unobjectionable. If, on the other hand, the scholarly life is what most others would also be inclined to favor, then my success in leading such a life without physical work is necessarily parasitic upon the (morally motivated or coerced) sacrifice by others producing the necessities for human existence.[52]

Here, Pogge suggests that whether it is permitted to make a living as a scholar depends on what others prefer. If they would want this too, while in fact they make food for you, it seems you are unfairly free riding on them. My account, in contrast, is not about what others in fact prefer but about whether you, the agent, make your conduct conditional on that.

The worry now is that most of us fail to do this. Instead, we act on "To make a living, I will work on abstract philosophical problems regardless of what other people prefer" or "I will let others produce food even when they prefer to be philosophers too." These will not pass the test. If everyone were to do philosophy unconditionally, no one would produce any food, and I would not be able

52 Pogge, "The Categorical Imperative," 190.

to stay alive (an implicit purpose of mine). Should acting on these maxims, then, be morally problematic?

Parfit also considers these cases: "We can imagine fanatical, unconditional maxims whose universal acceptance would lead us all to become childless underemployed Icelandic dentists who starved themselves to death. . . . Kant's formulas mistakenly condemn our acting on these maxims."[53] In response, Parfit suggests that acting on such maxims should be fine as long as enough people do not actually act on them. For example, Parfit permits Kant to act on "To devote my life to philosophy, I will not have children regardless of whether others do have them" because enough others do not act on this maxim.[54]

As I see it, we may well want to resist Parfit's position here and maintain that acting on certain unconditional maxims just is morally problematic. After all, you are relying on others to produce food for you (or have children, etc.) even when they would rather work on abstract problems too. What matters is not (only) that people's contributions to society in fact complement one another. What also matters is what everyone prefers to do and whether we are sufficiently sensitive to that. Kant acts wrongly (in the case imagined) because he does not care one bit about what everyone else wants.[55]

In the actual world, to be sure, not everyone wants to do philosophy, yet more people want this than there are available jobs. In such a situation, we do not seem to be able to make our conduct conditional on *everyone's* preferences, though it would be implausible to think that we thereby all act wrongly.

I think a promising alternative analysis of such cases is the following. In addition to coordination and free-riding cases, there is a third type of case: *competition cases*. Consider: "To get rich, I will finish first and win the prize." This maxim yields a practical contradiction. It is not possible to get rich if everyone

53 Parfit, *On What Matters*, 1:311.

54 Note that Parfit discusses maxims that are (or are not) conditional on what others *do*, not on what they prefer (as I have it). According to Parfit, doing philosophy unconditionally is permitted by his LN3: "We act wrongly unless we are doing something that we could rationally will everyone to do, in similar circumstances, if they can" (*On What Matters*, 1:311). Compare Brown's *vigilance principle*: "Citizens should do actions which are such that if not enough people do them public goods will suffer and there is a reasonable risk that not enough will do so" ("Reciprocity Without Compliance," 415).

55 Shahar argues that acting on the following maxim is permitted: to make the world a better place, I will not boycott animal products but spend my energy on other causes. Shahar, *Why It's OK to Eat Meat*. In light of the account developed here, we could say that the case is inconclusive—much like "To contribute to society, I will not produce food but spend my time as a philosopher" is undetermined. The question is: Do I make my conduct conditional on what other people prefer? The maxim "To make the world a better place, I will spend my energy on such and such causes regardless of what other people prefer" may still fail to be universalizable.

finishes first. (We might all have to split the money, or we might not get anything at all if there is no clear winner.)

Competition cases have received a compelling analysis by McCarty.[56] Some act descriptions are explicitly or implicitly relative to the actions of others. For example, not everyone can be the first to make it to the finish. As McCarty points out, in such cases, we lack full control over the act ("finishing first"), and much depends on what one's competitors do. For this reason, McCarty suggests, we should move the comparative terms in a maxim from the act description to the purpose description. So I do not act on "To get the prize, I will finish first" but on "To finish first (and win the prize), I will try my best and run the hardest I can." That yields no practical contradiction. In a world where everyone runs the hardest they can, I can still finish first. Winning may not be likely, of course, but that is why it is a competition case.

Similarly, then, people might act on "To get the job, I will try my best and work the hardest I can, i.e., regardless of what other people prefer." Such a maxim would pass the Kantian test.

5. Two Perspectives

As announced, I distinguish two perspectives: how an agent wants others to behave versus how others themselves actually want to behave. Contrast:

> *Flat Share 1*: Imagine a shared flat where all three flatmates strongly prefer a certain level of cleanliness. Two of them do their share of upholding this level of cleanliness, yet the third refuses.
>
> *Flat Share 2*: Imagine a shared flat where all three flatmates strongly prefer a certain level of cleanliness. This level of cleanliness is maintained thanks to the fact that two of the flatmates enjoy exercising around the house with a duster as their preferred way of staying in shape.

According to Trifan, the third flatmate—call him Immanuel—is an unfair free rider in Flat Share 1, but not in Flat Share 2.[57] In Flat Share 2, Immanuel's flatmates cooperate (here, clean) willingly, and according to Trifan, you are unfair only when others share a "free-riding preference" with you. In Flat Share 1, Immanuel makes an exception of himself exactly because he allows himself

56 McCarty, "False Negatives of the Categorical Imperative," 186–88.

57 Trifan, "What Makes Free Riding Wrongful?" 176.

to act on a preference (namely, to avoid unnecessary efforts to keep the house clean) that others share with him but do not act on.[58]

This is very different from the Kantian conception, which states that what matters is the agent's mindset—and Immanuel's mindset might be the same in both these cases. Moreover, in both, Immanuel seems to run into a practical contradiction since his preference to live in a clean house is frustrated if others act like him (i.e., if they also do not clean). Is this a *reductio* of the whole approach?

According to the account proposed in the foregoing, we should focus on Immanuel's maxims rather than on the preferences of his flatmates. But after a closer look, his maxims are not the same. Contrast:

e. "To have time for my hobbies, I will not help clean regardless of what my flatmates prefer."
f. "To have time for my hobbies and let others enjoy theirs, I will avoid helping to clean, but only when enough of my flatmates prefer to clean."

In Flat Share 1, Immanuel likely acts on e. Doing so runs into a practical contradiction. Immanuel's preference for a clean house—that is, if he possesses it—is frustrated after universalization. If his flatmates also refuse to do their part regardless of what the others prefer, their house will be a mess. In Flat Share 2, in contrast, Immanuel likely acts on f. Doing so passes the Kantian test. If Immanuel's flatmates refuse to do their part but only when enough others clean, their house will still be clean.

Are the two perspectives—the Kantian perspective on the agent versus Trifan's perspective on other agents—the same then? No. According to the Kantian account, Immanuel may still be unfair in Flat Share 2. After all, even when his flatmates actually prefer to clean, he might not care about that at all and still act on e rather than on f, i.e., refuse to clean regardless of what his flatmates prefer. And if he does, he will still run into a practical contradiction and be unfair. I think that is just what we should say, and this counts in favor of

58 Trifan builds upon Klosko, "Presumptive Benefit, Fairness, and Political Obligation," and Cullity, "Moral Free Riding." On Cullity's account, basically, Immanuel should pay his part as long as the demand to do so is fairly generalizable, i.e., it is reasonable to ask people to pay in all similar cases. In contrast to the Kantian analysis, Immanuel's own mindset—whether he cares about collective success and seeks to benefit from it—is not relevant in Cullity's account. Inspired by Klosko, Trifan agrees with this, specifically when it comes to required goods (e.g., a clean environment) as opposed to optional goods (e.g., a high level of cleanliness in the flat). My Kantian account does not make use of any distinction between required and optional goods, and it corresponds more to Nozick's subjective approach. According to Nozick, you have no obligation to contribute if you consider the costs of doing so higher than the benefits you will receive (*Anarchy, State, and Utopia*, 93–94).

the Kantian perspective. To be sure, though, this raises subsequent issues that we cannot address here.[59] This paper's very ambition instead was to save the Kantian test from a widespread objection.

Thus, can Kantians distinguish unfair free riding from innocent coordination? The proposal is this: in innocent coordination cases, people let their conduct depend on what others prefer (e.g., playing tennis whenever enough others do not want to play); in unfair free-riding cases, in contrast, people do not make their conduct conditional in this way but rather refuse to do their part regardless of what others prefer. The latter—but not the former—fails the Kantian test.[60]

Vrije Universiteit Amsterdam
j.j.w.wieland@vu.nl

REFERENCES

Albertzart, Maike. "A Kantian Solution to the Problem of Imperceptible Differences." *European Journal of Philosophy* 27, no. 4 (2019): 837–51.

Allison, Henry E. *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford University Press, 2011.

Bicchieri, Cristina. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms.* Oxford University Press, 2017.

Braham, Matthew, and Martin van Hees. "An Anatomy of Moral Responsibility." *Mind* 121, no. 483 (2012): 601–34.

Brown, Brookes. "Reciprocity Without Compliance." *Philosophy and Public Affairs* 48, no. 4 (2020): 382–421.

Cholbi, Michael. *Understanding Kant's Ethics*. Cambridge University Press, 2016.

Cullity, Garrett. "Moral Free Riding." *Philosophy and Public Affairs* 24, no. 1 (1995): 3–34.

Galvin, Richard. "Maxims and Practical Contradictions." *History of Philosophical Quarterly* 28, no. 4 (2011): 407–19.

59 For example, see Parfit's Mixed Maxims Objection in *On What Matters*, 1:293, discussed in Hoesch and Sticker, "Parfit's Mixed Maxim Objection Against the Formula of Universal Law Reconsidered." The issue is whether it is plausible to assume that acting on maxim e is wrongful even though in Flat Share 2, the action itself may not be considered problematic.

60 This paper benefited from discussions with many colleagues. In particular, I would like to thank Justin Bernstein, Fleur Jongepier, Ben Ferguson, Emma Turmaine, Martin van Hees, Julia Duetz, Johan Olsthoorn, Chris Kutz, and two reviewers of this journal.

Gauthier, David. *Morals by Agreement*. Oxford University Press, 1986.
Giubilini, Alberto, Thomas Douglas, and Julian Savulescu. "The Moral Obligation to Be Vaccinated: Utilitarianism, Contractualism, and Collective Easy Rescue." *Medicine, Health Care and Philosophy* 21, no. 4 (2018): 547–60.
Glasgow, Joshua. "Expanding the Limits of Universalization: Kant's Duties and Kantian Moral Deliberation." *Canadian Journal of Philosophy* 33, no. 1 (2003): 23–47.
Goodin, Robert E. "Laundering Preferences." In *Foundations of Social Choice Theory*, edited by Jon Elster and Aanund Hylland. Cambridge University Press, 1986.
Gunnemyr, Mattias, and Caroline T. Touborg. "Reasons for Action: Making a Difference to the Security of Outcomes." *Philosophical Studies* 180, no. 1 (2023): 333–62.
Hedden, Brian. "Consequentialism and Collective Action." *Ethics* 130, no. 4 (2020): 530–54.
Herman, Barbara. *The Practice of Moral Judgment*. Harvard University Press, 1993.
Hoesch, Matthias, and Martin Sticker. "Parfit's Mixed Maxim Objection Against the Formula of Universal Law Reconsidered." *Journal of Value Inquiry* 58, no. 1 (2024): 13–32.
Kagan, Shelly. "Do I Make a Difference?" *Philosophy and Public Affairs* 39, no. 2 (2011): 105–41.
———. "Kantianism for Consequentialists." In *Groundwork for the Metaphysics of Morals*, edited by Allen W. Wood. Yale University Press, 2002.
Kant, Immanuel. *Groundwork of the Metaphysics of Morals*, translated by Mary Gregor and Jens Timmermann. Cambridge University Press, 1998.
Kerstein, Samuel J. *Kant's Search for the Supreme Principle of Morality*. Cambridge University Press, 2002.
Kleingeld, Pauline. "How to Use Someone 'Merely as a Means.'" *Kantian Review* 25, no. 3 (2020): 389–414.
Klosko, George. "Presumptive Benefit, Fairness, and Political Obligation." *Philosophy and Public Affairs* 16, no. 3 (1987): 241–59.
Korsgaard, Christine M. "Kant's Formula of Universal Law." *Pacific Philosophical Quarterly* 66, nos. 1–2 (1985): 24–47.
Lomasky, Loren E., and Geoffrey Brennan. "Is There a Duty to Vote?" *Social Philosophy and Policy* 17, no. 1 (2000): 62–86.
McCarty, Richard. "False Negatives of the Categorical Imperative." *Mind* 124, no. 493 (2015): 177–200.
Nefsky, Julia. "Collective Harm and the Inefficacy Problem." *Philosophy Compass* 14, no. 4 (2019): 1–17.

———. "How You Can Help, Without Making a Difference." *Philosophical Studies* 174, no. 11 (2017): 2743–67.

Nozick, Robert. *Anarchy, State, and Utopia*. Basic Books, 1974.

Nussbaum, Martha. "Adaptive Preferences and Women's Options." *Economics and Philosophy* 17 (2001): 67–88.

Nyholm, Sven. "Kant's Universal Law Formula Revisited." *Metaphilosophy* 46, no. 2 (2015): 280–99.

O'Neill, Onora. *Acting on Principle*. Cambridge University Press, 1975. Second edition, 2013.

Parfit, Derek. *On What Matters*. Vol. 1. Oxford University Press, 2011.

———. *Reasons and Persons*. Oxford University Press, 1984.

Pettit, Philip. "Free Riding and Foul Dealing." *Journal of Philosophy* 83, no. 7 (1986): 361–79.

Pogge, Thomas W. "The Categorical Imperative." In *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, edited by Paul Guyer. Rowman and Littlefield, 1998.

Scanlon, Thomas M. *What We Owe to Each Other*. Harvard University Press, 1998.

Sensen, Oliver. "Universalizing as a Moral Demand." *Estudos Kantianos* 2, no. 1 (2014): 169–84.

Shahar, Dan C. *Why It's OK to Eat Meat*. Routledge, 2022.

Sneddon, Andrew. "A New Kantian Response to Maxim-Fiddling." *Kantian Review* 16, no. 1 (2011): 67–88.

Timmermann, Jens. *Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge University Press, 2007.

Trifan, Isabella. "What Makes Free Riding Wrongful? The Shared Preference View of Fair Play." *Journal of Political Philosophy* 28, no. 2 (2020): 158–80.

Wieland, Jan Willem. "Cooperation, Kantian-Style." *Inquiry*, ahead of print, February 3, 2024. https://doi.org/10.1080/0020174X.2024.2312215.

Wood, Allen W. *Kant's Ethical Thought*. Cambridge University Press, 1999.

Journal of Ethics and Social Philosophy
VOL. 29, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.3436

NAVIGATING NONIDENTITY

SCANLONIAN CONTRACTUALISM AND TYPES OF PERSONS

D. Valeska Martin

SCANLON'S moral contractualism presents a compelling framework for understanding our moral obligations towards one another.[1] According to this view, our actions should be guided by principles that are justified to others, in the sense that they cannot reasonably reject these principles. However, an open question remains as to whether Scanlonian contractualism can adequately address our obligations to future generations, specifically in light of the nonidentity problem.[2]

Many actions and policies that have long-term negative consequences for the well-being of future generations are to be understood as nonidentity cases in which the existence of the affected future persons is contingent upon our actions. In these cases, we are not making any future individual worse-off than she would otherwise be; rather, our actions fail to bring into existence other future individuals who would be better-off. The person-affecting structure of Scanlonian contractualism is tested to its limits here. While the principles allowing such actions may intuitively appear morally unjustifiable, it remains unclear on what grounds a future person, whose existence depends on the adoption of a principle and who would lead a worthwhile life, would have reason to reject that principle.

Scanlon has suggested that moral contractualism should respond to the nonidentity problem by showing that future persons have a reasonable objection not as the particular individuals but in terms of the "reasons that any individual would have in virtue of being in a certain position."[3] A more detailed

1 Scanlon, "Contractualism and Utilitarianism" and *What We Owe to Each Other*.

2 The nonidentity problem has been brought to wider attention by the work of Derek Parfit. See especially Parfit, *Reasons and Persons*, ch. 16. To my awareness, the earliest discussion of the problem can be found in Schwartz, "Obligations to Posterity."

3 Scanlon, "Responses to Forst, Mantel, Nagel, Olsaretti, Parfit, and Stemplowska," 143.

account of this idea is given by Rahul Kumar.[4] He suggests that the nonidentity problem can be avoided because the justifiability of moral principles depends on whether they can be rejected from the perspective of a relevant standpoint or type of person, characterized in general terms. On this view, a future person is wronged when she is treated according to a principle that is rejectable from the perspective of her type. I call this the *types-of-persons approach*.

This paper puts the types-of-persons approach under closer scrutiny. In section 1, I will briefly present Scanlonian contractualism and the nonidentity problem, followed by the idea of types of persons in section 2. I suggest two alternative interpretations of the types-of-persons approach that differ in how the reasons for objection in nonidentity cases are characterized: we could assume that the reasons on grounds of which the principles are unjustified are (a) reasons of types of persons or (b) type-based reasons of token persons. I spell out these versions in more detail in sections 3 and 4, respectively, arguing that both are incompatible with Scanlonian contractualism. I thus conclude that the types-of-persons approach does not provide a satisfactory solution to the nonidentity problem for Scanlonian contractualists.

1. Scanlonian Contractualism and the Nonidentity Problem

The basic idea of Scanlonian contractualism is that as agents, we stand in a relation of mutual recognition in which we owe each other justification. On this view "an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement."[5] When we deliberate about what we ought to do, we consider what reasons the potentially affected individuals have to object to the principles allowing and disallowing the action in question. According to Scanlon, reasons to reject such principles are plural. They can be based on a person's well-being, interests, or fairness, for example. When we act against a person's reasonable rejection, we wrong her because the action is unjustified to her.[6]

4 Kumar, "Wronging Future People," "Risking Future Generations," and "Who Can Be Wronged?" This view has gained further support in Finneron-Burns, "Contractualism and the Non-Identity Problem" and *What We Owe to Future People*.

5 Scanlon, *What We Owe to Each Other*, 5.

6 This is the standard reading of Scanlonian contractualism as a relational theory, i.e., a theory concerned with directed obligations: we owe it to a person to act in a certain way, and we wrong her if we fail to do so. Katz has argued that to avoid the nonidentity problem, Scanlonian contractualism could also be interpreted in an impersonal "moral wrong" reading ("Contractualism, Person-Affecting Wrongness and the Non-Identity Problem"). However,

Two aspects are of particular importance in the context of this paper. First, to consider what the potentially affected persons could object to the adoption of the principles allowing or disallowing an action, Scanlon assumes that we ought to think about the various *standpoints* of the affected individuals and the *generic reasons* that can be attributed to them: we shall take into account what kinds of reasons people in certain situations plausibly have due to the position they are in.[7]

Second, it is only the *personal* reasons of *individuals* that count as reasons for the rejection of principles.[8] This idea can be spelled out in terms of the *individualist restriction* and the *impersonalist restriction*. Broadly speaking, the individualist restriction excludes objections based on interpersonal aggregation or the claims of groups, and the impersonalist restriction excludes objections based on impersonal considerations, such as the intrinsic value of a natural landscape.[9] I come back to this in more detail in section 4.

The nonidentity problem challenges Scanlonian contractualism's ability to say that we wrong future persons in nonidentity cases. Consider Parfit's case:

> *Depletion*: As a community, we must choose whether to deplete or conserve certain kinds of resources. If we choose Depletion, the quality of life over the next two centuries would be slightly higher than it would have been if we had chosen Conservation. But it would later, for many centuries, be much lower than it would have been if we had chosen Conservation.[10]

Because large-scale policies like Depletion and Conservation affect the lives of so many people, they have a profound effect on who and how many people will be born in the future. We can therefore assume that over the course of a few generations, none of the future persons who would exist if we chose Depletion would also exist if we chose Conservation, and vice versa. Furthermore,

---

an impersonal interpretation of Scanlonian contractualism seems to amount to an entirely different theory, and importantly, it can then claim no longer that future persons are wronged in nonidentity cases (but only that we act impersonally wrong).

7 Scanlon, *What We Owe to Each Other*, 202–4.

8 See especially Scanlon, *What We Owe To Each Other*, 229, 241.

9 This terminology is from Parfit. For the individualist restriction, see Parfit, *On What Matters*, 2:193; for the impersonalist restriction, see Parfit, *On What Matters*, 2:214. Scanlon adopts the terminology, albeit not consistently; see Scanlon, "Replies," 429, and "Contractualism and Justification."

10 Parfit, *Reasons and Persons*, 361–62.

it is stipulated that the lives of future people in both alternative outcomes, the depleted and the conserved world, would still be worth living overall.[11]

While it seems intuitive that we ought not to choose Depletion, it is not clear whether Scanlonian contractualism yields this verdict. Generally, it seems open to take into account the reasons of future persons just as those of the currently alive.[12] However, the future individuals in the depleted world are not made worse-off by the principle allowing Depletion than they would have been under the alternative principle prohibiting Depletion (and neither would their interests be fulfilled under the alternative principles, nor would they be treated fairly, or whatever you think is the appropriate basis for objection). These particular persons would not exist at all if we adopted the principles prohibiting Depletion. So the central question for Scanlonian contractualism is: Can future persons reasonably reject the principles allowing an action even if their particular existence depends on the adoption of these principles?

