JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXII · NUMBER 1
May 2022

ARTICLES

DISCUSSION

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

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

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

Journal of Ethics and Social Philosophy
VOL. 22, NO. 1 · MAY 2022
https://doi.org/10.26556/jesp.v22i1.1698

LAW AND VIOLENCE

Alexander Guerrero

CRIMINAL LAW, legal and political institutions, and efforts aimed at reforming those institutions all mark a significant difference between violent and nonviolent criminal actions. Violent crimes are typically met with more severe punishments and more extensive collateral consequences than nonviolent crimes—even when the violent crimes cause less harm. Advocates for criminal justice reform make their case by pointing to the high numbers of people incarcerated for nonviolent offenses and offering reform proposals that would significantly alter the treatment of nonviolent offenders—with the implicit or explicit suggestion that this is the heart of the injustice, and that things should stay as they are for those convicted of violent offenses.[1] The United States Sentencing Commission's 2016 *Report to Congress on Career Offender Sentencing Enhancements* made the case that sentencing enhancements should only be triggered by crimes of violence, and that they should no longer be triggered by convictions for drug trafficking.[2] Recent state efforts to re-enfranchise those convicted of felonies and to make record expungement easier have been barred to those convicted of a violent crime.[3] And in the midst of the COVID-19 pandemic, calls to release people from jails and prisons have focused almost entirely on "nonviolent" offenders.[4] David Sklansky argues in impressive detail in his recent book that "no distinction plays a larger role in contemporary American criminal law than the line between violent and nonviolent offenses."[5] Despite a

1 For example, see Silva, "Clean Slate"; Outlaw, "Time for a Divorce."

2 United States Sentencing Commission, *Report to the Congress*.

3 For example, the 2019 New Jersey criminal justice reform act allows for easier record expungement, except for those convicted of a violent criminal offense (see State of New Jersey, "Governor Murphy Signs Major Criminal Justice Reform Legislation"). An executive order in Kentucky restored voting rights to 140,000 convicted felons, but limited to those convicted of nonviolent offenses (Wines, "Kentucky Gives Voting Rights to Some 140,000 Former Felons").

4 Reported, with other examples, in Kim, "Why People Are Being Released from Jails and Prisons during the Pandemic."

5 Sklansky, *A Pattern of Violence*, 41.

decade of significant discussion of criminal justice reform, the refrain remains: violent crime is different; those convicted of violent crimes are different; and it is appropriate to punish and respond to violent crime differently.

In this article, I argue that the violent/nonviolent distinction cannot bear the normative weight currently placed on it and that we should jettison thinking in terms of violent crime and move to thinking in terms of wrongful harm caused and risked. I argue that, if we do this, our current practices of sentencing and punishment require revision, and that we should make those revisions. The basic argument is that there are moral constraints on punishment; that these are provided by (a) the amount of wrongful harm caused or risked and (b) facts about agent culpability; and that there is no consistent relationship between a crime being violent and how much wrongful harm was caused or risked by that crime, nor is there a close relationship between whether a crime involves violence and the degree of culpability of the agent committing the crime. In the conclusion of the article, I offer an error theory concerning our commitment to treating violent crime differently than nonviolent crime, attempting to explain why we see this distinction as important in the criminal law and suggesting that morally better categorizations are available to us.

It is worth stressing that I am not arguing that violent crime is not incredibly harmful in some cases, or that all violent crime should be punished less than it currently is. The right response could be—and in some cases probably will be—to increase penalties for very harmful nonviolent crime, rather than to lessen penalties for very harmful violent crime. Nor is it my suggestion that it is never appropriate to pay attention to the specific nature of the criminal offense in terms of the kind of wrongful harm that is caused or risked. It might be appropriate, for example, to have greater restrictions on future gun ownership for those who are convicted of a weapons offense. That is very different than the categorical difference in treatment that we currently see in the United States.

If successful, this argument would have substantial implications for current law and policy.[6] The differential punishment of violent crime is central to the mass incarceration crisis in the United States. Michelle Alexander suggests that "the uncomfortable reality is that arrests and convictions for drug offenses—not violent crime—have propelled mass incarceration."[7] John Pfaff labels this the "Standard Story" regarding mass incarceration. His recent book makes a powerful case that this story is not the full story. Through analysis of state and federal data, Pfaff demonstrates that more than half of the increase in state prison

6 My focus throughout is on the US context. Although the argument applies more broadly, the issues are of perhaps distinctive importance in the US.

7 Alexander, *The New Jim Crow*, 102.

growth in the 1980s through 2010 came from people serving time for violent offenses.[8] If we are serious about addressing mass incarceration and rethinking the role prisons are playing in our society, we must also reconsider the way we are responding to violent crime and to those convicted of violent offenses. We must not shy away from talking about violent offenses—what Pfaff calls the "third rail" of criminal justice reform. Questioning the normative weight currently placed on violence as a category in law must be a central part of that conversation.

I

Let me provide an overview of the central argument of this article. The argument begins with an empirical fact about the significance of being convicted of a violent crime in the United States:

1. *Violence and Law*: The United States legal system marks a categorical difference between violent crime and other crime that is materially significant in terms of sentencing and punishment, including: sentence length; eligibility for probation and parole; eligibility for government-provided benefits, employment opportunities, and civic roles; and eligibility for alternatives to incarceration including probation, and diversion from incarceration into substance abuse or mental health treatment.

The moral significance of this writing of violence into the law is evident if we attend to some general claims about the morality of punishment. In particular, consider the following four claims:

2. *Proportionality*: A necessary condition of a punishment being permissibly exacted upon *S* is that the severity of the punishment is proportional to the crime for which *S* has been convicted.
3. *Equality*: Two people should not be punished substantially differently unless there is a morally significant difference between them in terms of (a) their culpability for committing the crimes or (b) the crimes for which they have been convicted.
4. *Proportionality and Harm*: Assuming two equally culpable offenders, proportionality of punishment for an action should be tied to the wrongful harm caused or risked by that action: the greater the wrongful harm caused or risked, the greater the maximum permissible severity of punishment.

8 Pfaff, *Locked In*, 31–36, 187–90.

5. *Equal Harm, Equal Punishment*: Assuming two equally culpable offenders, the quantity of punishment for two crimes, *C1* and *C2*, should not differ substantially unless *C1* and *C2* differ substantially in wrongful harmfulness caused or risked.

These claims about the morality of punishment have significant implications for the use of violence as a significant legal category, which we can see by noticing important, underappreciated facts about violence and all the actions that are included in that category.

6. *Wide Variation in Harmfulness of Violence*: Violent criminal action is not a uniform category such that all or most actions in that category cause or risk causing a similar amount of wrongful harm.
7. *Violent Action Not Systematically More Harmful than Nonviolent Action*: It is not true that all or almost all violent criminal actions are more wrongfully harmful than nonviolent criminal actions.
8. *No Positive Correlation between Violence and Culpability*: Those who commit violent offenses are no more likely to be culpable for offending, nor are they likely to be more culpable, than those who commit nonviolent offenses.

And we are not forced to use the category of violence as a matter of administrative convenience.

9. *Better Categories Possible*: There are usable categorizations of actions that do a better job sorting actions by their wrongful harmfulness than the violent/nonviolent categorization.

Taken together, with some details filled in, we reach the following:

> *Conclusion*: We should jettison the use of the category of violent crime for purposes of punishment—including the assignment of collateral consequences and the availability of parole and diversion from incarceration—and instead use categorizations that better track wrongful harm caused and risked.

The rest of the article explains and defends these claims.