2. The Types-of-Persons Approach

Scanlon has not given an explicit account of how to respond to the nonidentity problem. When mentioning it in *What We Owe To Each Other*, he suggests only that it touches substantive questions to be dealt with "within the morality of right and wrong."[13] More recently, Scanlon has suggested that a response to the nonidentity problem requires viewing the reasons for objecting to a principle not as reasons that are raised by particular persons but rather as reasons that could be raised by any individual due to her position. He clarifies:

> The objections that are relevant in the process of contractualist justification are not objections of particular individuals. Rather, they are reasons that any individual would have in virtue of being in a certain position, namely the position of being affected in a certain way by a principle. . . . The position in question is that of living under conditions that are very bad in some way, due to policies that this principle would allow (for example, bad because of pollution, increased temperatures, or high sea levels) but who would not have existed if some alternative policy had

11 Parfit, *Reasons and Persons*, 362.

12 Scanlon assumes that in principle, those who can be wronged are "all those who do, have, or will actually exist" (*What We Owe to Each Other*, 187). This is usually accepted. See Kumar, "Wronging Future People," 255–56; Hurley and Weinberg, "Whose Problem Is Non-Identity?" 724; Gibb, "Relational Contractualism and Future Persons," 138; Finneron-Burns, "Contractualism and the Non-Identity Problem," 1152–53, and *What We Owe to Future People*, 56–63.

13 Scanlon, *What We Owe to Each Other*, 186.

> been followed. It is not obvious to me that people in this position do not have such an objection, although I do not have a worked-out view of the matter.[14]

What is needed, then, is a "worked-out view" of that objection based on the reasons any individual in such a position would have. Such an account can be found in the work of Rahul Kumar.[15] His central claim is that Scanlonian contractualism avoids the nonidentity problem because it is concerned with whether the principles allowing an action are reasonably rejectable from relevant standpoints or types of persons, characterized in more general terms. Like Scanlon in the quote above, Kumar thus presses the point that Scanlonian contractualism is concerned with what can be objected from standpoints rather than from particular individuals. He argues:

> The general point here, which is key to understanding Contractualism's claim to immunity to the non-identity problem, is that for there to be a fact of the matter concerning what it is one owes another to whom one stands in a particular type of relationship, it is enough that the other to whom consideration is owed be characterizable in normative terms, by a relevant type description. There need be no fact of the matter concerning the particular token identity of that individual, as her token identity is irrelevant for fixing what it is she is owed as a matter of respect for her value as a person.[16]

Applying this view to Depletion, we can assume that the principles allowing Depletion are reasonably rejectable from the perspective of the type of a future person. Importantly, if we do choose Depletion, then any particular future individual in the depleted world can claim to have been wronged because she "instantiates" the relevant type.[17] This claim that, very roughly, the nonidentity problem is avoided by reference to types of persons, is what I call the types-of-persons approach.

14 Scanlon, "Responses to Forst, Mantel, Nagel, Olsaretti, Parfit, and Stemplowska," 143.

15 Kumar, "Wronging Future People," "Risking Future Generations," and "Who Can Be Wronged?" His view has received further support by Elizabeth Finneron-Burns in "Contractualism and the Non-Identity Problem" and *What We Owe to Future People*, ch. 2.

16 Kumar, "Wronging Future People," 261–62.

17 Kumar, "Risking Future Generations," 252. Note that this is an important element that distinguishes Kumar's view from Caspar Hare's. Hare argues that we can have moral obligations to *de dicto* people (for example, an obligation to "my future child" in the *de dicto* rather than the *de re* sense), but he does not assume that the *de re* future child is wronged by a violation of a *de dicto* obligation. See Hare, "Voices from Another World."

While I am sympathetic to this view, it is not clear how it is supposed to work in detail. On closer inspection, it seems to me that two alternative interpretations of the types-of-person approach are available, and they have been conflated so far. To see this, it is helpful to spell out the objections to a principle in terms of reasons. Schematically, we may say that in a given context *C*, proposition *P* is a reason for subject *S* to object to the principle in question. For example, suppose that the adoption of a certain principle would harm me, and this is a reason for me to object to that principle. Then the fact that I would be harmed is the proposition, and I am the subject. Based on this understanding of reasons for objection, we can distinguish two different versions of the types-of-persons approach.

The first option (a) is to claim that types are the subject for which some consideration is a reason to object to a principle. So in the understanding of a reason introduced above, *S* is a type of person rather than a token person. With regard to Depletion, for example, we might say that the type *future person* could object to the principle allowing Depletion because it makes the type worse-off, violates its interests, is unfair, or is in some other form disrespectful towards that type. On this view, moral deliberation is essentially concerned with types of persons and their objections. The alternative option (b) is to say that the type is only part of the proposition *P*, while the subject *S* is still a token person. On this view, we can maintain that it is particular individuals and their reasons that ultimately matter for the justifiability of moral principles. An individual may nonetheless have reason to object to a principle based on how the principle affects her type more generally. For Depletion, for example, we may say that any token future person could object to the principles allowing Depletion on the grounds that it makes persons of her type worse-off, is harmful to her type, or imposes higher risks on her type. In sections 3 and 4 below, I will discuss these two alternative versions of the types-of-persons approach one by one and argue that both are, in their own ways, incompatible with central features of Scanlonian contractualism.

3. THE REASONS OF TYPES OF PERSONS

The first option to spell out a types-of-persons solution is to claim that contractualist reasoning about wrongness essentially takes place at the level of types of persons rather than of token persons. According to this view, there are reasons to reject Depletion attributed to the standpoint or type *future person* in such generality; and this is all we need to know in order to assess whether the principle in question is unjustified. As Scanlonian contractualism allows a plurality of reasons for objection, there are in principle different options to spell out the

objection in more detail. For example, it could be argued that the type *future person* could object to the principle allowing Depletion because it significantly lowers its well-being, violates its interests, imposes higher risks on it, or treats the type unfairly. A particular future individual in Depletion would then be wronged by not being treated according to the behavior owed to her type in that type of situation. However, I believe this understanding of the types-of-persons approach conflicts with some of the basic ideas of Scanlonian contractualism.

From an exegetical perspective, note that such an understanding of the role of types of persons does not correspond to the concept of standpoints originally introduced in *What We Owe To Each Other*. The notions of standpoints and generic reasons are here primarily an answer to an epistemic problem.[18] Scanlonian contractualism requires deliberating agents to take into account a very broad range of considerations. This is why Scanlon suggests that we must rely on "commonly available information about what people have reason to want" when assessing the rejectability of a principle.[19] We therefore have to refer to generic reasons that are "reasons that we can see that people have in virtue of their situation, characterized in general terms, and such things as their aims and capabilities and the conditions in which they are placed."[20]

Elizabeth Ashford ascribes a more substantive, moralized role to standpoints, restricting reasonable rejection to the kind of objections "everyone has reason to be concerned with."[21] This rules out reasons that persons have due to very peculiar or inadequate individual preferences. But irrespective of whether standpoints serve a merely epistemic or some further moral function in Scanlonian contractualism, standpoints represent token individuals on both readings. Deliberating about standpoints or types is, on this view, only a way of taking into account the morally relevant reasons that the affected token individuals plausibly have.

Putting exegetical questions aside, however, the first version of the types-of-persons approach raises additional formal and substantial questions. To deliberate about types of persons might seem quite in line with the generality of moral principles and the way in which we commonly speak of persons when we deliberate about what is right or wrong. Kumar insightfully suggests that the contractualist system of moral principles can be seen as analogous to a legal

18 Scanlon, *What We Owe to Each Other*, 202–4. See also Gibb, "Relational Contractualism and Future Persons," 146.

19 Scanlon, *What We Owe to Each Other*, 204.

20 Scanlon, *What We Owe to Each Other*, 204.

21 Ashford, "The Demandingness of Scanlon's Contractualism," 277.

system in this respect.[22] Principles define what we can legitimately expect from each other in certain situations, and just like legal principles, moral principles refer to persons and situations in a generalized manner by specifying the types of persons and types of situations in which they hold. Kumar's example to illustrate this is that of an employer and an employee: when we think about what principles for the regulation of employment relationships are justified, we think about an employer's and an employee's interests in different situations in a generalized way. We can meaningfully discuss which conduct employer and employee owe each other by referring to types of persons and types of situations. And when a particular employee is said to have been wronged, then this claim is based on the assessment that the relevant type descriptions are applicable to her and her employer in their actual situation.[23]

However, while this might be plausible when deliberating about employers and employees, the relation of the type and the token person is not so straightforward in nonidentity cases. While the type of a future person has a reason to object to the principle in question, in nonidentity cases, the token future person herself does not have that reason. (Otherwise, I assume, we would not need a types-of-persons approach to solve the nonidentity problem in the first place.) This raises the question of whether the token person is in fact an instantiation of that type.[24] But even if it is, to rely on the type's reasons in such a case seems not very attractive for Scanlonian contractualism as a moral theory.

First, to claim that the wrongness of an action is grounded in the reasons of types of persons is at odds with the view that reasons are relative to persons. In Scanlonian contractualism, reasons are the most fundamental normative concept.[25] Scanlon characterizes a reason for action as a four-placed relation between a person, a proposition, the circumstances, and an action.[26] So what I have schematically introduced as the subject of a reason, on this view, must always be a person. And this assumption is not unique to Scanlon's view but widely shared.[27] However, when reasons are relative to particular persons, and reasons are the foundation of normativity in Scanlonian contractualism, then ultimately, the reasons of particular persons should be guiding our deliberation.

22 Kumar, "Risking Future Generations," 260.

23 Kumar, "Risking Future Generations," 261.

24 See Huseby, "Person-Affecting Moral Theory, Non-Identity and Future People," 204.

25 Scanlon, *What We Owe to Each Other*, 17.

26 Scanlon, *Being Realistic About Reasons*, 120.

27 I am not aware of any scholar who denies that a reason is always a reason for a person. See, e.g., Raz, *Engaging Reason*; Dancy, *Ethics Without Principles*; Schroeder, *Slaves of the Passions*; Skorupski, *The Domain of Reasons*; Parfit, *On What Matters*, vol. 1; Scanlon, *Being Realistic About Reasons*; Snedegar, *Contrastive Reasons*.

This is important because it shows what to do when we come to see that on closer inspection, a person herself does not actually have the reason we ascribed to her standpoint. What is morally relevant in such a case must surely be the actual person's reasons and not the reasons ascribed to an abstract type description. I assume that this is plausible in the employer example as well. If we see that a particular employer and employee in fact have reasons very different from those we ascribed to the general roles of employer and employee, it seems problematic to take the general descriptions as decisive for assessing what they owe to each other. In a legal system, perhaps all that matters is whether your case falls under the circumstances and roles defined by the law, but in morality, I believe, we should rather look at the actual individuals. It is their reasons and interests that ultimately matter for the justifiability of moral principles.

Second, even if some may disagree with the claim that morality must be concerned with the reasons of token persons, this is not an option for Scanlonian contractualism. In particular, the view that moral deliberation is concerned with types of persons rather than with token persons clashes with the idea of what we owe to each other. Contractualism's central idea is that in a relation of mutual recognition, we respect the value of another person by recognizing her as a reasoning agent and thus owe it to her to take her reasons into account.[28] If in deliberation we take into account the reasons ascribed to a type description of her, even if she herself does not have these reasons, this does not seem to amount to the kind of respect we owe to her. Thus understood, contractualism ultimately seems to be a theory of what is owed to types of persons rather than of what is owed to persons.[29]

That being said, standpoints and generic reasons still have an important epistemic (or perhaps even further moral) function in contractualist reasoning, and I do not want to imply that this feature should be dismissed. We should bear in mind, however, that in a Scanlonian framework, what is ultimately morally relevant must be the reasons of particular individuals, not those of abstract types. Therefore, the idea that Scanlonian contractualism avoids the nonidentity problem because types of persons are the subject of the reasons for rejection is not plausible. This is not the end of the types-of-persons approach, however. Even if the relevant subjects in moral deliberation should be token persons, it may still be the case that a token person can reject a principle based on how it affects her type more generally. This idea is discussed in the following section.

28 See also Ashford, "The Demandingness of Scanlon's Contractualism," 277; Gibb, "Relational Contractualism and Future Persons," 11–12.

29 See also Parfit, *On What Matters*, 2:235–36.

4. TYPE-BASED REASONS OF TOKEN PERSONS

In the previous section, I have taken the types-of-persons approach to make the claim that types of persons are the subject of the relevant reasons for objection in moral deliberation. But perhaps the types-of-persons approach is better understood as making a claim only about the proposition. On this interpretation, a defender of the types-of-persons approach might agree that it is the reasons of particular persons that ultimately matter morally but argue that particular persons have certain reasons to reject principles *due to their type*. This interpretation of the types-of-persons approach could avoid the concerns raised in the previous section and may also seem closer to Scanlon's suggestion that contractualism should look at the reasons persons have "in virtue of being in a certain position."[30]

On this view, a person can object to a principle on the ground that it negatively affects her type. Regarding Depletion, for example, any token future person could object to a principle allowing Depletion because it would leave a person of her type (the type *future person*) significantly worse-off compared to the alternative principles, because it would violate her type's interests or impose higher risks on her type. The reason to object to a principle thus is a particular person's reason, but its proposition refers to how the person's more general type is affected rather than how she herself as a token would fare under this principle and its alternatives. I call these reasons *type-based reasons* to reject principles.

Note that the type-based reasons I have in mind here are comparative, that is, they refer to how the type of person is comparatively affected by a principle relative to its alternatives. Instead, one could also argue for a noncomparative account of these reasons, for example by referring to a violation of noncomparative interests, noncomparative harm, or a sufficiency threshold of well-being. A noncomparative response to the nonidentity problem, however, is an entirely different view that deserves its own in-depth discussion.[31] I assume that the main motivation for adopting a types-of-person approach is to incorporate the kinds of comparisons that we can make with respect to the types but not

30 Scanlon, "Responses to Forst, Mantel, Nagel, Olsaretti, Parfit, and Stemplowska," 143.

31 For a defense of noncomparative objections, see Suikkanen, "Contractualism and Climate Change"; and Wallace, *The Moral Nexus*. See also Kumar's brief remarks in "Risking Future Generations," 253. The idea of noncomparative objections recurs in an older debate on noncomparative harm, most prominently defended by Shiffrin, "Wrongful Life, Procreative Responsibility, and the Significance of Harm"; Harman, "Can We Harm and Benefit in Creating?" For critical discussions, see, e.g., Hanser, "Harming and Procreating"; Boonin, *The Non-Identity Problem and the Ethics of Future People*, especially ch. 3, sec. 3; and McMahan, "Climate Change, War, and the Non-Identity Problem."

with respect to the token persons. The following discussion is thus limited to a comparative understanding of type-based reasons.

Generally, I believe the idea of type-based reasons has quite an intuitive appeal, but it has the following problem. It assumes, for example, that a person can object to a principle on the ground that it leaves persons of her type worse-off than persons of that type would otherwise be, even though she cannot object on the ground that the principle makes *her* worse off than *she* would otherwise be. This is what allows the view to solve the nonidentity problem. However, it is in tension with the contractualist requirement that objections must be the personal objections of individuals.

As introduced earlier, in Scanlonian contractualism, it is only the personal reasons of individuals that count as reasons for the rejection of principles. For this constraint, Parfit has coined the terms *individualist restriction* and *impersonalist restriction*. Following his definition, the individualist restriction requires that "in rejecting some moral principle, we must appeal to this principle's implications only for ourselves and for other single people."[32] This excludes the interpersonal aggregation of benefits or burdens or referring to the claims of groups, for instance. The impersonalist restriction claims that "in rejecting some moral principle, we cannot appeal to claims about the impersonal goodness or badness of outcomes."[33] This excludes, for example, reasons referring to the intrinsic value of nature or other impersonal considerations. As Scanlon argues, if someone wants to destroy the Grand Canyon to build a theme park, I might object to the principles allowing such a project on the ground that it would deprive me of enjoying that natural landscape. But the value of the natural monument itself is nothing that makes the action unjustified *to me*. While there might be intrinsic value in landscapes or ecosystems, this is not considered to be part of the contractualist domain of morality, the morality of what we owe to each other.[34]

The common idea behind these constraints is that the reasons must be, as Scanlon suggests, reasons that a person has "on his or her own behalf."[35] Its roots lie in the contractualist ideal of justifiability to each person, and it is intricately linked to the relational structure it takes morality to have. What we owe to each other is to take the reasons of each person into account, and to ground a contractualist wronging, these must be reasons that the person has on her own behalf, otherwise it would not be a wronging of *her* in particular. The two

32 Parfit, *On What Matters*, 2:193.

33 Parfit, *On What Matters*, 2:214.

34 Scanlon, *What We Owe to Each Other*, 219–20.

35 Scanlon, "Replies," 429.

restrictions are thus closely related, and a Scanlonian solution to the nonidentity problem should be compatible with both.

My main concern in this paper will be with the impersonalist restriction, but let me first take a brief look at the individualist restriction. On Parfit's definition given above, the individualist restriction requires that the objections to a principle be based on its implications for "single people."[36] On the types-of-persons approach, the relevant objection to a principle appeals to its implications for types of persons. Quite simply, a type of person does not seem to be a single person. Rather, it is a generalized description of a person that could be instantiated by different individuals. In Parfit's terms, such a description is a "general person" and "general people are *not* individuals. A general person is a vast group of possible individuals, or people, one of whom will be actual."[37] In a way, the whole point of the types-of-persons approach is that objections must not be based on how a single individual is affected by a principle and its alternatives. So on this interpretation, the types-of-persons approach is a violation of the individualist restriction.

It is not so clear, however, that the individualist restriction should be understood in this way. Its central role is usually taken as barring interpersonal aggregation, and the types-of-person approach is compatible with this constraint. For example, if some people in the depleted world were miserable, and some others were very well-off, the types-of-persons approach would presumably not aggregate their well-being. Instead, the question here is whether an objection can be based on how different possible individuals of the same type would be comparatively affected. Rather than interpersonal aggregation, this is what we could perhaps call an *interpersonal comparison*. Such comparisons may be, from a Scanlonian perspective, problematic in much the same way as interpersonal aggregation, but this is far from obvious.[38] So on a narrower interpretation of the individualist restriction, one in which it excludes only interpersonal aggregation, the types-of-persons approach is compatible with it. Apart from that, it is quite contested whether Scanlonian contractualism should uphold the individualist restriction even in this narrower interpretation. In recent years,

36 Parfit, *On What Matters*, 2:193.

37 Parfit, *On What Matters*, 2:236, emphasis in original. In the context of this quote, the general person Parfit discusses is "your future child," and he explicitly refers to Kumar's view in an endnote to this sentence in *On What Matters*, 2:754.

38 Parfit suggests that failing to distinguish between two individuals by treating them as parts of the same general person ignores the "separateness of persons," a concept usually invoked in opposition to interpersonal aggregation. Parfit, *On What Matters*, 2:236, 754. Settling this question requires a deeper discussion of the impermissibility of interpersonal aggregation in Scanlonian contractualism, which I cannot provide here.

even Scanlon has expressed doubt that the strict exclusion of aggregative reasons can be defended.[39] I therefore do not want to rest my arguments on the individualist restriction, on either interpretation. A violation of the impersonalist restriction seems to be the more pressing concern.

So, how does the impersonalist restriction relate to the idea of type-based reasons? Scanlon's formulation that reasons for objection must be "on a person's own behalf" suggests a narrow interpretation, including only reasons based on how the person herself would be affected by a principle, how it would affect her own interests or her own well-being, and so on, rather than how it would affect those of her type more generally. However, Scanlon is not very clear on this, and following Parfit's definition, there might be room for type-based reasons. A lot hinges on how 'impersonal' and 'personal' are understood. On one interpretation of the restriction, the exclusion of an appeal to the *impersonal* goodness or badness of outcomes can be taken to exclude only reasons that do not refer to (what is good or bad for) persons, such as reasons referring to the intrinsic value of nature. On this understanding, type-based reasons would be valid reasons to object to principles. They are *personal* in that they refer to how persons are affected; and even in a stronger sense, they are not about just any persons but specifically about those of one's *own* type.

Without explicitly discussing the impersonalist restriction, Kumar addresses this question in a few of his papers. He accepts "the initial intuition that a person's being wronged consists in something having been done to *her*, the force of which needs to be accounted for in light of the implications for *her life*."[40] He adds, though, that this does not require that a person's objection is based on how she in particular is affected. Instead, a particular person is entitled to be treated according to what is owed to her based on more general considerations: "her entitlement, on the contractualist account, is one that she has in virtue of her circumstances being relevantly characterized by a more general description, one that could be instantiated by an indefinite number of others."[41] Similarly, in his work on risk, Kumar claims that what we owe to a person is to treat her in ways that are owed to a person *like her* more generally: "The objection is general: it is an objection not just to *you* doing what had negative consequences for *me*, but to *anyone* under this type of circumstance engaging in that type of conduct, on the grounds that it could have certain implications for another

39 Scanlon, "Contractualism and Justification."

40 Kumar, "Who Can Be Wronged?" 109, emphasis in original. See also Kumar, "Risking Future Generations," 248.

41 Kumar, "Risking Future Generations," 252.

individual (like me)."[42] This, Kumar emphasizes, "does not abandon the contractualist ideal of always conducting oneself in a way whose permissibility is justifiable to each person."[43]

Even though this sounds like an attractive way for Scanlonian contractualists to account for nonidentity cases, I am not convinced. It is one thing to consider "general objections" in the sense that all token persons in a given position share that objection to a more general principle and to be able to consider that objection even if we do not know the identity of the persons who are or will be in that position. It is another thing to attribute a general objection to a person when we know that, as a token person, she does not have that objection to the principle in question. Therefore, in the following, I will argue that type-based reasons should be excluded on a plausible understanding of the impersonalist restriction because they do not plausibly ground a personal wronging of the token person.