II

Although a philosophical discussion of the concept of violence might be of interest, I will focus on the ordinary conception of violence that figures into actual

law, as it is that conception that is used to sort crimes into categories, and it is that conception that I argue cannot bear the normative weight currently placed on it.[9] Although jurisdictions differ in the details, a basic characterization of violence is found in US federal law (which influences most sub-jurisdictions within the US), which defines a "violent felony" as any crime punishable by imprisonment of greater than one year that "has as an element the use, attempted use, or threatened use of physical force against the person of another; or … burglary, arson, or extortion, involves use of explosives, or otherwise involves conduct that presents a serious potential risk of physical injury to another."[10] Violent crimes other than felonies are just those that are otherwise similar but that have shorter sentences attached to them. Notably, this definition includes physical force against persons, attempts and threats, and both intentional and reckless action.[11]

9 For a helpful discussion of the understanding of "violence" in law, and the changing understanding of violence over time, see Ristroph, "Criminal Law in the Shadow of Violence."

10 18 USC § 924(e). The last clause, known as the "residual" clause, has proven difficult for courts to interpret and apply. In *Johnson v. United States*, the Supreme Court declared it unconstitutionally vague (135 S.Ct. 2551 (2015)).

11 The recent law in the United States has focused on trying to interpret a particular string of language, which appears in a number of different places in federal law definitions of what constitutes a "violent felony": "a crime … that has as an element the use, attempted use, or threatened use of physical force against the person of another; or … burglary, arson, or extortion, involves use of explosives, or otherwise involves conduct that presents a serious potential risk of physical injury to another." 18 USC § 924(e). This provides a rough guide to how to think about "violence" in law, but it leaves a number of questions unclear, as the Supreme Court itself has said.

In *Begay v. United States*, 553 US 137 (2008), the Supreme Court explained that "the provision's listed examples illustrate the kinds of crimes that fall within the statute's scope. Their presence indicates that the statute covers only similar crimes." The Supreme Court further reasoned that the listed crimes "all typically involve purposeful, 'violent,' and 'aggressive' conduct." Importantly, the Supreme Court also stressed that the correct way to discern whether a crime was "violent" or not was to look (somehow!) at "ordinary" instances of that crime, not at the particular facts in a particular case. The court used this reasoning to find that felony driving while intoxicated is not a "violent felony" for purposes of the Armed Career Criminal Act (ACCA).

As noted above, the clause in the ACCA (and other similar statutes) that states that violent felonies will include those crimes that "otherwise [involve] conduct that presents a serious potential risk of physical injury to another" has come to be known as the "residual" clause. This clause has proven particularly difficult to interpret and apply. Many crimes can arguably fit under the "serious potential risk of physical injury" standard, and so federal courts were frequently divided over which crimes were covered by the residual clause. Additionally, courts were instructed to consider an "ordinary" case of a crime, although they do not have any evidence about what "ordinary" or typical versions of these crimes look like. So, it is no surprise that the Supreme Court has had to struggle with the residual clause. In *Johnson v. United States*, the Supreme Court finally declared ACCA's residual clause unconstitutionally

In section v, I enumerate and discuss the main categories of violent and nonviolent crime at greater length.

There are many ways in which a crime being classified as violent or nonviolent can make a difference to the sentencing and punishment of those convicted of that crime. For the purposes of this article, I include under the heading of "punishment" all of the following: initial sentence length, total time served for the offense (not just the initial sentence, but also factoring in the availability or likelihood of parole), legally mandated collateral consequences of the conviction, and facts about the nature of the legal punishment, including whether one is incarcerated or is instead permitted to be on probation or enter an alternative diversion program.[12] Being convicted of a violent offense can affect severity of sentencing and direct punishment, resulting in longer sentences and serving as a distinctive kind of trigger for mandatory minimum or repeat offender extended sentences.[13] It can affect whether a person will be eligible for various government-provided benefits, employment and volunteer opportunities, and civic roles—even after completion of their sentence.[14] Perhaps most significantly, it

vague. In *Welch v. United States*, 136 S.Ct. 1257 (2016), the Supreme Court made the decision retroactive, potentially putting many sentences into question if they relied on convictions under the residual clause. The Supreme Court has found other incorporations of this clause, as in the Immigration and Nationality Act, also unconstitutionally vague (Sessions v. Dimaya, 138 S.Ct. 1204 (2018)). To my mind, the difficulty in codifying a precise definition of "violence," as well as discerning how courts should decide whether a particular instance of a crime was "violent," provides just one more reason to jettison the significance of this category. It should at least rebut worries that it is considerably easier to do this than to engage in what I will later recommend: analysis of the wrongful harm caused or risked.

12 Some of these are perhaps controversial as "punishment" for theories of punishment that focus on what the state is trying to express or communicate through punishment (such as Wringe, *An Expressive Theory of Punishment*). Even for those theories, there is a plausible case that there are expressive dimensions to these other components. Making that case in full would require more discussion, but I will suggest later in the paper that our differential attitudes toward violence are a significant part of the explanation of the use of violence as a distinctive category in law, and we are plausibly communicating a message about the nature of an offender's wrongdoing to society at large when we treat those convicted of a violent crime differently in all of these ways. The argument of the article is at least partly that this message is inapt and misplaced.

13 The ACCA, passed in 1984, imposes a mandatory minimum sentence of fifteen years for any person illegally possessing a firearm who has three prior convictions for violent felonies (18 USC § 924(e)). The United States Sentencing Guidelines (USSG) career offender enhancement applies when a defendant is facing prosecution for either a serious drug crime or a crime of violence, and has at least two prior convictions for either serious drug crimes or crimes of violence. If these conditions are met, then USSG § 4B1.1 provides for a guideline range "at or near the maximum [term of imprisonment] authorized."

14 In many jurisdictions, all people convicted of a felony—nonviolent or violent—lose civic

can affect an individual's eligibility for alternatives to incarceration: probation and parole and various other forms of diversion from incarceration into substance abuse or mental health treatment. Here I will canvas some of these, to highlight significant examples in support of 1.

People convicted of violent crimes serve more time in prison than those convicted of nonviolent offenses. Those convicted of violent offenses (roughly 30 percent of people admitted to state prisons) spend an average of 3.2 years in prison, whereas the overall average (including those convicted of a violent offense) is only 1.7 years.[15] Additionally, although people convicted of a violent crime make up only a third of prison admissions, they make up more than half of the people in prison at any time. As Pfaff puts it, "violent offenders take up a majority of all prison beds, even if they do not represent a majority of all admissions."[16]

Why is this? Some of this difference is a function of initial sentence length (sometimes due to enhancements and mandatory minimums). But a significant component is the expanded use of parole for everyone except those convicted of a violent crime. Pfaff notes that of the three hundred thousand people admitted to prison in 2003 in seventeen states, only 3 percent had not yet been released or paroled by the end of 2013.[17] And of that 3 percent, almost 85 percent had been convicted of a violent crime. Some of this is because of a difference in average initial sentence length. But there is also this significant factor: parole is rarely granted to those who have been convicted of a violent crime. And this is despite a general trend toward an increased use of parole. As Pfaff summarizes the situation:

> After years of limiting and restricting [parole], states have started to rely on parole more extensively. Such reforms are in fact perhaps the most widely adopted type of prison reform to date. In almost all cases, however, these changes have been limited to people convicted of nonviolent crimes.[18]

Marc Morjé Howard makes a similar point. He argues that parole could be safely

rights, including the right to vote or serve on a jury. There are also collateral consequences that apply specifically to those who are convicted of violent crimes. Under US law, those convicted of violent felonies receive lifetime bans on Section 8 and other federally subsidized housing, and local housing authorities are authorized to refuse housing to individuals who have "engaged in any … violent criminal activity" (42 USC § 13661(c) (2006); 24 CFR § 906.203(c) (2010)). And there are similar barriers to those convicted of violent crimes in terms of federal and state employment and licensing permits (see 45 CFR 2522.205, 2540.200).