In nonidentity cases, we are inclined to accept a personal wronging based on type-based reasons because it leads to the intuitively right verdict. However, in other cases, this seems quite problematic. Consider the following case:

> *Medical Program*: You have two alternative therapy options to choose from for setting up a program to treat patients suffering from a rare and painful disease. Treatment *A* is more effective in improving the patients' well-being, but it is also a lot more expensive, leaving you with less budget for other health expenses. Treatment *B* is a different, less expensive therapy that also helps its patients, but it leaves them worse-off than patients would be under treatment *A*. The alternative therapy options are applicable only to persons with certain medical characteristics, so you know in advance that the patients who would be admitted to the alternative programs are nonidentical.

With the idea of type-based reasons in mind, one could argue that the type *future patient* is worse-off in *B* than in *A*, and this is what grounds the objection to the *B* treatment program from any of the affected token persons. However, we know that the decision we are about to make will determine not only the quality of treatment for future patients but also who will be treated and who will not. We should therefore be very careful about what we take the relevant type to be.

When we know in advance that a person would be treated only in the *B* program, considering the type *future patient* does not seem to adequately capture

42 Kumar, "Risking and Wronging," 50, emphasis in original.

43 Kumar, "Risking and Wronging," 50–51.

her perspective. She has a strong interest in the implementation of the *B* program to receive treatment for her painful disease. This consideration is lost when we take into account only how the general type *future patient* is affected. It seems that we should therefore use more specific type descriptions that capture these relevant reasons of the token persons, for example, the types *sick person who would be treated only in the B program*, *sick person who would be treated only in the* A *program*, and *person who would benefit from the money invested in other services*. This way, we see what is in fact owed to whom.

This supports the original interpretation of standpoints as representing token persons in moral reasoning, as discussed in the previous section. On this view, I have suggested that once we have evidence that a token person has reasons different from what we ascribed to her type, we should correct our belief about what the relevant standpoints are. Otherwise, we run the risk of defining entirely arbitrary standpoints or, worse, characterizing standpoints according to what we believe should be the result of the moral deliberation.

I thus take it that an adequate type description must be narrow enough so that the reasons of the token persons correspond to their type's reasons (with regard to the principle at stake, at least). This means that type-based reasons that a person has with regard to her type but not with regard to herself should not plausibly count as personal reasons that can ground a wronging of that person.[44] They are rather based on an inadequately broad description of her that obscures the reasons that are decisive for what we owe to her. Regarding nonidentity cases, I believe this strongly suggests that we should dismiss standpoint descriptions as general as suggested by the types-of-persons approach. An adequate standpoint description of a future person should presumably include the facts that she would not exist if the alternative principles were adopted and that she is therefore not made worse-off than she otherwise would have been (or adversely affected in any other comparative sense).[45]

But perhaps it is too early to draw conclusions. In nonidentity cases, we are concerned with future persons who are not yet born, and we thus have very limited knowledge about them. It could be argued that this is a crucial

44 Allowing personal reasons to be based on interpersonal comparisons is what Harney calls, in a welfarist context, a *pseudo-person affecting view*. See Harney, "The Interpersonal Comparative View of Welfare."

45 Note that I am claiming only that the standpoint description should be narrow enough so that it does not include reasons based on the type being comparatively affected in a way that the token person is not. I do not claim that the standpoint description should include an interest in existence (so that, for example, a future person could object to a principle that it prevents her from existing). Elizabeth Finneron-Burns argues convincingly that the latter is implausible ("Contractualism and the Non-Identity Problem," 1156–58, and *What We Owe to Future People*, 73–77).

consideration. Notably, Kumar combines his types-of-persons approach with an *ex ante* view for cases of risk or uncertainty.[46] On an *ex ante* view, what is decisive for the justifiability of a principle that allows us to impose risks on others are the objections based on their prospects (*ex ante*) rather than on how persons fare in the outcome (*ex post*). I cannot discuss here the more general plausibility of *ex ante* and *ex post* contractualism, but in many cases, it seems reasonable to adopt an *ex ante* interpretation.[47] Does the *ex ante* reasoning provide a rationale for considering the general standpoint of a future person in cases like Depletion?

In support of this, note that in Medical Program, it may seem much more plausible to consider the standpoint *future patient* when we do not know who the patients in the alternative programs would be.[48] On an *ex ante* view, considering the objections of the affected persons requires us to take seriously their prospects rather than looking at the outcomes. For a person suffering from a painful disease, about whom we cannot know whether she would be admitted to program *A* or program *B*, the fact that patients would receive better treatment in *A* than in *B* does indeed seem to be a relevant consideration. This holds even though we know that there are in fact only persons who would be admitted to *A* and persons who would be admitted to *B*. In other words, even though we know that each token person can be treated in only one of the programs, due to our epistemic limitations, it is reasonable to take into account a type of person who could be treated in either program. For this type, implementing the *B* program implies a higher risk of not receiving effective treatment, and this may be an objection to the principles allowing us to implement the *B* program. This type description may be generalized or abstract but is nonetheless very relevant for what we owe to the token persons.

Considering this, it may not seem odd at all that a person who is treated in the *B* program can be wronged by the choice of *B*, even though she benefited from the unjustified choice. The wronging consists in the action being unjustified to her, regardless of the outcome of that action. A person can be wronged in virtue of a risky behavior even when she happens to benefit from it and does

46 See especially Kumar, "Risking and Wronging" and "Risking Future Generations." See also Finneron-Burns, "Contractualism and the Non-Identity Problem," 1158, and *What We Owe to Future People*, ch. 6.

47 For defense of this claim, see, e.g., Frick, "Contractualism and Social Risk"; and Kumar, "Risking and Wronging." While Scanlon argues for an *ex post* perspective in *What We Owe to Each Other*, ch. 5, he later accepts the *ex ante* view and credits Frick's paper with having changed his mind ("Reply to Zofia Stemplowska").

48 I thank an anonymous reviewer for pressing this point.

not regret that the choice was made.[49] This kind of *ex ante* reasoning, it may be argued, is what we should apply to cases like Depletion. This way, we can see that a token person in the depleted world who neither is worse-off nor regrets our choice can nonetheless be wronged by us, in virtue of us having imposed unjustified risks on a relevant type description of her.

However, on closer inspection, I believe that the *ex ante* reasoning for cases of risk does not provide a justification for considering the general type *future person* in nonidentity cases and thus does not vindicate the types-of-persons approach. There is a crucial difference between Depletion and Medical Program. In Medical Program, when we do not know to which of the programs a sick person would be admitted, it seems reasonable to consider the standpoint of a person who could be treated in either of the programs. Importantly, though, in this case, there is a token person about whom we do not know whether she would be treated in *A* or in *B*. In order to justify ourselves to that token person, we have to take into account that, for all we know, she might end up in *A* or in *B*. Thinking about the standpoint of such a future patient, I assume, is therefore not in conflict with the impersonalist restriction. The reasons attributed to that standpoint do correspond to the token person's reasons in light of the epistemic limitations.

This is not the case in Depletion. There is no such person who, for all we know, could end up in either the depleted or the conserved world. Although we have a fair amount of epistemic uncertainty about future persons, we do *not* have the kind of epistemic uncertainty involved in Medical Program.[50] In Medical Program, to justify our actions to a token person about whom we do not know in which program she would be treated, we should take into account that the principle allowing the *B* program imposes on her a higher risk of not receiving effective treatment. In Depletion, there is no token person about whom we do not know whether she would live in the depleted or the conserved world and on whom the principles allowing Depletion thus impose higher risks. To consider such a standpoint and to attribute the standpoint's reasons to a token person are thus to attribute reasons to a person that she does not have on her own behalf, neither *ex ante* nor *ex post*. This is what makes the types-of-person approach incompatible with the impersonalist restriction, as I have argued. The *ex ante* view does not change this, nor does it provide a justification for considering a standpoint so broadly constructed in nonidentity cases.

From this, I take it that the second version of the types-of-persons approach, according to which token persons have type-based reasons, is incompatible

49 Kumar, "Risking Future Generations," 249–50, 254.

50 See also Gibb, "Relational Contractualism and Future Persons."

with the impersonalist restriction. A reason that refers to how a person's type is comparatively affected by a principle and its alternatives when this is not how the token person herself is affected is not a reason that she has on her own behalf that can ground a wronging of her. Perhaps a version of Scanlonian contractualism without an impersonalist restriction or with a revised one could be defended.[51] I cannot engage in a discussion of such a revised contractualism here, but the theoretical costs seem significantly high. The impersonalist restriction lies at the core of Scanlonian contractualism. While Scanlon has shown himself somewhat open to reconsidering the individualist restriction (to capture cases in which numbers seem to matter after all), he seems to be a lot more reluctant to loosen or give up on the impersonalist restriction.[52] And I believe there is good reason for this. It seems impossible to dismiss the restriction without losing the central ideal of justifiability to each person and thereby the resources for grounding a personal wronging in the first place.

5. Conclusion

To conclude, the types-of-persons approach is not a plausible option for Scanlonian contractualism to account for our obligations to future persons. I have suggested two alternative options to spell out such a view. On closer inspection, however, both versions prove to be incompatible with central ideas of Scanlonian contractualism. The assumption that types of persons and their reasons are what normatively matters does not align well with the concept of standpoints as originally advanced; it clashes with a plausible conception of reasons within the contractualist framework and with the idea of what we owe to each other. On this reading, Scanlonian contractualism would no longer adequately represent the idea of what we owe each other in the sense of what we owe to the persons with whom we stand in a relation of recognition rather than in the sense of what we owe to types. The assumption that the particular persons have type-based reasons to object to principles conflicts with the impersonalist restriction. A reason that is based on how a person's type is comparatively affected by a principle and its alternatives when the token person herself is not affected in that way is not a reason that the token person has on her own behalf that can ground a wronging of her. I therefore conclude that the types-of-persons approach does not provide a solution to the nonidentity problem for Scanlonian contractualism.

51 Parfit argues that Scanlonian contractualism should dismiss both the individualist and the impersonalist restrictions. Parfit, *On What Matters*, vol. 2, ch. 10.

52 Scanlon, "Contractualism and Justification," especially 35–36.

Where does this leave us? If a solution to the nonidentity problem on contractualism's own terms proves to be impossible, then perhaps a pluralist view is the only plausible way for Scanlonian contractualists to explain our obligations to the future. There may be different ways to outline such a view, but it raises new, difficult questions that need to be answered. And importantly, it means understanding the diminishment of future living conditions not as a wrong done to future persons but as an impersonal wrong. Before we become pluralists, a noncomparative account of the future persons' objections might thus be worth looking into in more detail. The notions of types of persons or standpoints, however, do not help Scanlonian contractualists to solve the nonidentity problem.[53]

Humboldt-Universität zu Berlin
martinde@hu-berlin.de

REFERENCES

Ashford, Elizabeth. "The Demandingness of Scanlon's Contractualism." *Ethics* 113, no. 2 (2003): 273–302.

Boonin, David. *The Non-Identity Problem and the Ethics of Future People*. Oxford University Press, 2014.

Dancy, Jonathan. *Ethics Without Principles*. Oxford University Press, 2004.

Finneron-Burns, Elizabeth. "Contractualism and the Non-Identity Problem." *Ethical Theory and Moral Practice* 19, no. 5 (2016): 1151–63.

———. *What We Owe to Future People: A Contractualist Account of Intergenerational Ethics*. Oxford University Press, 2024.

Frick, Johann. "Contractualism and Social Risk." *Philosophy and Public Affairs* 43, no. 3 (2015): 175–223.

Gibb, Michael. "Relational Contractualism and Future Persons." *Journal of Moral Philosophy* 13, no. 2 (2016): 135–60.

Hanser, Matthew. "Harming and Procreating." In *Harming Future Persons*, edited by David Wasserman and Melinda A. Roberts. Springer, 2009.

Hare, Caspar. "Voices from Another World: Must We Respect the Interests

53 I am very grateful to Kirsten Meyer, Jonas Harney, Johanna Privitera, Gustaf Arrhenius, Krister Bykvist, Simon Hollnaicher, Valentin Salein, Louis Kohlmann, and the reviewers of this journal for their incredibly helpful comments on earlier versions of this article. I would also like to thank the participants in research seminars and workshops at HU Berlin, LMU Munich, and the Institute for Future Studies in Stockholm for valuable discussions and encouragement.

of People Who Do Not, and Will Never, Exist?" *Ethics* 117, no. 3 (2007): 498–523.
Harman, Elizabeth. "Can We Harm and Benefit in Creating?" *Philosophical Perspectives* 18, no. 1 (2004): 89–113.
Harney, Jonas. "The Interpersonal Comparative View of Welfare: Its Merits and Flaws." *Journal of Ethics* 27, no. 3 (2023): 369–91.
Hurley, Paul, and Rivka Weinberg. "Whose Problem Is Non-Identity?" *Journal of Moral Philosophy* 12, no. 6 (2015): 699–730.
Huseby, Robert. "Person-Affecting Moral Theory, Non-Identity and Future People." *Environmental Values* 19, no. 2 (2010): 193–210.
Katz, Corey. "Contractualism, Person-Affecting Wrongness and the Non-Identity Problem." *Ethical Theory and Moral Practice* 21, no. 1 (2018): 103–19.
Kumar, Rahul. "Risking and Wronging." *Philosophy and Public Affairs* 43, no. 1 (2015): 27–51.
———. "Risking Future Generations." *Ethical Theory and Moral Practice* 21, no. 1 (2018): 245–57.
———. "Who Can Be Wronged?" *Philosophy and Public Affairs* 31, no. 2 (2003): 99–118.
———. "Wronging Future People: A Contractualist Proposal." In *Intergenerational Justice*, edited by Axel Gosseries and Lukas H. Meyer. Oxford University Press, 2009.
McMahan, Jeff. "Climate Change, War, and the Non-Identity Problem." *Journal of Moral Philosophy* 18, no. 3 (2021): 211–38.
Parfit, Derek. *On What Matters*. 2 vols. Oxford University Press, 2011.
———. *Reasons and Persons*. Oxford University Press, 1984.
Raz, Joseph. *Engaging Reason: On the Theory of Value and Action*. Oxford University Press, 1999.
Scanlon, T. M. *Being Realistic About Reasons*. Oxford University Press, 2014.
———. "Contractualism and Justification." In *Reason, Justification, and Contractualism: Themes from Scanlon*, edited by Markus Stepanians and Michael Frauchinger. De Gruyter, 2021.
———. "Contractualism and Utilitarianism." In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams. Cambridge University Press, 1982.
———. "Replies." *Ratio* 16, no. 4 (2003): 424–39.
———. "Reply to Zofia Stemplowska." *Journal of Moral Philosophy* 10, no. 4 (2013): 508–14.
———. "Responses to Forst, Mantel, Nagel, Olsaretti, Parfit, and Stemplowska." In *Reason, Justification, and Contractualism: Themes from Scanlon*, edited by Markus Stepanians and Michael Frauchinger. De Gruyter, 2021.

———. *What We Owe to Each Other*. Harvard University Press, 1998.
Schroeder, Mark. *Slaves of the Passions*. Oxford University Press, 2007.
Schwartz, Thomas. "Obligations to Posterity." In *Obligations to Future Generations*, edited by Richard I. Sikora and Brian Barry. White Horse, 1978.
Shiffrin, Seana. "Wrongful Life, Procreative Responsibility, and the Significance of Harm." *Legal Theory* 5, no. 2 (1999): 117–48.
Skorupski, John. *The Domain of Reasons*. Oxford University Press, 2010.
Snedegar, Justin. *Contrastive Reasons*. Oxford University Press, 2017.
Suikkanen, Jussi. "Contractualism and Climate Change." In *Canned Heat: Ethics and Politics of Climate Change*, edited by Marcello Di Paola and Gianfranco Pellegrino. Routledge, 2014.
Wallace, R. Jay. *The Moral Nexus*. Princeton University Press, 2019.

Journal of Ethics and Social Philosophy
VOL. 29, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.3128

SHOULD ALL FREEDOM BE BASIC?

Seena Eftekhari

In a recent paper, Jessica Flanigan claims that the logic internal to Rawlsian high liberalism requires elevating nearly all freedoms to the list of basic rights.[1] 'High liberalism' refers to a set of liberal views, of which Rawls's justice as fairness serves as one notable example, that weds a commitment to freedom with a commitment to the common good.[2] On the one hand, Rawls's commitment to freedom is reflected in the lexical priority (referred to as the priority of liberty principle) he attaches to the list of freedoms protected as basic rights by his first principle of justice. Because these rights are supposed to protect citizens' most fundamental interests, the priority of liberty principle prohibits absolutely the violation of basic rights for the sake of protecting nonbasic rights or for the sake of promoting either perfectionist values or the common good. Basic rights may be violated only for the sake of protecting other basic rights. They consequently constrain the extent to which the state may legitimately interfere with individual freedom for the sake of promoting the common good—that is, for the sake of promoting things such as equality of opportunity, economic growth and a fairer distribution of wealth, public health, campaign finance reform, national defense, etc. Rawls specifies the list of basic rights very narrowly for this reason, as including too many freedoms would undermine his commitment to the common good, and it is also partly for the same reason that he excludes capitalist economic freedoms. The viability and much of the appeal of high liberalism depend on its ability to synthesize these two commitments into a harmonious whole in which individual freedom does not completely eclipse or marginalize the normative and political importance of the common good.

Insofar as Flanigan argues that nearly *any* freedom should count as basic in this Rawlsian sense, then her argument threatens Rawlsian high liberalism by completely marginalizing its commitment to the common good.[3] When

1 Flanigan, "All Liberty Is Basic." I use the terms 'freedom' and 'liberty' interchangeably.

2 For a survey of high liberalism in contrast to classical liberalism and libertarianism, see Freeman, "Capitalism in the Classical and High Liberal Traditions."

3 Her argument is therefore more threatening than the argument pressed by John Tomasi, who argues that capitalist economic freedoms ought to be included within the list of basic

nearly any freedom qualifies as a basic right, there will be little if any political and legal space remaining for the state to promote the common good. To the extent that her argument requires marginalizing the common good for the sake of individual freedom, it thereby implies that the very logic internal to Rawlsian high liberalism is responsible for paving the way to what are standard libertarian conclusions, such that in the end there will be little theoretical and practical differences between the two. For if nearly all freedoms are to be protected as basic rights, then it turns out that the primary if not exclusive function of the state becomes the protection of individual freedom.

My aim in this paper is to survey potential Rawlsian high-liberal responses to Flanigan's argument. To do so, the paper is organized as follows. Section 1 provides a very brief exposition of Rawls's view. As has been mentioned already, a key feature of basic rights is that they are not to be violated for the sake of promoting the common good, so the section also explains in greater detail what is meant by the 'common good'. Section 2 draws out the implications of Flanigan's argument for the viability of the high-liberal project. Flanigan considers and rejects several responses offered by high liberals in her paper, so section 3 considers two novel responses. I argue that both responses depend on controversial value judgments, making them unattractive from the high-liberal perspective. Section 4 turns to the most promising response. I argue that to stave off the implications that follow from Flanigan's argument, Rawlsian high liberals should abandon the priority principle and preserve their commitment to the common good. Doing so not only allows them to preserve the more egalitarian aspects of their views but also brings additional benefits. Preserving their commitment to the common good allows Rawlsian high liberalism to operationalize the harm principle (section 4.1); to determine a coherent, compossible scheme of basic rights (section 4.2); and to adjudicate conflicts between citizens' basic rights (section 4.3). Abandoning the priority principle thus preserves high liberalism's theoretical flexibility in a way that allows it to address important political problems. Lastly, section 5 rejects an objection that attempts to preserve the priority principle.

1. RAWLS'S VIEW

1.1. The Connection Between Personhood and Basic Rights

For Rawlsian high liberals, there are two important differences between basic rights and nonbasic rights. The first important difference is that the basic rights

rights. Tomasi, *Free Market Fairness*. For criticisms of Tomasi's arguments, see Arnold, "Right-Wing Rawlsianism"; Patten, "Are the Economic Liberties Basic?"; and Melenovsky and Bernstein, "Why Free Market Rights Are Not Basic Liberties."

protected by Rawls's first principle of justice possess absolute normative and political status relative to the rights not protected by that principle (but not to other basic rights). This difference in status between the two categories of rights is reflected in the scope of reasons capable of justifying their violation. In virtue of their absolute normative and political status, basic rights may not be violated for the sake of protecting nonbasic rights, promoting the common good, or promoting perfectionist values.[4] Rawls refers to this feature of basic rights as a commitment to the priority of liberty. The priority of liberty principle expresses the idea that the rights (i.e., the freedoms) protected by the first principle of justice take absolute priority over other rights and the second principle of justice within Rawls's theory. Nonbasic rights, in contrast, may be violated for the sake of promoting the common good.

The second difference between basic and nonbasic rights explains why the former carry absolute weight. Rawls singles out his basic rights in virtue of their connection to his conception of personhood. Under the Rawlsian framework, fully cooperating moral persons, or citizens, possess two moral powers: the capacity for a sense of justice and the capacity to form, revise, and pursue a conception of the good. What distinguishes basic rights from nonbasic rights is that basic rights are necessary for the adequate development and exercise of the two moral powers.