15 Pfaff, *Locked In*, 188.

16 Pfaff, *Locked In*, 188.

17 Pfaff, *Locked In*, 188–89.

18 Pfaff, *Locked In*, 198.

expanded to those who have been convicted of violent crime, but state legislators and parole board members (typically political appointees) are unwilling to risk implementing reforms or make parole decisions that might result in a person convicted of a violent crime then committing a violent crime while on parole.[19]

An additional explanation comes from the costs of mass incarceration more directly. The Violent Crime Control and Law Enforcement Act of 1994 included many provisions that contributed to the mass incarceration crisis, one of which was creation of the Violent Offender Incarceration and Truth-in-Sentencing Incentive Grants Program. This program provided millions of federal dollars in grants to states to build or expand correctional facilities, provided that the states had sentencing guidelines in place that required those convicted of *violent* crimes to serve no less than 85 percent of their sentences. As a result of this program, by 1999, twenty-eight states and the District of Columbia had adopted sentencing guidelines forcing those convicted of violent crimes to serve no less than 85 percent of their sentences, and three states required such people to serve 100 percent of their sentences.[20]

As mass incarceration has come under more widespread criticism on moral and economic grounds, a wide variety of alternative courts and alternatives to incarceration have been created or expanded, in addition to the expanded use of parole. Many of these alternatives are foreclosed to people charged with or convicted of a violent crime.

One of the most common alternatives comes in the form of drug courts. There are now over three thousand drug courts in the United States, with drug courts in all fifty states.[21] These courts aim to divert people into substance abuse treatment, rather than incarceration, as an acknowledgment that many people who engage in crime have significant substance abuse problems, and that these problems often are at the root of their criminal conduct. These courts have a range of criteria for eligibility, but most have a requirement that people not be charged with or have a conviction for a violent crime. This is due largely to the aforementioned Violent Crime Control and Law Enforcement Act of 1994, which authorized billions of dollars for anti-crime programs with specific funds allotted for the implementation of drug court programs, but eligibility criteria limited participation in these programs to *nonviolent* drug-involved offenders.[22]

Along with an increasing realization that incarceration is not the best response to substance abuse problems, so, too, there is increasing awareness that

19 Howard, *Unusually Cruel*.

20 Ditton and Wilson, "Truth in Sentencing in State Prisons."

21 National Association of Drug Court Professionals, http://www.nadcp.org/about/.

22 Saum, Scarpitti, and Robbins, "Violent Offenders in Drug Court."

mental health problems might not be best addressed by incarceration. Mental health "diversion" programs place offenders in treatment programs rather than incarcerating them. The record of these programs is generally good in terms of treating offenders with mental health issues and reducing recidivism. Unfortunately, the main legislation addressing this issue in the past twenty years offers substantial federal grant money for diversion programs, such as mental health courts, but only if they are barred to those charged with or convicted of violent offenses.[23] Thus, it is little surprise that many local and state diversion programs ban people with mental health problems who committed a violent offense.[24]

Almost every jurisdiction allows some people to be sentenced to supervised probation, rather than time in jail or prison, and almost all of these jurisdictions have consideration for "public safety" as an explicit, statutorily mandated consideration. Consideration of "public safety" has almost always focused on safety with respect to threats of violence. As a result, nonviolent felonies such as white-collar crimes or simple drug possession typically have a much better chance of qualifying for supervised probation than violent felonies. It is common to have an explicit bar on probation for people convicted of violent felonies.[25]

III

Recall two of the general claims about the morality of punishment presented above:

2. *Proportionality*: A necessary condition of a punishment being permissibly exacted upon *S* is that the severity of the punishment is proportional to the crime for which *S* has been convicted.
3. *Equality*: Two people should not be punished substantially differently unless there is a morally significant difference between them in terms of (a) their culpability for committing the crimes or (b) the crimes for which they have been convicted.