Rawls identifies several rights that are to be protected by his first principle of justice. These rights include freedom of thought and liberty of conscience, political liberties, freedom of association, integrity of the person, and the rights and freedoms covered by the rule of law.[5] Notably, Rawls excludes any economic rights—such as a right to own the means of production, a right protecting freedom of contract, or other entrepreneurial rights—from the first principle. According to him, these rights are not necessary for the development and exercise of the moral powers and, for that reason, do not qualify as basic rights.

1.2. The Common Good

In expressing his commitment to the priority of liberty, Rawls tells us that the basic rights protected by the first principle of justice "have an absolute weight with respect to reasons of public good."[6] Basic rights cannot be violated for the sake of promoting the common good. What exactly constitutes the common good, however? Rawls does not expand on precisely what sorts of reasons count as "reasons of public good" in *Political Liberalism*, although

4 Rawls, *Political Liberalism*, 294–95.

5 Rawls, *Political Liberalism*, 291.

6 Rawls, *Political Liberalism*, 294.

he does provide a few examples: reasons that pertain to economic efficiency and growth, as well as a discriminatory selective service act for the purpose of raising an army. In the revised edition of *A Theory of Justice*, he provides a bit more detail, as he uses the term to refer to the interests shared by citizens from their position as citizens.[7] More specifically, these shared interests include both the scheme of basic rights and the fair equality of opportunity principle. These brief comments suggest that there may be a tension between how the concept is employed in *Political Liberalism* and the way in which Rawls understands it in *A Theory of Justice.* For it appears that in *Theory*, the scheme of basic rights comprises one part of the common good rather than a categorically different set of interests. One way to resolve the tension here is by appealing to the priority of liberty principle: even if the scheme of basic rights were one part of the common good as is suggested by *Theory*, Rawls's commitment to the priority of liberty would nonetheless prohibit this part of the common good from being violated or sacrificed for the sake of promoting other parts of the common good, such as fair equality of opportunity or a particular distribution of wealth.

In either case, because Rawls does not specify the concept in sufficient detail and because the concept plays an important role in the paper, I provide some general remarks as to how I understand it. To do so, I will follow—albeit with one important qualification—the account laid out by Waheed Hussain in his *Stanford Encyclopedia* article on the concept. In his article, Hussain identifies five central features of the common good: (1) the common good provides a shared account of practical or deliberative reasoning for citizens of a political community; (2) the common good refers to a set of facilities—political institutions, human artifacts (hospitals, schools, roads, etc.), and the environment—that serve some common interests; (3) the set of interests included with the common good are a *privileged* set of common interests; (4) the common good expresses a solidaric concern on the part of citizens to give the same status to the interests of other citizens as they do to their own in their deliberation; and finally, (5) "most" conceptions of the common good, according to Hussain, do not take an aggregative view of citizens' interests.[8]

For the purposes of this paper, the one "modification" I propose is to allow aggregative concerns to form a legitimate part of our understanding of the concept. Given that Hussain concedes that some conceptions of the common good already allow aggregation to play a role, it really is not a modification at all. It is nonetheless important to emphasize the aggregative aspect, because although Hussain tends to minimize its importance, he recognizes that different

7 Rawls, *A Theory of Justice: Revised Edition*, 82–83, 271.

8 Hussain, "The Common Good."

conceptions of the common good will have different ways of specifying the privileged set of common interests. For example, some conceptions of the common good—those he refers to as "private individuality" conceptions—define that set by appealing to the common interest that individuals have in pursuing lives as private individuals. Similarly, a "distributive" conception of the common good does *not* require that citizens abstract from their sectional or private interests in evaluating competing political policies. Importantly, both conceptions are more sensitive to the possibility of conflict between citizens' interests. For instance, although citizens may possess a common interest in having *a* scheme of basic rights that allows them to pursue private lives, there will be significant disagreement over which scheme of rights to implement. Even though all citizens possess a common interest in being able to live as private individuals, the exact contours of their lives will obviously differ under different schemes of rights, and this will explain why one scheme will be in the interest of some citizens but another less so. Consequently, the support citizens may express for one scheme over another will often be motivated by their sectional or private interests.

While some might be inclined to reject aggregative conceptions of the common good as viable candidates, they are important insofar as it is not always possible to arrange the political, economic, and social institutions to accommodate all reasonable forms of life (a point Rawls concedes, as we will see). In circumstances like these, having to determine which interests to promote based on aggregation would be invaluable and likely unavoidable. In fact, as we will see in the disputes between high liberals and their opponents, the arguments offered by both often rely on making appeals to the common good in precisely this fashion.

2. The Implications of Flanigan's Argument for the Distinction between Basic and Nonbasic Rights

In the second part of Flanigan's paper, she extends the reasoning within Rawls's framework to cover nearly all other freedoms. The argument is relatively straightforward: just as some citizens may find religion to constitute a fundamental part of their identities and to play a central role in expressing authorship over their lives, many other pursuits and freedoms will play a similar role for other citizens. Flanigan illustrates this point with the example of the fervid Green Bay Packers football fan, Owen. Insofar as Owen "identifies as a Packers fan above all else," his support for the Packers shapes his sense of self; it infuses all aspects of his life with vibrancy and color—in a literal sense too, as he even desires to buried in a casket painted with team colors—to make life meaningful for him. It is no accident that Flanigan describes Owen's fandom in quasi-religious terms. After all, as

she describes him, Owen would be willing to give up both religious protections and the right to vote before he would be willing to give up his right to support the Green Bay Packers. The freedom to pursue a life centered around athletic fandom, like the freedom of religious worship and the freedom to start and run a business, is *important* for *some* individuals to develop their moral powers.[9] And this argument does not apply only to athletic fandom. Insofar as *any* liberty *could* play this sort of role within an individuals' life, Flanigan concludes that it should therefore be elevated to a basic right.[10]

How does Flanigan begin with this claim and arrive at the conclusion that nearly any liberty could play this role? The answer lies in assumptions found within Rawlsian high liberalism about the nature of value and how it is created in the world—assumptions that, when taken independently, appear to be uncontroversial and relatively innocuous but that produce radical implications when combined.

The first assumption is that under a liberal, democratic regime, citizens possess equal moral worth in virtue of their possession of the two moral powers. One important aspect of this moral worth, as Rawls describes it, is that citizens are "self-authenticating sources of valid claims [who are] able to make claims on their institutions so as to advance their conceptions of the good."[11] These claims made by citizens include at minimum claims against interference with their freedom and their ability to pursue their (reasonable) conceptions of the good.

The second assumption is that the world is devoid of intrinsic, objective value until the choices of citizens (and groups of citizens) imbue the world with value. Various ritualistic practices and activities, objects, lifestyles, and beliefs that were once meaningless and devoid of value suddenly become infused with value through the choices of citizens, particularly when they become part of

9 I say that the basic rights are *important* rather than *necessary* because the former seems to be Flanigan's view, but there is an ambiguity here that is not satisfactorily cleared up by her comments. The ambiguity stems from the fact that it is possible to conjoin different modalities with different referential scopes. As some examples, it is possible that basic rights must be *necessary* for *everyone* to develop their moral powers, *necessary* merely for *some* citizens, or *merely important* for *anyone*. In the end, I do not believe that much rides on requiring one modality over another. For even if the basic rights must be *necessary*, as long as they must be necessary only for *some* individuals, then Flanigan's argument would still go through. I merely mention all this to forestall any potential objections insofar as Flanigan presents her argument as being internal to Rawls's position despite employing a different modality.

10 Flanigan, "All Liberty Is Basic," 466–70. Although Flanigan often refers to examples of basic liberties that she believes *do* play such a central role in the lives of individuals, she denies that the status of any liberty ought to be determined by popular support for it at a given moment of time (467).

11 Rawls, *Justice as Fairness*, 23.

citizens' life plans and conceptions of the good. Think of how religious believers of different denominations imbue all sorts of practices, behaviors, and objects with sacramental value. The story is similar for fanatical football fans like Owen. For him, a life centered around supporting the Packers—including all the sundry activities from the superstitious pregame rituals to the communal weekly tailgating, the momentary merging of the self with other fans into a collective body cheering in unison towards a common goal, and so on—all become infused with tremendous value. The practices, activities, lifestyles, and beliefs imbued with value comprise and represent fundamental aspects of their life plans in which their identities and sense of self are firmly rooted.

The combination of these two assumptions leads to very radical implications. If individuals are self-authenticating sources of valid claims against political institutions, then nearly every single act of valuation by a citizen—every time a citizen imbues a practice, activity, or object with value as part of their life plans—expresses a *politico-legal act.* Citizens are not merely making a personal declaration about what matters to them; in doing that, they are also making a declaration to both other citizens and the state. According to Flanigan's argument, the mere fact that they imbue (or can imbue) nearly any practice or activity with value is sufficient to elevate the associated freedoms to the status of basic rights. Nearly every single instance of valuation has the effect of creating correlative obligations on the part of both other citizens and the state to respect and not interfere with the exercise of those basic liberties. It is no exaggeration to say that such a view would make every citizen "a law unto himself."[12]

The radical implications become more vivid when one remembers that what distinguishes basic rights from nonbasic rights is the scope of reasons capable of justifying violations. Basic rights may be violated only when it is necessary to protect other basic rights, not for any other reason. Taken in the context of Flanigan's argument, it follows not only that citizens' acts of valuation create correlative obligations on the parts of others to not interfere but that those same acts—insofar as they elevate the status of those freedoms to *basic* rights—also prohibit the state from interfering with those liberties for the sake of promoting the common good. In establishing the scheme of basic rights, citizens' acts of valuation thereby determine the bounds of permissible legislative and judicial behavior in the future. In the end, individual acts of valuation take normative and political priority over other all other political values and social interests.

In making this argument, Flanigan's aim is to demolish the distinction between basic rights and nonbasic rights. She can thus claim not only that her

12 *Employment Division v. Smith*, 494 U.S. 872 (1990), 879, 885, 890 of the majority opinion.

view prioritizes freedom to an even greater extent than (unmodified) Rawlsian high liberalism does but also that, in doing so, she "strengthen[s] the presumption of liberty that high liberals already endorse."[13] What makes her argument so challenging from the perspective of Rawlsian high liberals is that their own commitments to personhood and to the priority of liberty principle pave the way to what are essentially libertarian conclusions. For if nearly all freedoms become basic, and their protection takes priority over all other political values and social interests, the common good can be pursued only within the extremely narrow confines established by the (ever-growing) list of basic rights.

It is probably in the spirit of softening this blow that Flanigan claims her argument does not entail that public officials must abandon their commitment to social justice. For those reasonably skeptical of this claim, she insists that institutions can still be arranged to benefit the worst-off: "The citizens who are most vulnerable to abuses of state power are the worst-off, so they are the most likely to benefit from policies that protect them from government interference."[14] I must confess my perplexity with this claim. For under her view, the central purpose of the state is to protect the exercise of basic rights; therefore, what constitutes "abuses of state power" would be the violation of individuals' basic rights. Her view thus has the great irony of turning capitalist magnates into the worst-off, as their economic liberties in particular would be the most vulnerable to "abuses by state power" for a state that seeks to promote a fairer distribution of wealth.

Notwithstanding these cursory remarks made by Flanigan, there is little encouragement to be gleaned for Rawlsian high liberals, and the issue stems in large part from their commitment to the priority of liberty principle. As we have seen, that principle attributes *absolute* status to basic rights in relation to various components of the common good—whether those components be equality of opportunity, economic growth, distributive justice, campaign finance reform, public health measures, and so on. The principle prohibits violating citizens' basic rights for any of those reasons. Thus, the state's ability to promote the common good is limited in direct proportion to the number of freedoms included within Rawls's list of basic rights. Consider the recent debate over the status of capitalist economic freedoms. If entrepreneurial freedoms and private ownership over the means of production are basic rights, then the state may not interfere with those rights for the sake of promoting either nondiscriminatory business practices or distributive justice. Flanigan's argument further exacerbates the problem as it elevates not merely capitalistic economic freedoms to basic rights as John Tomasi's argument does but nearly

13 Flanigan, "All Liberty Is Basic," 471.

14 Flanigan, "All Liberty Is Basic," 472.

any freedom. Once those freedoms have been elevated to the status of basic rights, it follows that the state is prohibited from interfering with them for the sake of promoting the common good. There is little reason for high liberals to be optimistic about this situation, as the odds that a policy will not conflict with some citizens' basic rights when nearly any freedom counts as basic are very low. Policies would have to be crafted to avoid such conflicts, but it is not clear what political and legal space, if any, remains for policies that could manage to avoid these conflicts. Prioritizing—or we might more accurately say fetishizing—individual freedom in this way marginalizes the normative and political significance of other worthy political values and goals besides freedom. The outcome is the total disruption of the harmonious balance of political values and aims that is so integral to the high-liberal project. And once the commitment to the common good is marginalized for the sake of individual freedom in this way, then it appears that Rawlsian high liberalism shares with standard versions of libertarianism the same all-encompassing normative and political commitment to individual freedom.

3. Two High-Liberal Responses to Flanigan's Argument

Instead of revisiting the responses considered by Flanigan in her paper, I explore three additional responses that high liberals might offer. The first argument will take Rawls's response to the critiques originally raised by H. L. A. Hart and apply that response to Flanigan's argument. The second argument will rely on recasting Rawls's original strains-of-commitment argument for the purpose of restricting the list of basic rights. In section 4, I will consider the third argument, which entails abandoning the priority of liberty principle altogether. In this section, I argue that the first two responses fail because each implicitly relies on controversial value assumptions about what constitutes the good life, thereby rendering those responses incompatible with the Rawlsian high-liberal desire to be neutral on these matters. Thus, if Rawlsian high liberals want to avoid the implications that follow from Flanigan's argument and to preserve their commitment to the common good, the best alternative would be to abandon the priority principle.

3.1. A First Response: Taking Rawls's Response to Hart and Applying It to Flanigan

The untoward implications that follow from incorporating too many freedoms within the first principle of justice were brought to Rawls's attention by H. L. A. Hart, and Rawls attempts to address the issue in subsequent work.[15] Rawls's

15 See Hart, "Rawls on Liberty and Its Priority."

response to Hart, first given in the 1981 lecture "The Basic Liberties and Their Priority," can also be seen as anticipating Flanigan's more recent argument.[16] One might thus attempt to marshal his argument in that work in order to counter Flanigan's argument.

In that essay, Rawls insists that the scope of the rights included under the first principle must be specified in narrower terms if the first principle of justice is going to be not only internally coherent but also compatible with the ambitions of the second principle. To do so, Rawls appeals once more to his idea of personhood. Just as that idea played a fundamental role in determining the general list of basic rights, he returns to the two moral powers to further delimit the scope of the rights included within the first principle:

> The notion of the *significance* of a particular liberty … can be explained in this way: a liberty is more or less significant depending on whether it is more or less essentially involved in, or is a more or less necessary institutional means to protect, the full and informed and effective exercise of the moral powers in one (or both) of the two fundamental cases.[17]

Consider the right to freedom of speech. Rawls does not hold the view that any form of speech should be protected by the first principle of justice; in fact, he cites commercial speech and various forms of advertising as examples that do not.[18] Although free speech may be exercised in different ways, only some forms are necessary for the development and exercise of the two moral powers. Those forms that are necessary constitute the "central range of application" of the more abstract right to freedom of speech.[19] The point behind this idea of the "central range of application" is to specify the abstract rights in more specific terms by determining which forms of speech (or of other rights on the list) are significant in the sense described in the quotation above and should, for that reason, be protected by the first principle of justice. The additional step of specifying the precise scope of every single basic right must always be completed.[20] Most importantly, and this point cannot be understated, the additional work of determining the precise scope of the rights protected by the first principle will have very significant implications for the scope of legitimate legislative and judicial action. Because Rawls did not view commercial speech and advertising

16 The lecture appears in Rawls, *Political Liberalism*, 289–371.

17 Rawls, *Political Liberalism*, 335–36.

18 Rawls, *Political Liberalism*, 363–68.

19 Rawls, *Political Liberalism*, 298.

20 This is all part of the successive process of specification outlined by Rawls (*Political Liberalism*, 336–40).

as falling within the "central range of application" of the abstract right to freedom of speech—in virtue of the fact that he did not believe either necessary for the development and exercise of the moral powers—then those forms of speech could be interfered with for the sake of promoting the common good.[21]

It should now be possible to see how this idea could be employed to rebut Flanigan's argument. In brief, the claim is that the many exercises of freedom pointed out by her—such as Owen's sports fandom—do not fall within the central range of application of any of the rights enumerated in the first principle because they fail to hold the requisite modal connection with the moral powers. As a consequence, those exercises of freedom would not qualify as basic and could therefore be interfered with for the sake of promoting the common good.

The problem with this sort of response on behalf of high liberals like Rawls is it relies on controversial assumptions about what sort of life plans are "significant," and which freedoms are connected to those life plans. They would have to argue that some freedoms are not connected in the right way to the development and exercise of the two moral powers. But on what sort of basis can they make an argument like this? In formulating their life plans, citizens like Owen signal which freedoms are in fact important to them. Rawls's belief that it is possible to specify *a priori* the significance of various freedoms is at odds with the liberal assumption, mentioned in the previous section, that the choices made by citizens are responsible for instilling value into the world. If there are no objective values prior to the choices made by citizens, but rather all values are, in a Sartrean fashion, a consequence of their choices, then those citizens become the final arbiters in determining which freedoms possess the right sort of connection to the development and exercise of their moral powers. In offering this sort of response to Flanigan's argument, Rawlsian high liberals would be implicitly relying on the sort of perfectionist considerations they expressly seek to exclude from their theories and in doing so would bring their versions of liberalism closer to perfectionist versions like John Stuart Mill's. Perfectionist liberals like Mill tend to instrumentalize and consequently attenuate the significance of some freedoms—particularly capitalistic economic freedoms—based on their beliefs about the diminished value of lives centered around economic

21 Consider the implications within this context of the United States Supreme Court's decision to treat corporations' financial contributions to political campaigns as an exercise of speech protected by the First Amendment of the Constitution in *Citizens United v. Federal Election Committee*, 588 U.S. 310 (2010). As another example, consider how freedom of the press/speech has become conflated with the commercial freedoms of corporate media owners so that any state regulation of or infringement with their commercial freedoms is now construed as an attack upon a fundamental American freedom. See Phelan and Dawes, "Liberalism and Neoliberalism," 9, 19.

pursuits.[22] It is thus difficult to deny the point raised by critics of Rawls like Tomasi and Flanigan, who argue that his identification of basic rights implicitly harbors value judgments about different life plans. According to these critics, a neutral approach—one compatible with the aspirations of Rawls's own theory—would exclude neither capitalistic economic freedoms nor the freedoms associated with sports fandom (and many others) from the first principle of justice.

3.2. A Second Response: Recasting Rawls's Strains-of-Commitment Argument

A second solution may be found in Rawls's idea of the strains of commitment, which are the "strains that arise in such a society between its requirements of justice and the citizens' legitimate interests its just institutions allow."[23] This idea serves three purposes within Rawls's theory. Its first and most general function is to reinforce the selection of the two principles of justice over utilitarianism. Its ability to perform this function depends on the fact that the selection of principles of justice is made only once and for perpetuity, which means that the parties in the original position must see those they represent as always being able to abide by the agreement made there.[24] If the principles of justice selected in the original position turned out, once the veil was lifted and the principles implemented, to generate significant strains between the requirements of justice and the pursuit of citizens' *legitimate* interests, then citizens could not reasonably be expected to abide by the requirements of justice. That justice as fairness does not violate the strains of commitment, whereas utilitarian principles of justice do, is, Rawls insists, a significant advantage of his theory.

Rawls also employs the idea to support the lexical priority of the first principle of justice over the second principle. In *A Theory of Justice*, Rawls writes that "in order to secure their unknown but particular interests from the original position, [the parties in the original position] are led, in view of their strains of commitment, *to give precedence to basic liberties*."[25] Robert Taylor further elaborates on this connection: "political principles that place fundamental interests (such as the religious interest) at even the slightest risk, by refusing lexical priority to the liberties that protect them, make the strains of commitment intolerable."[26] This would be intolerable because the failure to attribute lexical priority to something like freedom of religious worship—thereby allowing trade-offs between

22 For a summary of this pattern within the history of liberal thought, see Gaus, *The Modern Liberal Theory of Man*, 238–39.

23 Rawls, *Political Liberalism*, 17.

24 Rawls, *Justice as Fairness*, 103.

25 Rawls, *A Theory of Justice*, 475 (cross-reference removed, emphasis added).

26 Taylor, "Rawls's Defense of the Priority of Liberty," 252.

it and the common good—reflects a failure to grant it the appropriate weight it requires. Considering that freedom of religious worship plays such a pivotal and fundamental role in the lives of so many citizens, it would not be reasonable to expect those citizens to abide by principles of justice that permit infringement upon religious freedom for the sake of promoting the common good.[27]

What is of particular interest is that at least in one part of *Political Liberalism*, Rawls employs the strains of commitment as an argument for *including* certain freedoms within the first principle.[28] Now, it must be admitted that in these passages, Rawls does not explicitly mention the strains of commitment. Nonetheless, it is clear from his comments that an implicit appeal to that idea is performing the argumentative work. To see why, consider that in these passages, Rawls says that the parties in the original position are motivated to protect freedom of religious worship for two reasons. The first is that the veil of ignorance prevents them from knowing the determinate conceptions of the good held by those they represent. But if this were the only reason, then the parties may be tempted to gamble by supporting a dominant religion and prohibiting or restricting the freedom of religious worship for all others. Rawls, however, denies that the parties in the original position would be willing to take such a gamble: "If the parties were to gamble in this way, they would show that they did not take the religious, philosophical, or moral convictions of persons seriously, and in effect, did not know what a religious, philosophical, or moral conviction was."[29] Similar to the support offered for the lexical priority of liberty, Rawls again appeals to the special importance of religious faith to explain its inclusion within the first principle. Because religious faith is so important to so many individuals, the parties would not risk limiting freedom of religious worship on the chance that those they represent end up belonging to a minority religion. Given the importance of religious freedom, the strains of commitment would be too much in circumstances where it is not protected to reasonably expect citizens to abide by the requirements of justice. Gambling with this freedom demonstrates, as Rawls says, an ignorance of the nature of religious, moral, and philosophical convictions.