Let me expand upon and clarify these claims. Consider two different theories of the morality of punishment: hybrid theories and retributivist theories. On a hy-

23 Mentally Ill Offender Treatment and Crime Reduction Act of 2004 (MIOTCRA), Pub. L. No. 108-414, 118 Stat. 2327 (2004). MIOTCRA permits a small amount of grant money to be used to address treatment for mentally ill violent offenders, but only through in-prison programs. But in-prison programs have been much less effective—little surprise given that prisons exacerbate mental illness.

24 See, for example, California: Mentally Ill Offender Criminal Reduction Act 32 (MIOCR) S.B. 1485, 1997-98 Leg., Reg. Sess. § 1(g) (Cal. 1998).

25 See e.g., Tex. Code Crim. Pro. Ann. art 42.12, § 6.

brid theory, there are constraints on who may be punished (only those convicted of an offense), how much they can be punished (only in an amount that is proportional to the gravity of their offense), and how much punishment can differ between similarly culpable persons convicted of the same offenses (not much). Subject to those constraints, the theory says: determine who should be punished and how much they should be punished based on the consequences of punishment.

On a simple retributivist theory, punishment is justified and morally appropriate if and only if it is deserved. This view sets out moral desert of punishment as both a necessary and sufficient condition for punishment being appropriate. Thus, the answer to the targeting question is: people should be punished if and only if they deserve to be punished. To the quantity question, the answer is: people should receive an amount of punishment equal to what they deserve. The hard question then becomes: What is the appropriate basis or bases of desert of punishment?

Both hybrid and retributivist theories make a central place for Proportionality and Equality (on some views, it may be that Equality, or a nearby variant, is just entailed by Proportionality). These are also both intuitive principles, corresponding with common judgments about cases. I will not argue for them further here.[26] Still, despite this, or perhaps because of it, little has been said about the details of Proportionality and Equality with respect to these two questions:

a. What is it that punishment should be proportional *to*?
b. When we maintain that like cases should be treated alike in terms of punishment, what are the relevant dimensions of similarity in the cases that matter, morally?

IV

When thinking about proportionality and equality of punishment, we must keep two questions distinct. One: *Is this person culpable or fully culpable for committing the crime*? Two: *How morally bad is the crime*? Plausibly, proportionality and equality considerations attach both to *how culpable* the person was and *how bad* the action was. I will say more about culpability later, but I mean something like "moral responsibility for acting" where (on different theories) that can include facts about an agent's intentions, will, control, and causal responsibility.

26 Simple consequentialist views about punishment reject both Proportionality and Equality, but that is a significantly counterintuitive position, and I take it that few who wish to defend the use of the category of violence do so on pure consequentialist grounds. In fact, consequentialist considerations cannot support the use of the category of violence, as follows independently from claims 6–9.

Here, I will focus on the second kind of consideration, how bad the action was, and I will assume that the offenders in question are equally culpable—both 4 and 5 have explicit clauses setting aside culpability. This might be worrying if violent offenders were always or typically more culpable for what they do than nonviolent offenders. But I argue against that suggestion below, when discussing 8, and so I leave questions of culpability to the side.

I will argue that what is significant about an action for proportionality and equality analysis is the *wrongful harm* caused or risked by the action. Focus on this kind of harm analysis forces us to consider more seriously the true harmful consequences of crime. This is not easy. There are difficulties in quantifying different kinds of harm in a way that allows comparisons, for example. But this is already done in other legal contexts (torts, for example), as well as in medical contexts in which assessments must be made about the costs and benefits of interventions and allocations of limited resources. People draw on empirical surveys, studies based on revealed preferences, and other admittedly imperfect methods to do this.[27] Furthermore, that it is difficult does not mean that it is not what we ought to be doing, or even what we are (very roughly) trying to do when we distinguish between, say, misdemeanors and felonies, or between different sentencing guideline ranges for different crimes. We should be doing this kind of wrongful harm analysis more explicitly and we should be using this analysis to guide our thinking about proportionality and equality. Let me say more to clarify and defend this view.