In all three cases, the force behind the strains of commitment depends on the assumption that some beliefs or convictions carry heightened importance, at least

27 "Here it is fundamental that affirming such views and the conceptions of the good to which they give rise is recognized as non-negotiable. . . . They are understood to be forms of belief and conduct the protection of which we cannot properly abandon or be persuaded to jeopardize for the kinds of considerations covered by the second principle of justice" (Rawls, *Political Liberalism*, 311–12).

28 Rawls, *Political Liberalism*, 310–15.

29 Rawls, *Political Liberalism*, 311.

for some individuals. Some convictions rise above the level of mere superficial fancies, likes, or preferences, and it is only the beliefs that belong to the former that can produce the sorts of genuine motivational strains between citizens' desire to follow the requirements of justice on the one hand and their desire to pursue their legitimate convictions on the other. For this reason, not only do the freedoms protecting those beliefs need to be included within the first principle of justice, but that principle also must be attributed priority over the second principle.

This suggests a possible route for employing the strains of commitment as a means for restricting the number of freedoms included within the first principle of justice. This response on behalf of Rawls relies on drawing a substantive distinction between life plans centered on religious, moral, and philosophical beliefs from life plans based on "less weighty" beliefs, such as (according to this argument) being a fan of the Green Bay Packers. The idea would be that because the life plans that fall into the second category are not capable of invoking the strains of commitment like those in the first category are, then the freedoms associated with those life plans would not merit the same sort of privileged protection.

Now, at first glance it appears this response also runs afoul of the assumption of the subjectivist basis for value creation that was highlighted earlier. If individuals are the sources of value in the world, then on what extrapersonal basis can it be argued that some of their acts of valuation—specifically, those nonreligious, nonmoral, and nonphilosophical acts—fail to generate, we might say, "enough" or the "right sort" of value needed to invoke the strains of commitment as a relevant consideration? It is interesting to note that contemporary society already harbors these implicit value distinctions in some contexts in a way that is not altogether consistent with that subjectivist assumption insofar as exemptions are routinely made for religious believers that are not similarly extended to followers of many secular life plans. The question of course is whether Rawlsian high liberals can accommodate this thought, especially once it has been conceded that such freedoms are important for some individuals to develop and exercise their moral powers. Another way to articulate the end goal here is by asking whether it is possible to draw a principled distinction between religious, moral, and philosophical convictions from more "prosaic" (if that is the right word) life plans, such as sports fandoms.

In posing and trying to answer this question, Rawlsian high liberals are led down a thorny road. For whatever consideration is appealed to as a basis for such a distinction, it will have the consequence of conflicting with the way that many citizens relate to their life plans. After all, according to this argument, those life plans that fall on the prosaic side of the distinction would not be capable of generating the strains of commitment, despite citizens' own protestations about how their identities and their sense of self-worth are inextricably

bound up with them. This does not mean that it is impossible to draw such a distinction or even that such a distinction is implausible (there may indeed be some intuitive plausibility to it), but it highlights that this sort of response will rely on controversial value assumptions.

Having made these cautionary remarks, I will briefly consider some potential bases for drawing this distinction between religious, moral, and philosophical convictions from more prosaic life plans.

One possibility is to adopt a quasi-empirical, historically oriented approach. The approach is quasi-empirical insofar as observations taken from an empirical and historical standpoint would be used as a guide for drawing a value distinction. For instance, when we look to the history of Western civilization, we observe that it has continually been riven by religious persecution and war.[30] Yet notwithstanding centuries of religious intolerance, persecuted religious believers from all denominations have persevered in their faith, even to the point of death. That so many individuals throughout history have been willing to go to such great lengths—including the loss of their own lives—for the sake of their religious beliefs gives us good reason to believe that the prohibition of religious worship would invoke the strains of commitment.[31] Of course, this way of drawing the distinction establishes an awfully high threshold for what would count as the sort of legitimate interests capable of invoking the strains of commitment, and it would, as a result, exclude many of the freedoms Flanigan wants to include within Rawls's first principle. If the distinction were to be grounded on a relatively less dramatic or extreme basis—say, whether individuals would be willing to lose other freedoms through imprisonment instead of whether they would be willing to face death for their beliefs—then freedoms like sports team fandom might not pass the test, giving Rawlsian high liberals the conclusion they seek.[32] It is interesting to note, however, that economic liberties (capitalist or socialist, depending on our historical point of reference)

30 Ronald Beiner sees Rawls as providing a genealogy of liberalism in which the paramount concern has been to allay the potentially combustible effects of religious pluralism on political and civil society. See Beiner, *Civil Religion*, ch. 23.

31 In congruence with the hypothetical argument offered here, Taylor finds the strains of commitment to be a relevant consideration when it comes to supporting the priority of freedom of religious worship but not many other freedoms ("Rawls's Defense of the Priority of Liberty," 252–53).

32 One could, following Rawls, attempt to draw the distinction by defining the moral, philosophical, and religious convictions by focusing on the idea of comprehensiveness. According to Rawls, a comprehensive doctrine includes conceptions of what is of value in human life, ideals of personal character, and friendship, and it should inform our conduct and our life as a whole (*Political Liberalism*, 13). But it is difficult to see why many modern life plans would fail to satisfy this notion of comprehensiveness insofar as they do pertain to many if not all

would seem to pass this second test, as many individuals have historically been willing to face imprisonment or worse for their beliefs about the injustice of different economic systems.

In either case, appealing to the strains of commitment as a way of rebutting Flanigan's argument requires that Rawlsian high liberals make judgments about the worthiness of different life plans from an objective, normative standpoint—worthiness in the sense that only some are so important that they can plausibly be said to invoke the strains of commitment. Notice also that either of the two bases considered for this distinction is a matter of stipulating how citizens would respond to prohibition and oppression as a means for gauging the importance of those life plans.[33] This is, to say the least, both an idiosyncratic and morally questionable method for drawing such a distinction.

Nonetheless, the lesson to be drawn from these two initial responses is that it is incredibly difficult to draw a distinction between life plans and their associated freedoms without relying implicitly or explicitly on controversial value judgments. Since Rawlsian high liberals seek to avoid having to make these value judgements, another response is necessary. The third response I consider allows them to avoid this problem, but it requires abandoning the priority of liberty principle.

4. A THIRD RESPONSE: ABANDONING THE PRIORITY OF LIBERTY PRINCIPLE

The most effective response to Flanigan's argument that Rawlsian high liberals might give is to abandon the priority of liberty principle and to thereby preserve their commitment to the common good. Since it is difficult to delimit the number of freedoms protected as basic rights without relying on value judgments, abandoning the priority principle means that it would be permissible for the state to interfere with these basic rights for the sake of promoting the common good. In this way, the high-liberal commitment to freedom need not entirely marginalize the commitment to the common good.

In addition, however, I want to suggest that there are compelling independent reasons for abandoning the priority principle and preserving the high-liberal commitment to the common good within the context of Flanigan's argument. Putting aside for the moment the state's ability to promote what are likely more controversial aspects of the common good, such as equality of opportunity or a fairer distribution of wealth, Flanigan's argument raises

of these areas of life. Owen's sports fandom is not implausibly characterized in such a way by Flanigan.

33 One might also wonder whether most philosophical or moral beliefs would even pass such stringent tests.

doubts that the state can promote a more basic aspect of the common good shared by citizens: establishing a compossible scheme of basic rights. Her argument also raises questions about how the state could adjudicate conflicts between basic rights when nearly all freedoms are basic. In this section, I argue that the concept of the common good will play an important role in addressing both problems.

4.1. What Is the State Permitted to Do?

Before I turn to the two problems highlighted above, I want to focus on the more general issue of what the state may do when it adopts Flanigan's argument. Thus far I have focused on the fact that her argument undermines the state's ability to promote political aspirations like equality of opportunity, a fairer distribution of wealth, public health, etc. While high liberals want to preserve the state's ability to promote these aspects of the common good, those working from competing political persuasions may not appreciate the value of doing so. However, Flanigan's argument threatens the state's ability to operate at a more fundamental level, and any political theory that can accommodate a more modest position for individual freedom in relation to the common good will possess the necessary flexibility to address both this issue and the other two problems mentioned above. Preserving the commitment to the common good thus reflects a distinct advantage of (Rawlsian) high liberalism in comparison to competing political theories.[34]

34 Left-libertarianism may present another alternative here. Left-libertarians, in contrast to right-libertarians or what I have referred to here as "standard libertarianism," couple a commitment to self-ownership with a commitment to a more egalitarian distribution of external resources. Generally speaking, the latter commitment is not defended by an explicit appeal to the common good, although there may be an indirect appeal to the common good lurking somewhere in the background. Determining to what extent left-libertarianism can address the problems I discuss and what sort of role the concept of the common good will play is beyond the scope of the current paper. Much depends on how left-libertarians construe the stringency of their commitment to a more egalitarian distribution of external resources and what exactly that commitment will look like in relation to the demands generated by their commitment to self-ownership (and there are questions about what self-ownership would look like under such a view). It is possible that left-libertarianism will turn out to be more accommodating to incorporating the concept of the common good within its framework. (Although as I suggest in note 41 below, right-libertarians may have reasons for incorporating the concept within their framework as well.) If these very general comments regarding both strands of libertarianism are correct, this could imply greater theoretical flexibility for both left and right versions of libertarianism, putting them relatively more on par with Rawlsian high liberalism, at least regarding this aspect. But these are very controversial points, and my focus is on how Rawlsian high liberals can respond to Flanigan's argument. I thank a reviewer for urging me to address these issues.

To support this claim, I turn to a well-known case considered by the United States Supreme Court: *Employment Division v. Smith*.[35] This case is noteworthy because it clearly demonstrates that the tension between individual freedom and the common good rears its head in a relatively more muted form (compared to Flanigan's argument) in virtue of the privileged protection afforded to freedom of religious worship. The latitude afforded to religious worshipers to imbue various practices, actions, etc. with sacramental value makes conflicts between religious worshipers and the interests of the state—i.e., the common good—inevitable. And given the inevitability of these conflicts in contemporary society, any political theory, whether liberalism or libertarianism, must consider how to handle them. The greater theoretical and political flexibility that comes from preserving a normative and political commitment to the common good thus remains a virtue even if Flanigan's argument were rejected; her argument, however, exacerbates the problem significantly and in doing so further accentuates the importance of that flexibility.

In this particular case, Alfred Smith and Galen Black were fired from their jobs with a private drug rehabilitation organization because they ingested peyote—a crime under Oregon state law—as part of a religious ceremony for their Native American church. They subsequently filed for unemployment compensation but were rejected because they had been fired for work-related misconduct. They argued that the denial of unemployment benefits violated their freedom of religious worship, which is protected by the First Amendment. In a six-to-three decision, the Supreme Court ruled in favor of the Department of Human Resources, arguing that the refusal to provide unemployment benefits did not violate Smith's and Black's free exercise rights. Without pretending to offer any judgments regarding the merits of the decision, I want to look briefly at the reasoning provided by the court in both the majority opinion (written by Justice Antonin Scalia) and in the concurring opinion (written by Justice Sandra Day O'Connor).

Justice Scalia offers two main lines of reasoning explaining the court's decision to uphold the nonprovision of employment benefits to Smith and Black. The first line of reasoning employs the distinction between neutrality of aim/intention and neutrality of effects to determine when a law unduly burdens religious worshippers: although a law may have incidental effects that burden (some) religious worshippers, as long as it was not designed with the intention of doing so, then compliance with the law takes priority over free exercise.[36]

35 *Employment Division, Department of Human Resources of Oregon v. Smith* 494 U.S. 872 (1990).

36 *Employment Division v. Smith*, 494 U.S. at 878–82. According to Justice Scalia, the only case in which a neutral law (neutral of aim/intention) could be judged as violating the First

The second line of reasoning offered by Justice Scalia is meant to support employing the neutrality distinction as the correct and *only* basis for determining the decision in this case. It is illuminating for my purposes here to understand why. Justice Scalia denies that the strict scrutiny standard employed in *Sherbert v. Verner* is relevant to the case at hand.[37] According to the strict scrutiny standard, state actions that burden or violate a constitutional protection such as freedom of religious worship are justifiable only by showing that those actions serve a compelling state interest and that they are either very narrowly tailored or the least restrictive actions open to the state. Strict scrutiny, in other words, places the presumption in favor of protecting constitutional freedoms, while it places the burden on the state to show that it has a compelling interest in interfering with those freedoms. Importantly, the strict scrutiny standard adopted by the court is considerably weaker than Rawls's priority principle insofar as it allows for the possibility of legitimately interfering with or violating a basic right for a compelling state interest. It thus represents a more modest alternative to Rawls's position, and it is the view adopted by Justice O'Connor in her concurring opinion in *Employment Division v. Smith*. There, she argues that the strict scrutiny standard must be applied to all "generally applicable laws that [have] the effect of significantly burdening a religious practice," regardless of whether the laws are neutral in aim/intention.[38]

In contrast, and precisely because strict scrutiny is so demanding as it places the burden on the state to justify its actions, Scalia argues that strict scrutiny applies only to a narrow range of circumstances; otherwise, the government would be held severely hostage to the religious beliefs of citizens:

> The government's ability to enforce generally applicable prohibitions of socially harmful conduct, like its ability to carry out other aspects of public policy, "cannot depend on measuring the effects of a governmental action on a religious objector's spiritual development."... To make an individual's obligation to obey such a law contingent upon the law's coincidence with his religious beliefs, except where the State's interest is "compelling"—permitting him, by virtue of his beliefs, "to become a law unto himself"—contradicts both constitutional tradition and common sense.[39]

Amendment is if the law conflicts with the free exercise clause *in conjunction* with other constitutional protections. He refers to these scenarios as "hybrid" ones.

37 *Sherbert v. Verner*, 374 U.S. 398 (1963).

38 *Employment Division v. Smith*, 494 U.S. at 894.

39 *Employment Division v. Smith*, 494 U.S. at 885.

If strict scrutiny were required in a case like this, then it would have to be applied to *all* state actions that infringed upon citizens' freedom of religious worship, no matter how unimportant—to avoid having the court determine which aspects of religious belief are important. Adopting such a demanding position would be, Scalia insists, "courting anarchy." If the state fails to demonstrate a compelling interest, this position would permit religious exemptions from all sorts of sundry civil obligations, including compulsory military service, the payment of taxes, health and safety regulations, child neglect laws, compulsory vaccination laws, drug laws, traffic laws, minimum wage laws, child labor laws, animal cruelty laws, environmental protection laws, and laws providing for equal opportunity in the context of race.[40] Thus, in rejecting the application of the strict scrutiny standard in this case, the majority adopts a position in which the presumption of protection holds in favor of the state's compelling interests over the protection of individual freedom.

My intention here is not to evaluate the soundness and accuracy of the reasoning offered by the court as it pertains to the unique circumstances of the case itself. What is important for my purposes is that that reasoning—found in the majority and minority opinions as well as in the dissenting opinion, which claims that the state failed to demonstrate a compelling state interest—shows that the court recognizes the necessity of adopting a less absolutist and more nuanced position regarding basic rights and freedoms than what is found in either Rawlsian high liberalism or standard libertarianism. Rather than simply adjudicating these inevitable conflicts in favor of individual freedom, as both those views would require, the court has adopted a more nuanced position because it has been more sensitive to the implications that follow from the valuational autonomy afforded to religious believers as part of the basic freedom of religious worship. With almost no check on this freedom, the court rightly recognizes that the state would become entirely hostage to the beliefs of religious worshippers.

Consider the situation in the context of Flanigan's argument, where the same valuational autonomy is extended to *any* citizen. Obviously, the state may not promote something like a fairer distribution of wealth. But what is it permitted to do? On the one hand, the state is still permitted to interfere with these freedoms for the sake of protecting direct harm to citizens, and libertarians at least would likely argue that nothing more than this is necessary.[41] Yet I argue

40 *Employment Division v. Smith*, 494 U.S. at 889.

41 Determining the status of actions that merely carry the risk of harming others is more complicated, however. Appealing to the common good may be helpful in determining which risky actions to permit and which to prohibit. If so, this could provide standard libertarians with a reason for incorporating the concept within their framework.

that even here, the state will have to rely on a conception of the harm principle shaped by considerations pertaining to the common good.

We can begin with Flanigan's own comments to see why. She appeals to Mill's harm principle to exclude life plans centered on directly harming other individuals, and this is why I have said throughout that her argument allows *nearly* any freedom to become basic. It is for this reason the examples she focuses on refer to freedoms that are relatively "self-regarding" and therefore seemingly innocuous. These freedoms include not only Owen's sports fandom but also Snoop Dogg's recreational drug use, helmetless motorcycling, and even gardening. In focusing on these examples, she conveniently sidesteps ambiguities inherent in Mill's harm principle. For the sake of space, I put to the side issues regarding Mill's distinction between self-regarding and other-regarding actions, and here I focus only on the idea of risk.[42]

Not all actions of the set of actions that harm others are the same. Some actions are clearly intended to harm others (whether they do so successfully or not); some harm is the product of culpable negligence; and some actions may merely increase the risk of harm to others.[43] Insofar as these actions are different, the reasons for prohibiting them will also differ. In the first case, the clear presence of the intent to harm justifies prohibiting any action of that kind—regardless of whether the action results in harm. In the second case, the harm must transpire as a consequence of culpable negligence on the part of an individual. In the third case of risky actions, these actions are like the first category insofar as it is not necessary for harm to occur; there is only some chance of harm. The two categories differ, however, in that risky actions do not involve the intention to inflict harm; the harm is rather an incidental side effect of the action. It is this last fact that makes the normative and legal status of risky actions more complex and ambiguous. For there are many actions that increase the risk of harm to others, but it is not self-evident whether these actions should be prohibited. The matter is further complicated by the fact that these actions may form integral parts of citizens' life plans, which, according to Flanigan's argument, would make them basic rights. Consider the example

42 Ben Saunders persuasively argues that a distinction between consensual and nonconsensual harm is more fundamental to Mill's harm principle. Yet this distinction raises in turn questions about what counts as consent and what counts as a harm. (If it were merely the fact that something is not consented to that makes it a harm, then this would be extremely broad.) See Saunders, "Reformulating Mill's Harm Principle."

43 These distinctions are not meant to be exhaustive, and it is possible to make finer distinctions, such as those made by the legal system in the United States. However, since my aim is only to distinguish between more paradigmatic examples of harm, such as physical assault, from risky actions, it is not necessary to delve into the finer details. I thank an anonymous reviewer for suggesting that I expand on these points.

of a group of adrenaline junkies whose life plans are centered on engaging in high-speed racing on busy streets. While the activity obviously increases the risk of harm to others, as the junkies defy all traffic laws, their intention is not to harm others: racing in busy streets adds an additional layer of challenge and therefore increases the level of skill needed. We could even say, in the same vein, that the mere act of driving increases the risk of harm to others.

Yet I suspect that Flanigan, as well as many liberals and libertarians, would want to prohibit things like high-speed street racing (and similarly, drunk driving) while allowing driving within a system of tightly defined traffic laws. The issue is the basis on which they can prohibit the former while permitting the latter. Given that these actions only increase the risk of harm without any accompanying intention to harm others, the justification for prohibiting them cannot be the intent to cause harm. After all, high-speed street racing is not the same as trying to run someone over with a vehicle, even though the same harm may transpire in both cases.

One way to distinguish between the prohibition of high-speed street racing and drunk driving from the nonprohibition of driving is by relying on socially acceptable thresholds of risk. Depending on the action in question, the severity of the potential harm, and the countervailing social interests promoted by prohibiting the action, society may be willing to tolerate lower or higher thresholds of risk for different actions. It may be (as a conjecture) that the risk of harm that comes about from driving—even within a system of tightly defined traffic laws—is still *higher* than the risk of harm in circumstances where either high-speed street racing or drunk driving is permissible. (We could imagine that the number of individuals engaging in these two activities exceeds their current number once they are made legally permissible, yet that number would still pale in comparison to the total number of drivers on the road, so we could imagine that the risk of being harmed from either of those two activities remains considerably lower than the risk of being harmed by other drivers merely driving about.) The severity of the harms in each case is comparable, yet as a society, we are willing to tolerate the (higher) risks associated with general driving, while we are not willing to tolerate the risks (even if they are lower) that come about from high-speed street racing and drunk driving. This is because allowing driving at all serves the common good. And in the very same way, it also serves the common good to ensure that the activity of driving is conducted within a system of tightly defined traffic laws, for without these laws in place, we would be, as Scalia puts it, "courting [vehicular] anarchy."

What this shows is that the valuational autonomy attributed to citizens as part of Flanigan's argument cannot always be sufficient for elevating freedoms to the status of basic rights. Even when those freedoms are integral to citizens'

life plans, determining their status will depend, at least when they carry the risk of harming others, on whether they promote or conflict with the common good.

4.2. Determining a Coherent, Compossible Scheme of Basic Rights

Although the previous subsection focused on the application of the harm principle only within the context of risky actions, the same insight generalizes to nonrisky actions. When citizens are capable of unilaterally elevating nearly any freedom to the status of basic rights, they will attempt to elevate freedoms that are often incompatible and that cannot all be protected under a unitary institutional apparatus. For instance, some citizens may have life plans (or fundamental philosophical views) that involve a refusal to pay taxes. Elevating this freedom may make it difficult for the state to enforce and protect other citizens' freedoms. Protecting corporations' right to speech may undermine the integrity of the political system and the fair value of political rights. Some citizens may desire to live under a capitalist framework and to exercise free market freedoms, whereas others may desire to live under a socialist framework and to exercise socialist economic freedoms.

While Flanigan does not devote attention to this problem, others have focused more directly on it. C. M. Melenovsky and Justin Bernstein, for instance, press this exact claim against Tomasi. They cite Rawls, who cites Isaiah Berlin in saying that there is no social world without loss—some (reasonable) life plans must and will lose out. It is for this reason that Melenovsky and Bernstein conclude that capitalist entrepreneurial freedoms may have to be excluded: "While it may seem intuitive that Amy should be able to open her own business, the fact that she cares deeply about it is not the appropriate ground to justify protection of rights that allow her to do so."[44] It is an unfortunate but inescapable limitation of the real world that even some reasonable and legitimate ways of life will be marginalized either by garnering weaker institutional protections or by being excluded from the social world altogether—in the sense that they are not afforded any institutional protections or prohibited entirely. Only by excluding some freedoms, the thought goes, can the list of basic rights be organized into a coherent, effective, and compossible scheme.

What we need to know is on what basis high liberals plan to determine which life plans fall outside the boundaries of the social world. Why do sports team fandoms and Amy's free market entrepreneurial freedoms lose out, but freedom of religious worship does not? For as we saw, Flanigan not implausibly describes Owen as willing to sacrifice both his religious freedom and his political freedoms far before he would sacrifice his ability to support the Green Bay Packers.

44 Melenovsky and Bernstein, "Why Free Market Rights Are Not Basic Liberties," 53.

As I already remarked upon the ineffectiveness of Rawls's solution to this problem in section 3.1, I will explain why the main argument made by Melenovsky and Bernstein is also unsatisfactory. As if they were responding directly to Flanigan, they argue that no *particular* life plan is an appropriate ground for justifying any basic right. Just a few sentences prior to making this point, they say that we protect religious freedom "because all would agree to guaranteeing the conditions necessary for the development and exercise of the two moral powers, and religious freedom guarantees those conditions."[45] Rawls also attempts to draw the same distinction.[46] The thought here seems to be that protecting religious freedom guarantees the social conditions necessary for the *moral powers* rather than for any *particular* life plan.

There is, however, no substantive distinction here that would serve as a basis for excluding some freedoms from the list of basic rights. In denying that a particular life plan is an appropriate ground for justifying any basic right, their concern rightly seems to be the sort of "bottom-up" approach suggested by Flanigan's argument. But Flanigan's argument is meant to exploit the logic internal to Rawls's view. On Rawls's view, religious freedom is necessary (or important—see note 9 above) for some citizens to develop their moral powers *because* these citizens have different yet particular life plans intimately connected with religion. The fact that citizens have these sorts of life plans makes it necessary to protect religious freedom. The justification for doing so is not that it allows them to become Christians or Muslims in particular but that it allows them to pursue their religious life plans. Since the justification itself is not tailored in such specific terms, then the right to religious freedom is also not specified so narrowly—in the sense that it permits citizens to practice only Christianity or only Islam, etc. But I fail to see how the examples mentioned by Tomasi or Flanigan are different in any relevant sense that makes them "more particular" in some problematic fashion. According to their views, capitalist entrepreneurial freedoms or sports fandoms are necessary (or important) for some citizens to develop their moral powers *because* these citizens have different yet particular life plans intimately connected with entrepreneurship or sports. Just as the justification for religious freedom is described in terms of allowing those citizens to pursue their religious life plans, the justification for these freedoms is the very same. Neither the justification nor the freedoms themselves is specified in such particular terms that citizens are permitted to run only certain forms of business or to support only certain teams within certain sports. It is not clear, in other words, in what sense following a religious life plan is any less "particular"

45 Melenovsky and Bernstein, "Why Free Market Rights Are Not Basic Liberties," 53.

46 Rawls, *Political Liberalism*, 10.

than following an entrepreneurial one or one dedicated to sports. In all three cases, these freedoms are necessary (or important) for individuals to develop and exercise their moral powers.

Notice too that it would not be satisfactory to make the argument that even without enshrining sports team fandom as a basic right, Owen is still free in a nonnormative sense to form a life plan centered around a sports team fandom. This response is problematic for several reasons. First, Owen's life plan would be subject to interference for the common good, whereas religious life plans would not be. This implies a diminished relative political and legal status and requires justification. Second, one could then make the exact same argument about being free in a nonnormative sense in the absence of formal freedoms with regards to religious freedom. Third, as both Melenovsky and Bernstein emphasize, basic rights are meant to protect the *social conditions* for the development and exercise of the moral powers. In other words, protecting these freedoms is not simply about protecting belief. As the Supreme Court notes in *Employment Division v. Smith*, the exercise of religion involves not only belief but also the *performance of physical acts*.[47] It is hard to see why other life plans should be treated any differently on this point. To underscore this last point, it is worth mentioning that when Rawls describes the moral power about life plans, he refers explicitly both to *forming* and to *pursuing* them.

Melenovsky and Bernstein's response thus runs into the same problems as the potential solutions canvassed earlier. In the absence of explicit value distinctions, appealing to vague criteria such as "particularity" or "significance" will fail, for such criteria will be overly inclusive unless applied arbitrarily and will leave us with the original dilemma as to which freedoms should be included and which should be excluded.[48] Given that these freedoms are necessary (or important) for the development of citizens' moral powers, we need some other basis for determining which freedoms to protect as basic rights. Appealing to the common good provides the grounds for solving this problem.

To see why, consider that while it is in the common interest of all citizens to have a list of basic rights in place, citizens will be divided over what sorts of rights and freedoms should be included within that list. Since there is no social world without loss, tough choices will have to be made, and since the freedoms among which we must choose all possess the correct modal connection to citizens' moral powers, these tough choices become more costly—the ability for some citizens to develop their moral powers will be affected negatively.

47 *Employment Division v. Smith*, 494 U.S. at 876.

48 See also Flanigan's responses to attempts made by high liberals to draw a distinction between freedoms on formal grounds: Flanigan, "All Liberty Is Basic," 460–62, 468–70.

Fortunately, some conceptions of the common good—such as the private individuality and distributive conceptions mentioned in section 1—are relatively more sensitive to this fact, as they permit appeals to aggregative considerations. Both theorists and legislators will need to rely on aggregative considerations as well as the importance of other goods outside of freedom that also constitute part of the common good to address this problem.

It is in fact quite common to see this sort of reasoning performing much of the argumentative work when it comes to debates about which freedoms should qualify as basic rights. Consider an argument that Samuel Freeman offers against elevating capitalist economic liberties to basic rights. Although he concedes that enshrining them as basic rights would allow *some* citizens whose life plans involve those freedoms to develop and exercise their moral powers, doing so would also make it difficult for the state to provide the *social minimum of goods* (i.e., goods outside of freedom) necessary for the adequate development and exercise of many other citizens' moral powers.[49] Flanigan's response to this argument offers the standard defense of free markets, which is to tout their ostensibly superior ability in promoting economic growth with all its subsidiary effects; she also argues that countries that best protect economic liberty do a better job of providing health care and social safety nets to their citizens—again, goods outside of freedom.[50] Not only do these arguments rely on aggregative considerations, but they also appeal to considerations not permitted by the priority principle and rely in the end on appeals to the common good—to the interests shared by all (or most) citizens in having access to goods *besides* freedom.[51]

Like these theorists, legislators will have to make hard decisions about which freedoms to include within the scheme of basic rights. In deciding whether to elevate freedom of contract to the status of a basic right (and potentially prohibiting mandatory minimum wage laws) or to protect employees' freedom to negotiate collectively, legislators will have to consider the social effects in each case, and they will have to determine whether protecting the former over the latter, or vice versa, would better promote the common good.

49 Freeman, *Rawls*, 395.

50 For the exchange between Freeman and Flanigan, see Flanigan, "All Liberty Is Basic," 463–64.

51 Samuel Arnold defends a similar view in his paper "Putting Liberty in Its Place," where he argues that goods besides freedom—he refers to goods such as opportunities for culture, leisure, education—also contribute to the development and exercise of the moral powers and should therefore be included within Rawls's first principle of justice. Arnold's argument thus rejects the priority principle, as his argument permits trade-offs between basic freedoms and these goods.

4.3. Adjudicating Conflicts Between Basic Rights

Once a scheme of basic rights has been established, conflicts between citizens' basic rights will still be inevitable. While we can organize the list of basic rights into a compossible set at an abstract level, in practice, citizens will exercise their rights in ways that obstruct or interfere with the ability of other citizens to exercise their rights. These conflicts are a consequence of the reasonable epistemic limitations on the part of theorists and legislators to articulate the contours of citizens' rights to avoid such conflicts altogether in the first place, as this would require predicting both every possible way each citizen might exercise their rights and all the possible effects those actions may have on the ability of other citizens to exercise their rights. These epistemic limitations are further compounded by the fact that the social, cultural, religious, economic, political, and technological aspects of the world are in continual flux and continually generating novel circumstances in which citizens must continue living their lives and exercising their rights. As these circumstances change, citizens' rights will be thrown against one another in unpredictable and conflicting ways, and Flanigan's argument will only exacerbate the likelihood of these conflicts.[52]

Once we recognize the likelihood of these conflicts and the practical necessity of adjudicating them, it becomes vital to develop conceptual and methodological tools for doing so. Jurists and legal philosophers have devoted considerable attention to this issue, and the primary result of this scholarship is the legal test of proportionality, which is employed as a necessary and sufficient test for adjudicating conflicts between basic rights.[53] One of the necessary components of this test—the component referred to as proportionality *stricto sensu* (balancing)—incorporates considerations that form part of the common good. In applying proportionality as a test for adjudicating these conflicts, it turns out that the common good plays an integral role in addressing this third problem.

52 I am therefore skeptical of specificationists who deny that these sorts of conflicts are possible once the rights in question have been articulated in specific detail. See Wellman, "On Conflicts Between Rights"; and Shafer-Landau, "Specifying Absolute Rights." It will be necessary, of course, to specify the rights to adjudicate conflicts when they do arise. But it is difficult to imagine that this process can be completed from the armchair beforehand, nor does it necessarily follow that once a right has been specified in more detail, the losing right no longer exists as a right at all. For an argument that rejects this latter point, see Waldron, "Rights in Conflict," 211–15. However these conceptual disputes are settled, the point remains that we would still need to know on what basis we may determine that one right should win out over another, and so my argument would still be relevant even if the specificationist thesis were true.

53 The literature on proportionality is quite extensive. For two thorough accounts, see Alexy, *A Theory of Constitutional Rights*; and Barak, *Proportionality*.

Before turning to an explanation of this test, it is necessary to note some qualifications concerning the following discussion. I do not intend to offer a full defense of proportionality here, as this would require determining whether basic rights are absolute and more thoroughly evaluating rival alternatives to proportionality, both of which are beyond the scope of this paper.[54] Instead, I want to highlight the role that the common good plays in the test—particularly in light of the fact that this test is firmly entrenched within the constitutional frameworks of many democratic countries.[55] Secondly, by contextualizing its application within the circumstances generated by Flanigan's argument, I hope to motivate its attractiveness. Appreciating its merits provides Rawlsian high liberals with yet another reason for abandoning the priority of liberty principle in favor of their commitment to the common good.

Jurists employ this evaluative test to determine whether the limitation of a basic right is justified or legitimate. Any statute or common law implemented for the sake of limiting a basic right must pass the test, whereas failure to do so renders the limiting statute or common law unconstitutional. Determining whether a statute or common law passes the test and is therefore proportional is a matter of applying its four components to the concrete circumstances at hand. These four components are: (1) proper purpose, (2) rational connection, (3) necessary means, and (4) proportionality *stricto sensu*. The first component, proper purpose, restricts the range over what sorts of reasons or purposes can limit basic rights. Although both constitutional theory and comparative law have recognized—in contrast to Rawlsian high liberalism—that the common good can serve as a proper purpose for limiting basic rights, the focus of our current discussion is on cases where one basic right is being limited for the sake of another.[56] The second and third components require that the means implemented by the limiting statute or common law are capable of sufficiently advancing the proper purpose while minimizing the extent to which the losing right is limited in comparison to alternative, rational means.

The fourth component—proportionality *stricto sensu*—is the most important. This component is inherently evaluative because it requires balancing the value of achieving the proper purpose on the one hand and the value of protecting the right on the other. Because this component involves an evaluative balancing act between conflicting values, it raises the question as to what sorts of considerations can be employed as a basis for rationally doing so—particularly

54 For a more complete defense of proportionality, see Barak, *Proportionality*, pt. IV, where he addresses both issues.

55 Barak, *Proportionality*, 132, 141.

56 On the issue of the common good serving as a proper purpose for limiting basic rights, see Barak, *Proportionality*, 249–59, 265–76.

when the two sides of the scale are values protected by rights of equal normative status. This balancing act involves a more consequentialist approach: applying it is a matter of evaluating the *social importance* of each right. Aharon Barak is careful to note, however, that this balancing procedure does not occur at an abstract level: it is not a matter of evaluating the general social importance of, e.g., religious freedom as opposed to sports team fandom.[57] The balancing that occurs at this judicial level is rather an evaluative judgment concerning the *marginal* social importance of protecting one right compared to the *marginal* social importance of protecting another right.[58] In a conflict between a concrete act protected by the right to religious freedom and a concrete act protected by the right to sports team fandom, proportionality allows courts to adjudicate this conflict by weighing the marginal social benefits and harms of protecting the former against the marginal social benefits and harms of protecting the latter.

Adjudicating conflicts between rights in this manner means that the common good will play an important role in the process. Proportionality requires that judges incorporate citizens' collective and aggregative interests in their decisions. Evaluating the marginal social benefits and harms between protecting one right over another will involve settling questions such as—but certainly not limited to: how many citizens will be affected directly in the decision to privilege one right over another; what effects the decision will have on citizens' ability to exercise other rights; and whether the decision will interfere with the state's ability to provide other common goods such as equality of opportunity, public health and safety, the integrity and fairness of the political system, etc. A distinct advantage of proportionality is that it allows judges to acknowledge the equal normative status of the rights in conflict but also to recognize that the marginal social importance of those rights can vary due to the unique features of a society and the peculiar details of the conflict itself.[59]

Rawls believed that he could solve the problems outlined in an *a priori* fashion by appealing to his conception of personhood. He thought it possible to distinguish the significance of free speech and political speech in contrast to commercial speech, for instance, by delineating the connection between the former and citizens' moral powers. Flanigan's argument casts doubt over the viability of Rawls's solution, as she argues that this connection holds for nearly any freedom. All freedom is basic, according to her. But this also means that (nearly) all freedom is of equal normative status. When we have to choose

57 This more abstract balancing would be more appropriate when it comes to determining which rights should be included within the list of basic rights in the first place, as we saw in section 4.2 above.

58 Barak, *Proportionality*, 349–52.

59 Barak, *Proportionality*, 359.

which values, freedoms, or rights to privilege or prioritize, the common good is what allows us to do so. It helps us to determine which rights should be included in a compossible scheme of rights, to adjudicate conflicts between those rights, to balance between values that are part of the common good—such as preserving the integrity of the political system versus protecting the speech of corporations—and to decide which risky behaviors to limit. It is difficult to see, in the context of Flanigan's argument, how these theoretical and practical political problems could be addressed without appealing to the common good. For these reasons then, Rawlsian high liberals ought to preserve their commitment to the common good and abandon the priority of liberty principle.

5. AN OBJECTION: PRESERVING THE PRIORITY PRINCIPLE

Since the priority principle permits the restriction of some basic rights for the purpose of protecting other basic rights, one might argue that my comments about how conflicts between rights are to be adjudicated are therefore consistent with it. The objection implies that the priority principle is satisfied if and only if our intention behind violating one basic right is the protection of another. But as mentioned in section 1, the priority principle also generates constraints over what sorts of *reasons* can be appealed to in order to justify the violation of a basic right.

The criterion set out by Rawls in *A Theory of Justice* suggests that the only permissible basis would be the quantity or extent of freedom. An example he employs is that it is permissible to limit freedom of speech (specifically, the right to interrupt speech) for the sake of promoting freedom of speech (the right to engage in discussion).[60] Limiting the former opens up freedom of speech to a greater extent. Hart finds this argument problematic, however. The central point behind his criticism of Rawls's view in *Theory* is that it attempts to resolve all conflicts in a purely formal or quantitative manner. Hart thinks that this method would prove effective only in the simplest cases (like Rawls's speech example) and that the resolution of many other conflicts would have to rely on considerations beyond the extent of freedom—considerations such as the value of the freedoms in conflict or the various social effects that would come about from protecting or privileging one freedom over another (similar to the proportionality test).[61]

60 Rawls, *A Theory of Justice*, 203.

61 Hart, "Rawls on Liberty and Its Priority," 542–47.

Rawls responds to this criticism in a number of interrelated ways.[62] The first step is to reject the notion that the list of basic rights is meant to maximize the development and exercise of the two moral powers. Instead, the basic rights are meant to guarantee, according to Rawls, the social conditions essential for the *adequate* development and exercise of the two moral powers. Once Rawls adopts the notion of a fully adequate scheme instead of the most extensive scheme of basic liberties, the second step of his response to Hart, as we have already seen, is to appeal to the idea of the "central range of application," which specifies the significance of concrete exercises of freedom by tying those freedoms to the adequate development and exercise of the two moral powers. Those concrete exercises of freedom that fall outside the central range of application would not merit status as basic rights.

But it is unclear how these changes address the main point behind Hart's criticism, which is only exacerbated with the addition of Flanigan's argument. Again, the first problem is that there are so many freedoms that satisfy the above criterion (of adequacy) that it is not possible to organize them all into a coherent, compossible scheme. As a consequence, it is necessary to exclude some of those freedoms from the list of basic rights, despite the fact that they possess the right modal connection to the moral powers. The question, to repeat it once more, is on what grounds theorists and legislators can exclude those freedoms. The next problem—adjudicating conflicts between basic rights—are conflicts between rights that *do* possess the proper connection to the moral powers. The question here too is on what grounds theorists and judges can then adjudicate these conflicts. In both cases, it appears that theorists, legislators, and judges will have to rely on the sorts of considerations highlighted by Hart—considerations either about the value of one freedom over another or about the broader social effects of protecting (or prohibiting) different freedoms.

Consider once more the choice confronting theorists and legislators as to whether capitalist economic liberties should be excluded from the list of basic rights despite their centrality to the life plans of at least some citizens. This example certainly differs from Rawls's example of freedom of speech. To note one important difference, while all citizens may concede the rationality of limiting the freedom to interrupt speech for the sake of promoting greater freedom of speech, rationality would not dictate the same consensus in the case of limiting the exercise of capitalist economic liberties. Whatever it would require in this case would depend, as Hart rightly points out, on the value associated with those freedoms for different citizens. It is difficult to see why rationality would dictate accepting limitations on or the prohibition of capitalist economic

62 Rawls, *Political Liberalism*, 331–40.

freedoms for citizens with life plans depending on these freedoms for the sake of promoting either an adequate level of socialist economic liberties or freedom of religious belief. Appealing to the adequacy criterion fails to resolve this issue.

Once we appeal to the common good as a basis for resolving these problems, we have moved beyond the sorts of considerations permitted by the priority principle. Our choice as to which freedoms should count as basic—when nearly any freedom possesses the right sort of connection to the moral powers—will depend on what best promotes the common good. And while it is also true that we are violating one basic right for the sake of protecting another basic right in accordance with the priority principle, our choice over which basic right to protect will also depend on what we ultimately think best promotes the common good.

6. CONCLUSION

Flanigan presses an important argument against Rawlsian high liberalism: she argues that the logic internal to the high-liberal position requires a significant expansion of the freedoms protected by Rawls's first principle of justice. In making this argument, not only does she place the high-liberal commitment to the common good on precarious footing, but she also reduces the conceptual space between Rawlsian high liberalism and standard versions of libertarianism, insofar as the overriding if not exclusive function of the state becomes the protection of individual freedom. I have considered several potential responses that high liberals could make, and I have argued that the strongest response would be for high liberals to abandon the priority of liberty principle. In doing so, Rawlsian high liberals will have to make greater normative space for considerations pertaining to the common good within their theoretical framework.[63]

Tufts University
seena.eftekhari@tufts.edu

REFERENCES

Alexy, Robert. *A Theory of Constitutional Rights.* Translated by Julian Rivers. Oxford University Press, 2002.

Arnold, Samuel. "Putting Liberty in Its Place: Rawlsian Liberalism Without the

63 I would like to thank Kurt Blankschaen and three anonymous reviewers for helpful feedback on previous drafts.

Liberalism." *European Journal of Philosophy* 26, no. 1 (2017): 213–37.
———. "Right-Wing Rawlsianism: A Critique." *Journal of Political Philosophy* 21, no. 4 (2013): 382–404.
Barak, Aharon. *Proportionality: Constitutional Rights and Their Limitations.* Cambridge University Press, 2012.
Beiner, Ronald. *Civil Religion: A Dialogue in the History of Political Philosophy.* Cambridge University Press, 2011.
Flanigan, Jessica. "All Liberty Is Basic." *Res Publica* 24, no. 4 (2018): 455–74.
Freeman, Samuel. "Capitalism in the Classical and High Liberal Traditions." *Social Philosophy and Policy* 28, no. 2 (2011): 19–55.
———. *Rawls*. New ed. New York: Routledge, 2007.
Gaus, Gerald. *The Modern Liberal Theory of Man.* St. Martin's Press, 1983.
Hart, H. L. A. "Rawls on Liberty and Its Priority." *University of Chicago Law Review* 40, no. 3 (1973): 534–55.
Hussain, Waheed. "The Common Good." *Stanford Encyclopedia of Philosophy* (Spring 2018). https://plato.stanford.edu/entries/common-good/.
Melenovsky, C. M., and Justin Bernstein, "Why Free Market Rights Are Not Basic Liberties." *Journal of Value Inquiry* 49, nos. 1–2 (2015): 47–67.
Patten, Alan. "Are the Economic Liberties Basic?" *Critical Review* 26, nos. 3–4 (2014): 362–74.
Phelan, Sean, and Simon Dawes. "Liberalism and Neoliberalism." *Oxford Research Encyclopedia of Communication* 26 (2018): 1–34.
Rawls, John. *Justice as Fairness.* Edited by Erin Kelly. Harvard University Press, 2001.
———. *Political Liberalism.* Columbia University Press, 1993.
———. *A Theory of Justice.* Clarendon Press, 1971.
———. *A Theory of Justice*. Rev. ed. Harvard University Press, 1999.
Saunders, Ben. "Reformulating Mill's Harm Principle." *Mind* 125, no. 500 (2016): 1005–32.
Shafer-Landau, Russ. "Specifying Absolute Rights." *Arizona Law Review* 37, no. 209 (1995): 209–25.
Taylor, Robert. "Rawls's Defense of the Priority of Liberty: A Kantian Reconstruction." *Philosophy and Public Affairs* 31, no. 3 (2003): 246–71.
Tomasi, John. *Free Market Fairness.* Princeton University Press, 2012.
Waldron, Jeremy. "Rights in Conflict." In *Liberal Rights: Collected Papers, 1981–1991.* Cambridge University Press, 1993.
Wellman, Christopher Heath. "On Conflicts Between Rights." *Law and Philosophy* 14, nos. 3–4 (1995): 271–95.

Journal of Ethics and Social Philosophy
VOL.21, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.3590

IN DEFENSE OF CLAIM RIGHTS

Michael Da Silva

Claim rights are defined in terms of correlative duties such that "*S* has a claim right against *T* that *T* ϕ iff *T* owes *S* a duty to ϕ."[1] The claim right model of rights suggests claim rights so defined mark a distinct phenomenon encompassing paradigmatic moral rights (where moral rights establish distinct entitlements that exist independent of legal recognition).[2] Claim rights apparently come "closest to capturing the concept of individual rights used in political morality," creating a "consensus" that claim rights are the "core" instances of rights.[3] One can thus use requirements for establishing claim rights, including correlativity, to evaluate philosophical uses of 'right'. For example, Sreenivasan suggests purported health rights are noncorrelative, and appeals thereto thus violate philosophical strictures on apt usage.[4]

This note defends the claim right model against recent criticisms that suggest any plausible specification thereof will prove (i) extensionally or explanatorily inadequate or (ii) unable to serve a distinct normative purpose intended by those invoking rights.[5] It argues that an accurate understanding of the model's purpose and explanatory and extensional targets defuses this purported dilemma. Claim rights can serve their intended taxonomic function, thereby making a distinct contribution to morality, while fulfilling apt explanatory/extensional desiderata.

1 Jonker, "Rights, Abstraction, and Correlativity," 122.

2 The phrasing of "entitlement" here is adopted from the basic formulation in Wenar, "Rights." Valentini discusses rights generally in terms of status ("Rethinking Moral Claim Rights"). I do not do so initially in order to avoid question-begging charges. Hohfeld's canonical account of correlativity was law specific ("Some Fundamental Legal Conceptions as Applied in Judicial Reasoning"). This work contributes to ongoing conversations about moral rights' claimed correlativity.

3 The former quote is in Waldron, *Theories of Rights*, 8. The latter is in Valentini, "Rethinking Moral Claim Rights," 433.

4 Sreenivasan, "A Human Right to Health?"

5 Past critiques focused primarily on extensional adequacy. See, e.g., Raz, "On the Nature of Rights"; MacCormick, "Rights in Legislation"; and Perry, "Correlativity." This recently proposed dilemma merits a distinct response.

1. THE CHALLENGE

Proponents and critics of the claim right model agree that a justified model should explain the contribution of claim rights to moral ontology, paradigmatic uses of rights language, and intuitions about when rights violations occur. Explanations should maintain a core of meaning that permits the term 'rights' to serve its intended normative function(s).[6] However, recent critics pose an apparent dilemma for proponents of the claim rights model. Any account of claim rights will, they argue, be extensionally or explanatorily inadequate or fail to identify the distinct normative role for claim rights intended by model proponents. Valentini identifies the general dilemma. Kahn's structurally similar work applies it to a key use case, human rights theory.[7]

The first lemma suggests any distinct articulation of claim rights over- or undergenerates rights or cannot account for paradigmatic invocations of rights. Kahn, for instance, suggests the claim right model cannot account for recognized human rights to education, health care, etc. whose fulfillment requires "coordinated action."[8] If they are interpersonal rights, it is difficult to identify who should hold corresponding duties. And appeals to governmental duties cannot establish "universal rights" since many governments lack the capacity to fulfill correlative rights.[9] This apparent explanatory failing implicates extensional adequacy as these cases purportedly exemplify human rights' intended purpose. Rights should signify "particularly significant, normative requirements of universal concern that should be met for individuals everywhere, and which should take priority over most other" concerns.[10] However, the claim right model renders education, health care, etc., into lower-priority goods without adequate reason. It cannot explain why correlative rights should

6 Compare, e.g., Valentini's distinctive moral position and consistency conditions ("Rethinking Moral Claim Rights"), Kahn's explanatory condition ("Beyond Claim-Rights"), my distinctiveness and action-guidingness conditions (Da Silva, "Correlativity and the Case Against a Common Presumption About the Structure of Rights"), and Jonker's requirements for an explanatory action-guiding concept that is not "irredeemably fragmented" ("Rights, Abstraction, and Correlativity"). I draw on each recent author, including my own prior work, below.

7 Kahn, "Beyond Claim-Rights" cites earlier Valentini (namely, "In What Sense Are Human Rights Political?") to highlight explanatory/extensional failings.

8 Kahn, "Beyond Claim-Rights," 162–63, 167.

9 See also, e.g., O'Neill, "The Dark Side of Human Rights"; and Sreenivasan, "Duties and Their Direction" and "A Human Right to Health?" Compare, e.g., Etinson, "Human Rights, Claimability and the Uses of Abstraction."

10 Kahn, "Beyond Claim-Rights," 172.

have priority over other "basic requirements of social justice," including those requiring collective responses.[11]

The second lemma suggests *any* explanatorily/extensionally adequate theory cannot maintain a distinctive normative purpose for claim rights. Kahn, for example, rejects correlativity and proposes discussing relevant human rights–related interests in terms of "*pro tanto* collectivization duties" requiring all persons to "make considerable efforts to achieve and maintain a sociopolitical order in which they are socially guaranteed for everyone."[12] However, those duties alone cannot maintain human rights as a unique moral contribution or explain many acts done in the name of rights. If human rights are equivalent to a broader range of social justice demands, their contribution to morality becomes obscure. There is a risk of a category collapse without clear corresponding explanatory gains that could warrant accepting that risk.

While one could reject Kahn's explanatory/extensional targets, Valentini contends that the basic dilemma is inevitable, as the claim right model's failings are structural: the term 'claim rights' does not denote a "distinct moral position … but a family thereof."[13] Per Valentini, no theory can account for the variety in "paradigmatic rights talk" and maintain both correlativity and a distinct moral position for claim rights.[14] One must either accept that claim rights cannot serve their intended normative role(s) or expand the concept until it is no longer distinct. I further detail Valentini's general version of the critique exemplified by Kahn when evaluating both below.

2. DEFENDING CLAIM RIGHTS

The claim right model can survive this challenge. The apparent dilemma rests on uncharitable or mistaken conceptions of relevant conceptual desiderata and the claim right model. The claim right model addresses the "taxonomical dimension" of rights theory concerning their relationship "to other normative phenomena."[15] This is distinct from the "explanatory dimension" regarding "what generally explains or grounds particular rights."[16] The model, then, purports to distinguish rights and claim rights from other moral phenomena like

11 Kahn, "Beyond Claim-Rights," 163.

12 Kahn, "Beyond Claim-Rights," 162.

13 Valentini, "Rethinking Moral Claim Rights," 434.

14 Valentini, "Rethinking Moral Claim Rights," 435.

15 Jonker, "Rights, Abstraction, and Correlativity," 122.

16 Jonker, "Rights, Abstraction, and Correlativity," 122. See also the distinction between the "form" and "function" of rights in Wenar, "Rights." Hohfeld, "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning," speaks to form.

justice. A claim right model specification should identify a common structure across use cases without categorically barring many seemingly licit cases if it is to play its intended taxonomic role. Yet arguments that no specification could account for paradigmatic rights talk rely on questionable claims about paradigm cases and, in Valentini's case, on a failure to recognize that a primarily taxonomic model does not aim or need to account for nontaxonomic discourse about rights' justification. Clarifying basic conceptual desiderata diffuses any apparent dilemma, and a standard view on claim rights can meet both the challenges connecting Valentini and Kahn and the critique they exemplify.

Understanding claim rights in light of the distinctive moral standing they provide helps fulfill properly articulated desiderata. Consider a view on which *S*'s claim right against *T* entails a (defeasible) duty to *S* to ϕ such that *T* wrongs *S* specifically by failing to ϕ and owes a second-order duty of explanation or compensation β where *T* fails to ϕ. *T*'s failure to β wrongs *S*. The initial duty being owed to the rights holder on the standard claim right model marks it as a *directed* duty. Duty bearers have specific duties to rights claimants *qua* rights claimants. This establishes a relationship between those parties providing the rights holder with a form of moral standing that makes a distinct contribution to moral ontology common to most rights talk. This is so even if accounting for facially apt talk requires calibrating rights for specific contexts. The focus on the correlative relationship between rights and directed duties is common to many accounts of claim rights, including Hohfeld's original specification.[17] The focus on second-order duties may be less standard (and potentially non-Hohfeldian) but maintains the structure of paradigmatic moral rights claims and articulates plausible implications of rights-based moral standing. The combined directed first-order and second-order duties likely identify a moral concept that can fulfill the desiderata above. However, the account is mere proof of concept for a general argumentative strategy. My defense of claim rights succeeds if this specification is implausible.

2.1. Fixing the Explanatory and Extensional Targets

The proposed dilemma fundamentally rests on a misunderstanding of relevant conceptual desiderata. The criticisms first mischaracterize their explanatory/extensional targets. Neither the practice or discourse of rights nor the nature of the claim rights model requires that rights encompass the highest-priority moral goods or explain all justificatory claims. And a well-calibrated model can address concerns with collective rights claims undergirding both critiques.

17 Hohfeld, "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning." On directed duties, see also Sreenivasan, "Duties and Their Direction" and related views in Jonker, "Rights, Abstraction, and Correlativity."

Kahn exemplifies this issue. Kahn believes the Universal Declaration of Human Rights (UDHR) catalogues paradigmatic moral rights and rights discourse for which any theory of rights must account.[18] The UDHR then marks human rights as the highest-priority moral goods and recognizes collective rights that cannot easily fit the claim right model. The claim right model thus cannot account for paradigmatically high-priority human rights and implausibly prioritizes "requirements of justice" that maintain correlativity over others that require collective responses.

With respect, however, Kahn's account of the intended purposes and paradigmatic instances of human rights is at best undermotivated. Kahn rejects contentions that human rights need not entail highest moral priority because they conflict with human rights practice.[19] Yet this move largely rests on references to the UDHR and individual statements by one United Nations body and by Amnesty International, respectively.[20] Kahn offers no independent normative reasons why rights or human rights should denote goods that must be categorically prioritized. Other core elements of human rights practice suggest that human rights do not have this categorical priority.[21] International legal rights admit many exceptions. Some need not be fulfilled immediately, as Kahn admits. The very social rights Kahn takes as central to her account are subject to a doctrine of "progressive realization" whereby states need to provide access to only a "minimum core" of goods immediately and then "take steps" to improve access over time. While Kahn suggests this problematically creates a two-tier system of rights, it is part of the "practice."[22] A principle requiring that a theory of rights account for the UDHR but not progressive realization is lacking.

While Kahn could alternatively reject appeals to (human) rights practice and simply seek to explain the existence of moral rights requiring a collective response, Kahn cannot account for the particular form of standing that even Valentini recognizes as core to rights talk. Rights typically aim to provide particular persons with distinct claims to the objects of their rights. It remains difficult to see how any individual has distinct standing to state, "The state uniquely wrongs me when it fails to establish a public health program."[23] Kahn thus risks collapsing relational and broadly structural claims: rights talk seeks

18 Kahn, "Beyond Claim-Rights," 162–63.

19 Kahn, "Beyond Claim-Rights," 168.

20 Kahn, "Beyond Claim-Rights," 162, 164.

21 Da Silva, "Correlativity and the Case Against a Common Presumption About the Structure of Rights."

22 For the two-tier worry, see Kahn, "Beyond Claim-Rights," 170.

23 See also Sreenivasan, "A Human Right to Health?"

to establish particular moral relations between specifiable parties, distinct from (admittedly related) claims to justice or just institutions.

A claim right model can, moreover, address many collective concerns motivating Kahn and Valentini while maintaining rights as a distinct moral phenomenon. Collective duty bearers could, for instance, fulfill a right to education.[24] Weaker forms of correlativity permitting multiple entities to fulfill the duty-bearer role or multiple options for means of fulfilling rights then specify broader ranges of prospective duty bearers and potential wrongs while maintaining rights holders' distinct individualized grounds for complaint/explanation/compensation. Claim rights so understood can also require structural change. Fulfilling directed duties plausibly correlative to social rights commonly requires collective action. If a right to vaccinations must be effected through governmental public health programs, realizing it necessitates system-level changes. The right remains a particularized claim to a vaccine, distinguishing it from other calls for change. Stating "I am uniquely wronged when I cannot access a vaccine necessary to safeguard basic health" is plausible even when avoiding that wrong would require collective action. Familiar injunctions to attend to the structural conditions of rights fulfillment need not entail direct rights *to that structure*.[25] Calls to effect a valid entitlement to a particular good for a specific person still denote a distinct phenomenon.[26] 'Claim rights' remains an apt descriptor.

These challenges exemplify a general problem: critics of claim rights often understandably but problematically mischaracterize their analytical target. Valentini further suggests that no characterization of rights can explain the way in which 'rights' refers to both justification statements concerning "moral reasons . . . to empower individuals" and status statements concerning the "empowered status individuals enjoy" as rights holders.[27] Only concerns with empowerment that are not distinct from claim rights explain both. This undergirds Valentini's

24 Kahn rejects this contention using the arguments about progressive realization that are rejected above/below ("Beyond Claim-Rights," 175).

25 Compare, for example, Ashford, "The Inadequacy of Our Traditional Conception of the Duties Imposed by Human Rights"; Pogge, *World Poverty and Human Rights*; and even the works by Valentini cited above.

26 Ashford's work accordingly may not support Kahn as claimed (*contra* Kahn, "Beyond Claim-Rights," 180n6). An anonymous reviewer suggests that the appeal to the "dynamic" aspect of rights in Raz also helps address this concern (Raz, "On the Nature of Rights," 200, 212). Rights can create new duties on the classic Razian scheme. This could plausibly entail new duties to address structural concerns. Rights should, moreover, be interpreted in particular social contexts. Those contexts can impact what rights bearers must do. The point here is distinct from but related to Raz's position.

27 Valentini, "Rethinking Moral Claim Rights," 438–40.

structural critique of the claim right model. Per Valentini, model proponents falsely assume that status and justification co-occur. For instance, children's claim rights to education aim to empower adults based on children's justificatory interests. If this is so, Valentini further argues, there are two options for how to proceed. One is to recognize that 'claim rights' denotes a family resemblance concept that "is not much more specific … than the broader notion of a right," which a distinct concept should disambiguate. Valentini believes that this option undermines the claimed distinct moral role of claim rights and so falls on one horn of the dilemma. The other option is to adopt an ad hoc account of "central" claim rights. Valentini believe this option unduly limits their extension and so falls on the other horn.[28]

Valentini's critique also mischaracterizes the operative desiderata for a theory of rights. Status and justification claims speak to different elements of rights and play different roles in rights discourse. The status-focused claim right model speaks to the taxonomic element, not to the justification-based explanatory element. Taxonomic and explanatory elements need not submit to a common conceptual explanation or schema. And otherwise distinct components of rights discourse need not be linked via a common interest in concepts like empowerment. A taxonomic theory disconnected from potential justifications would be problematic. But the claim right model lacks that defect. Indeed, the taxonomy/justification distinction further highlights why social rights, like the right to education in Valentini, need not undermine the model. On my proposal, one need explain only why a "child's right to education" would provide a noneducated child with specific standing to claim that a specifiable agent who could have secured access to education wronged them. Interests that would ground that claim are likely to change adults' powers, explaining why many assume a connection between status and justification claims. Yet the nonoccurrence of status and justification is unproblematic. The taxonomic element of rights focuses only on the status conferred upon claim right holders.

2.2. Maintaining Moral Distinctiveness

Understanding rights as conferring a particular kind of standing also identifies their distinct moral contribution. Claim rights confer a distinct form of "standing to claim the direct object of the right."[29] Rights holders can validly seek further explanatory or compensatory redress when the claim is unfulfilled.[30] This

28 Valentini, "Rethinking Moral Claim Rights," 443.

29 Jonker, "Rights, Abstraction, and Correlativity," 125.

30 A reviewer notes that infant human and nonhuman rights holders present challenges. Appeals to proxies/advocates address many challenges, as the reviewer notes, and some

basic concern and its attendant commitment to correlativity between the rights bearer claimant and the duty bearer addressee are common to claims across distinct normative domains, helping explain the diversity of rights claims without also over- or undergenerating rights. Where claim rights can be calibrated for distinct domains, the model retains plausibility.

A status-focused understanding of claim rights underlines the relational nature of characteristic rights claims and the distinct contribution to morality instantiated by such relations. A successful rights claimant must always identify what must be done to avoid wronging the claimant and the moral reasons why at least one of a specifiable set of persons are bound to do it for that claimant specifically. *Contra* Valentini, focus on particularized relations between specified parties and the ability of right holders to make particularized claims on duty bearers distinguishes claim rights from liberties and other Hohfeldian powers, permitting claim rights to serve their intended taxonomic function. Even if all rights confer standing, as Valentini contends, claim rights so defined confer a particular kind of standing with a distinct form. Claim right holders have specific standing to call on particular duty bearers to perform duties owed to them alone—and could have standing to demand explanation/compensation *from that person* for nonperformance.

Claim rights so defined further help differentiate the right and the good. One can, e.g., distinguish universal health care programs as good policies and as potential objects of rights. Policies can be justified (or even required) without being objects of rights. The proposed view acknowledges the right/good distinction without making assumptions about whether good policies can be valid objects of rights. It instead sets burdens for further work. Rights-based claims require explaining how failure to create such programs provides *particularized* grounds for complaint against those who can but do not create them. This is a distinct kind of moral argument made by real persons.[31] The claim right model permits evaluating it on its own terms.

Claim rights' variety on this model underlines the model's explanatory and extensional value rather than establishing problematic ambiguity. Distinctions between, for example, positive and negative rights or between perfect and imperfect duties, make it difficult to specify a common form of correlativity. If the claim right model cannot explain why many believe positive rights or imperfect duties are characteristic of core moral rights, this could undermine the model. Kahn and Valentini's shared interest in positive social rights to

claimants (e.g., rivers) may not have moral rights. I cannot resolve these problems here but likely need not where they apply to most theories of rights.

31 Compare Hassoun, "The Human Right to Health"; and Rumbold, "The Moral Right to Health."

education and health is thus notable. However, plausible calibrations of the model can and do account for the variety. Even "multifarious" claim rights are not irremediably fragmented, defusing ambiguity-related concerns. They nonetheless present sufficient conditions on valid invocations of rights, avoiding overgeneralization concerns.

While Valentini suggests 'claim right' cannot refer to a "family" of concepts if it is to disambiguate 'rights' and avoid moral confusion, a multifarious concept can be distinct if its paradigmatic forms share a common structure and implications and differ only in application(s). In previous work, for example, I challenged "strong correlativity" that requires a unique duty bearer who bears a specific duty because it cannot account for social rights.[32] However, I further demonstrated that this need not entail that rights lack a common structure. Rights in private and public law each feature a class of specifiable individuals who could be duty bearers and a set of acts from which they can choose to perform their duties. Even positive rights triggering imperfect duties thus maintain a kind of correlativity that is characteristic of the proposed claim rights model. They feature particular persons holding valid claims against others whose nonfulfillment creates second-order duties of explanation or compensation. Correlativity may not be identical across all domains, but it shares a basic form and impacts moral powers in the same ways. There is always a rights bearer, a duty bearer, and the prospect of second-order duties when claims are unfulfilled.

Jonker further suggests claim rights generally share a structure and purpose but vary in application. Per Jonker, claim rights of any kind, legal or moral, public or private, etc., always involve a relationship between rights and directed duties. Putative counterexamples to correlativity fail to recognize the diverse forms that this relationship can take. Rights and duties differ in their specificity and generality, creating "degrees of abstraction when it comes to rights and other entitlements: general (as opposed to particular) entitlements, unspecified (as opposed to specified) entitlements, and indefinite (as opposed to definite) entitlements."[33] A right at any level of abstraction will have a duty at the same level, maintaining the correlative form. But a general right to health need not entail a specific duty to provide a particular pill. One must "calibrate" the model for particular contexts and understand the level of abstraction applying in each. A claimed "right to insulin" requires one to establish a more specific duty than a right to health, which admits more candidate duty bearers and

32 Da Silva, "Correlativity and the Case Against a Common Presumption About the Structure of Rights."

33 Jonker, "Rights, Abstraction, and Correlativity," 147.

duties. The latter may, I add, only require "taking steps" towards an outcome without violating correlativity. Claim rights thus remain distinct.

This defense of claim rights does not beg questions about whether rights confer relevant status. Rather, correlative standing identifies a distinct normative position that explains relevant phenomena and clarifies the taxonomy of normative concepts. It now further provides a framework for debates about apt use. One can judge appeals to a purported right to health care by assessing whether any normative reasons can ground a valid claim fitting the correlative form.

3. Conclusion

Clarifying the targets for theoretical adequacy as well as the nature and intended role of claim rights and their constituent correlativity defuses the apparent dilemma for the claim right model. The shared form and function of paradigmatic right/duty pairs remain notable even when correlativity must be calibrated for specific contexts. Standing-focused claim rights can play their intended taxonomic role while fulfilling well-defined explanatory and extensional conceptual desiderata.

University of Southampton
m.da-silva@soton.ac.uk

References

Ashford, Elizabeth. "The Inadequacy of Our Traditional Conception of the Duties Imposed by Human Rights." *Canadian Journal of Law and Jurisprudence* 19, no. 2 (2006): 217–35.

Da Silva, Michael. "Correlativity and the Case Against a Common Presumption About the Structure of Rights." *Journal of Value Inquiry* 54, no. 2 (2020): 289–307.

Etinson, Adam. "Human Rights, Claimability and the Uses of Abstraction." *Utilitas* 25, no. 4 (2013): 463–86.

Hassoun, Nicole. "The Human Right to Health." *Philosophy Compass* 10, no. 4 (2015): 275–83.

Hohfeld, Wesley Newcomb. "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning." *Yale Law Journal* 23, no. 1 (1913): 16–59.

Jonker, Julian David. "Rights, Abstraction, and Correlativity." *Legal Theory* 29, no. 2 (2023): 122–50.

Kahn, Elizabeth. "Beyond Claim-Rights: Social Structure, Collectivization, and Human Rights." *Journal of Social Philosophy* 52, no. 2 (2021): 162–84.

MacCormick, Neil. "Rights in Legislation." In *Law, Morality and Society: Essays in Honour of H. L. A. Hart*, edited by P. M. S. Hacker and Joseph Raz. Oxford University Press, 1977.

O'Neill, Onora. "The Dark Side of Human Rights." *International Affairs* 81, no. 2 (2005): 427–39.

Perry, Ronen. "Correlativity." *Law and Philosophy* 28, no. 6 (2009): 537–84.

Pogge, Thomas. *World Poverty and Human Rights*, 2nd ed. Polity Press, 2008.

Raz, Joseph. "On the Nature of Rights." *Mind* 93, no. 370 (1984): 194–214.

Rumbold, Benedict E. "The Moral Right to Health: A Survey of Available Conceptions." *Critical Review of International Social and Political Philosophy* 20, no. 4 (2017): 508–28.

Sreenivasan, Gopal. "Duties and Their Direction." *Ethics* 120, no. 3 (2010): 465–94.

———. "A Human Right to Health? Some Inconclusive Scepticism." *Aristotelian Society Supplementary Volume* 86, no. 1 (2012): 239–65.

Valentini, Laura. "In What Sense Are Human Rights Political? A Preliminary Exploration." *Political Studies* 60, no. 1 (2012): 180–94.

———. "Rethinking Moral Claim Rights." *Journal of Political Philosophy* 31, no. 4 (2023): 433–51.

Waldron, Jeremy, ed. *Theories of Rights*. Oxford University Press, 1984.

Wenar, Leif. "Rights." *Stanford Encyclopedia of Philosophy* (Spring 2023). https://plato.stanford.edu/entries/rights/.

Journal of Ethics and Social Philosophy
VOL. 29, NO. 1 · NOVEMBER 2024
https://doi.org/10.26556/jesp.v29i1.3129

MORAL WORTH IN GETTIER CASES

Neil Sinhababu

For actions to have moral worth, must they be motivated by moral knowledge? Paulina Sliwa and J. J. Cunningham say yes.[1] Sliwa writes, "A morally right action has moral worth if and only if it is motivated by concern for doing what's right (conative requirement) and by knowledge that it is the right thing to do (knowledge requirement)."[2] Cunningham's Know How View requires morally worthy actions to be motivated by "one's knowing how to respond to reasons" that make the action right, "triggered by the agent's propositional knowledge of the particular normative reason at issue."[3] Both views require actions with moral worth to be motivated by moral knowledge rather than by more easily achieved epistemic states like justified true moral belief.

Gettier cases involve justified true belief that is not knowledge.[4] They suggest that moral knowledge is not needed for moral worth, as justified true moral belief serves equally well. The following Gettier case is a counterexample to the knowledge requirement:

> *Texting the Rabbi*: Ava faces a moral quandary. William loaned her a weapon. Now he is furiously pounding on her door and demanding it back. Unsure about what to do, Ava texts her rabbi, whom she knows to be an excellent source of advice on moral questions: "William is furiously pounding on my door and demanding his weapon. He might hurt someone, but it is his property. So would returning it be right?" The rabbi understands the situation and replies, "No." Seeing his reply, Ava forms the true belief that it is right not to return the weapon, as reasons to prevent harm are decisive.[5] She rightly does not return the weapon.

1 Sliwa, "Moral Worth and Moral Knowledge"; and Cunningham, "Moral Worth and Knowing How to Respond to Reasons." See also Sliwa, "Praise Without Perfection"; and Sliwa, "Moral Understanding as Knowing Right from Wrong."

2 Sliwa, "Moral Worth and Moral Knowledge," 394.

3 Cunningham, "Moral Worth and Knowing How to Respond to Reasons," 396.

4 Gettier, "Is Justified True Belief Knowledge?"

5 Plato, *The Republic*.

> Beth then faces the same quandary. William loaned her his other weapon. He furiously pounds on her door next, demanding it back. Beth proceeds just as Ava did, considering both options and texting the rabbi the same words. But the rabbi never sees her message, as a thief steals his phone right after he replies to Ava. Seeing Beth's message, the thief mischievously decides to answer. He flips a coin, and it comes up tails, so he replies, "No." Thinking the rabbi sent the message, Beth forms the true belief that it is right not to return the weapon, as reasons to prevent harm are decisive. She rightly does not return the weapon.

In Texting the Rabbi, Ava's and Beth's actions of refusing to return William's weapon both seem to have equal moral worth.[6] They face the same situations, which they deal with by seeking advice and acting on it in the same ways. The difference is that the wise rabbi's testimony causes Ava's belief, while the mischievous thief's randomly chosen answer causes Beth's belief, making them differ in knowledge. Yet Ava and Beth have equal reason to believe that their messages are from the rabbi and equal reason to believe what the messages say. What makes them differ in knowledge is too far beyond them to make them differ in moral worth.

Beth's justified true belief that it is right to not return the weapon falls short of knowledge, making Texting the Rabbi a Gettier case. Her belief has features common to Gettier cases. It violates safety conditions on knowledge, as she would acquire a false belief in nearby worlds where the thief's coin comes up heads.[7] A false lemma causes it, namely that the rabbi sent the message, when the thief actually sent it.[8] Many other things that Gettierize us, like Russell's stopped clock and Dharmottara's mirage, generate unsafe beliefs via false lemmas.[9] With Russell's stopped clock, the false lemma that the clock is running causes true belief about the time, which is unsafe because viewing the clock at

6 If this moral quandary seems so easy that Ava and Beth seem morally incompetent for asking for advice about it, you can imagine William as more judicious or more hotheaded until the dilemma becomes nontrivial. Hills raises the case of Ron the extremist, who asks his rabbi for advice about whether to murder ("Moral Testimony and Moral Epistemology"); and Sliwa responds by treating Ron as morally incompetent ("Moral Worth and Moral Knowledge"). This need not apply to Ava and Beth.

7 Sosa, "How to Defeat Opposition to Moore"; and Pritchard, "Anti-luck Virtue Epistemology." Becker connects safety formulations and reliabilism. See Becker, "Reliabilism and Safety."

8 Clark and Armstrong discuss false lemmas: Clark, "Knowledge and Grounds"; and Armstrong, *Belief, Truth, and Knowledge*.

9 Russell, *Human Knowledge*. As Nagel notes, the Indian philosopher Dharmottara developed Gettier cases around the year 770. See Nagel, *Knowledge*.

other times would generate false belief. With Dharmottara's mirage, the false lemma that one veridically perceives water causes true belief that there is water, which is unsafe because it would have been false without the unseen water nearby under a rock. While precisely stating how justified true belief differs from knowledge is notoriously difficult, false lemmas and violations of safety are common enough to Gettier cases to be suggested as accounts of the difference.[10] Their presence in Beth's case confirms the intuitive sense that she has justified true belief without knowledge.

Ava and Beth differ in whether they know that their actions are right, which Sliwa's formulation is concerned with. Ava receives genuine moral testimony from someone wise and gains moral knowledge from it. Beth unwittingly receives a random answer from a mischief maker, which is not a way of gaining knowledge about morality or most other topics. Because Ava knows that her action is right and Beth does not, Sliwa's account entails that Ava's not returning the weapon has more moral worth than Beth's. This makes Texting the Rabbi a counterexample to Sliwa's version of the knowledge requirement.

Ava and Beth also differ in whether they have propositional knowledge of how to respond to the reasons before them, which Cunningham's formulation is concerned with. Upon seeing the rabbi's reply and coming to know that it is right not to return the weapon, Ava gains the propositional knowledge that the reasons to prevent harm to others make it right not to return William's weapon, even though it is his property.[11] The right-making reason she is aware of, decisive in this case, is her reason to prevent harm to others. This triggers her knowledge of how to respond to these reasons—namely by not returning William's weapon even though it is his property. Upon seeing the thief's reply, Beth does not gain propositional knowledge about how to respond to the reasons before her. She merely has justified true belief that it is right to respond by not returning the weapon—not knowledge. So Cunningham's account entails that Ava's action has more moral worth than Beth's. That their actions have the same moral worth despite Ava's additional knowledge of normative reasons makes Texting the Rabbi a counterexample to Cunningham's version of the knowledge requirement.

10 For a difficult Gettier case for safety conditions, see Williams and Sinhababu, "The Backward Clock, Truth-Tracking, and Safety."

11 Cunningham's account needs to assign importance to something like knowing how to respond to all the reasons in combination or knowing which are decisive or stronger. One should not satisfy the knowledge requirement if one knows one has reason to save a dollar and reason not to kill, and believes that the reason to save a dollar is decisive or stronger. Even if laziness akratically prevents one from killing by combining with the weaker motivation not to kill, this refraining from killing is deficient in moral worth.

As Beth does some actions that the knowledge requirement correctly credits with moral worth, it is important to clarify which action the knowledge requirement gets wrong. Sliwa notes, "Since most actions are complex, agents who perform some morally wrong action may, at the same time, perform actions that are morally right. When these actions are motivated in the right way, the agent is morally praiseworthy for them."[12] As Sliwa's example of a donor to counterproductive charities suggests, Beth does some actions motivated by moral knowledge, to which the knowledge account ascribes moral worth. Beth knows that she should seek and follow advice from the rabbi to address her uncertainty, and the knowledge requirement correctly credits her for these actions. The same is true of Ava, whose equivalent actions have the same moral worth. The knowledge requirement faces problems specifically with Beth's action of not returning the weapon. Since she is Gettierized, the knowledge requirement treats this action of not returning the weapon as having less moral worth than Ava's not returning the weapon, when both actions intuitively seem to have equal moral worth.

Might Ava's and Beth's reliance on moral testimony in this case deprive both their actions of moral worth, rendering them equal? Alison Hills argues that mere moral testimony does not generate the kind of knowledge required for moral worth, perhaps because something else like understanding is required.[13] Many others argue that moral testimony fails to generate the full epistemic benefits of other testimony.[14]

Sliwa and Cunningham understand that knowledge accounts are poorly positioned to treat moral testimony as unusual in this way. Both accordingly treat action driven by knowledge of rightness as both necessary and sufficient for moral worth. Excepting testimonial knowledge would abandon sufficiency. Moreover, treating moral knowledge as having an atypical relation to testimony would concede advantages to rival accounts invoking alternatives to knowledge that more typically have that relation. Rather than pursuing this dubious strategy, Sliwa is faithful to the spirit of the knowledge account, accepting that "moral testimony can be a source of moral knowledge."[15] Breaking any links between moral testimony, moral knowledge, and moral worth would also push the knowledge requirement out of alignment with Timothy Williamson's

12 Sliwa, "Moral Worth and Moral Knowledge," 403.

13 Hills, *The Beloved Self*.

14 See McGrath, "The Puzzle of Pure Moral Deference"; and Fletcher, "Moral Testimony."

15 Sliwa, "Moral Worth and Moral Knowledge," 394.

knowledge-first research program.[16] It treats knowledge as necessary and sufficient for evidence and the justification of assertion, and does not generally make exceptions for testimonial knowledge.

Sliwa instead argues that the knowledge requirement prevents accidental performance of right action from conferring moral worth. She writes that requiring less than knowledge and settling for mere "justification and truth is to give up on the thought that morally praiseworthy actions are non-accidentally right."[17] This view treats Ava's knowledge as making it "not just an accident that she did the right thing."[18]

Gettier cases like Texting the Rabbi provide precision regarding the exact sense of "not just an accident" relevant to moral worth.[19] There is an obvious sense in which Beth's doing the right action is accidental. If the coin had come up heads, Beth would have received a different text message and acted wrongly. But Ava's right action is not far from being similarly accidental. If the thief had stolen the rabbi's phone a little earlier and randomly sent a "yes" message to Ava, she too would have acted wrongly. Right action performed on the basis of testimony can approach accidentality, as advisors might make rare and unpredictable mistakes or be impersonated by impostors. But the absence of mistakes and impostors should not be regarded as an accident that undermines moral knowledge and worth. That would prevent actions motivated by moral testimony from having moral worth and perhaps generalize into a broader skepticism that prevents other fallible sources of evidence from conferring moral knowledge and worth. The knowledge requirement is supposed to identify the counterfactuals under which right action is nonaccidental and can have moral worth. Sliwa writes, "The counterfactuals that matter are simply those that come from our best account of knowledge."[20]

Texting the Rabbi shows that the counterfactuals that matter for knowledge are not the ones that matter for moral worth. One way to put this is that knowledge attributions are sensitive to some differences of modal distance that do not affect moral worth attributions. Beth's belief is not knowledge because false belief is just a coin flip away, while Ava's belief is knowledge because false belief

16 Williamson, *Knowledge and Its Limits*, cited by Sliwa, "Moral Understanding as Knowing Right from Wrong." For discussion, see McGlynn, *Knowledge First?*; and Littlejohn, "How and Why Knowledge Is First."

17 Sliwa, "Moral Worth and Moral Knowledge," 402.

18 Sliwa, "Moral Worth and Moral Knowledge," 406.

19 See Isserow, "Moral Worth and Doing the Right Thing by Accident"; Johnson King, "Accidentally Doing the Right Thing"; and Coates, "Moral Worth and Accidentally Right Actions."

20 Sliwa, "Moral Worth and Moral Knowledge," 401.

is an earlier theft and a coin flip away. The additional modal distance provided by the timing of the theft gives Ava moral knowledge that Beth lacks. But this additional modal distance does not give Ava's action moral worth that Beth's action lacks. So moral worth does not require moral knowledge—making the knowledge requirement false.

The difference in what moral knowledge and moral worth require can be fruitfully expressed in explanationist terms as well.[21] Ava's belief is knowledge because well-considered testimony from a morally knowledgeable person explains it. It is not accidental that the rabbi knew what to do and gave Ava the right answer. Beth's belief is not knowledge because the random outcome of a coin flip explains it. It is accidental that the thief's coin came up tails, and he gave Beth the right answer. But what explains whether our actions have moral worth is too deeply inside us to be affected by the external accidents undermining knowledge in Gettier cases.[22] Morally worthy action indeed must not be merely accidental. But actions motivated by Getterized justified true belief rather than by knowledge can be nonaccidental in the sense required for moral worth.

Moral worth cannot be Gettierized. Knowledge can, as its necessary conditions extend far outside us. A mischievous thief can deprive us of knowledge, while leaving all our experiences, beliefs, and actions exactly the same. Moral worth is not so easily stolen away.[23]

National University of Singapore
neiladri@gmail.com

REFERENCES

Armstrong, David. *Belief, Truth, and Knowledge*. Cambridge University Press, 1973.

Arpaly, Nomy. *Unprincipled Virtue*. Oxford University Press, 2003.

Becker, Kelly. "Reliabilism and Safety." *Metaphilosophy* 37, no. 5 (2006):

21 For views invoking explanatory notions rather than modality, see Faraci, "Groundwork for an Explanationist Account of Epistemic Coincidence"; and Singh, "Moral Worth, Credit, and Non-accidentality."

22 Arpaly and Markovits suggest the agent's quality of will. See Arpaly, *Unprincipled Virtue*; and Markovits, "Acting for the Right Reasons."

23 This paper was improved by helpful feedback from Jessica Isserow, Summer Westerbur, and editors and referees from the *Journal of Ethics and Social Philosophy*. The late John Williams also deserves thanks for giving me confidence in my Gettier cases.

691–704.
Clark, Michael. "Knowledge and Grounds: A Comment on Mr. Gettier's Paper." *Analysis* 24, no. 2 (1963): 46–48.
Coates, Allen. "Moral Worth and Accidentally Right Actions." *Analysis* 81, no. 3 (2021): 389–96.
Cunningham, J. J. "Moral Worth and Knowing How to Respond to Reasons." *Philosophy and Phenomenological Research* 105, no. 2 (2021): 385–405.
Faraci, David. "Groundwork for an Explanationist Account of Epistemic Coincidence." *Philosophers' Imprint* 19, no. 4 (2019): 1–26.
Fletcher, Guy. "Moral Testimony: Once More with Feeling." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer-Landau. Oxford University Press, 2016.
Gettier, Edmund. "Is Justified True Belief Knowledge?" *Analysis* 23, no. 4 (1963): 121–23.
Hills, Alison. *The Beloved Self.* Oxford University Press, 2010.
———. "Moral Testimony and Moral Epistemology." *Ethics* 120, no. 1 (2009): 94–127.
Isserow, Jessica. "Moral Worth and Doing the Right Thing by Accident." *Australasian Journal of Philosophy* 97, no. 2 (2019): 251–64.
Johnson King, Zoe. "Accidentally Doing the Right Thing." *Philosophy and Phenomenological Research* 100, no. 1 (2020): 186–206.
Littlejohn, Clayton. "How and Why Knowledge Is First." In *Knowledge First: Approaches in Epistemology and Mind*, edited by Adam Carter, Emma Gordon, and Benjamin Jarvis. Oxford University Press, 2017.
Markovits, Julia. "Acting for the Right Reasons." *Philosophical Review* 119, no. 2 (2010): 201–42.
McGlynn, Aidan. *Knowledge First?* Palgrave Macmillian, 2014.
McGrath, Sarah. "The Puzzle of Pure Moral Deference." *Philosophical Perspectives* 23, no. 1 (2009): 321–44.
Nagel, Jennifer. *Knowledge.* Oxford: Oxford University Press, 2014.
Plato. *The Republic*. Translated and edited by C. D. C. Reeve. Hackett, 2004.
Pritchard, Duncan. "Anti-luck Virtue Epistemology." *Journal of Philosophy* 109, no. 3 (2012): 247–79.
Russell, Bertrand. *Human Knowledge: Its Scope and Limits.* Routledge, 2009.
Singh, Keshav. "Moral Worth, Credit, and Non-accidentality." In *Oxford Studies in Normative Ethics*, vol. 10, edited by Mark Timmons. Oxford University Press, 2020.
Sosa, Ernest. "How to Defeat Opposition to Moore." *Noûs* 33, no. 13 (1999): 141–53.
Sliwa, Paulina. "Moral Understanding as Knowing Right from Wrong." *Ethics*

127, no. 3 (2017): 521–52.
———. "Moral Worth and Moral Knowledge." *Philosophy and Phenomenological Research* 93, no. 2 (2016): 393–418.
———. "Praise Without Perfection: A Dilemma for Right-Making Reasons." *American Philosophical Quarterly* 52, no. 2 (2015): 171–85.
Williams, John, and Neil Sinhababu. "The Backward Clock, Truth-Tracking, and Safety." *Journal of Philosophy* 112, no. 1 (2015): 46–55.
Williamson, Timothy. *Knowledge and Its Limits*. Oxford University Press, 2000.