Wrongful Harm Caused or Risked

Here is an intuitive account of harm and harming: *A* harms *B* if and only if *A* causes *B* to be in a bad state—either absolutely bad, or bad relative to other relevant alternative states that *B* might have otherwise been in.[28] Non-wrongful harm cases are ones in which *A* causes *B* to be in a bad state by doing *X*, but *B* had no right or reasonable expectation that *A* not cause *B* to be in this state by doing *X*. Consider a case in which *A* rejects *B*'s offer of going on a date, leading *B* to be depressed. Or a case in which *A* and *B* are both fairly competing for a job, and *A* gets the job, resulting in *B* being unemployed.

Wrongful harm comes in a wide variety, but will include, prominently, things that you might impermissibly do to cause my physical body, things that I care

27 See e.g., Prieto and Sacristán, "Problems and Solutions in Calculating Quality-Adjusted Life Years (QALYs)."

28 There are cases that pose difficulties for the details of this account of harm; those details need not detain us here. For discussion, see Harman, "Harming as Causing Harm"; Shiffrin, "Harm and Its Moral Significance."

about, or things that I have rights over (such as property or ideas) to be destroyed or taken from me or made worse off in significant ways.

Some actions that cause wrongful harm to a primary victim also cause broader social and psychological effects constituting wrongful harms to people who might be called secondary victims. Mass public shootings, sexual assault, domestic violence, and terroristic racialized lynching all provide clear examples of this. When considering wrongful harm that is caused by an action, we should include these secondary harms, including harms to those other than the primary victim, such as psychological injuries due to increased anxiety or fear, offense, or broader effects on social position or social standing. Harm caused through these kinds of broader social effects is wrongful—it results in violations or diminutions of rights that people have to autonomy, equality, respect, social standing, and so on. Individuals have a right against intentional or terroristic infliction of emotional distress. (Some harms will not count as wrongful, because one does not have a right against being caused to suffer them in this way—say, the psychological distress gay marriage causes homophobic people.) More would have to be said to demarcate the precise contours of individual rights here, but secondary wrongful harms—either through broader social effects or through individual subjective experience and emotional and psychological distress—caused by an action should count on the ledger of that action for the purposes of proportionality judgments. Additionally, harm to secondary victims is often a foreseeable result of certain criminal actions, undercutting at least one potential objection to counting this kind of wrongful harm caused by the action for purposes of assessing proportionality and equality.

Another complication comes in countenancing not just the actual wrongful harm caused by criminal action, but also the harm that was risked. There are hard issues here about exactly how harm risked should be weighed in relation to wrongful harm caused. Some will tolerate and embrace a significant amount of moral luck on this score; others are wary of tolerating significant differences here. One possibility would be to identify ranges of likely or expected or standard consequences for various *types* of criminal actions and treat those as the harms risked even in cases in which little or no harm materializes. But there are complications.

Wrongful Harm and Proportionality

Claims regarding criminalization and harm are familiar from debates about the so-called harm principle offered by Mill and refined by many. Antony Duff has put forward a version of the harm principle in criminal law discussions, suggesting that "only conduct that wrongfully harms or threatens to harm others is a suitable candidate for criminalization."[29]

29 Edwards, "Theories of Criminal Law."

The fact that harm has been seen as central to questions of criminalization does not require that wrongful harm should also be morally central with respect to proportionality analysis, but it suggests that it might be a decent starting point. Most discussions of proportionality focus in a general way on the *badness* or *gravity* of the actions in question, without specifying more precisely what dimension of badness or gravity is relevant, or how those ideas are to be understood.[30]

Here is a hypothesis: if pressed to offer a rationale to explain "badness" or "severity" or "gravity" of offenses, most would settle on something like how much wrongful harm was caused or risked. Those theorists who have spoken to the issue usually cite harm caused or risked as a central factor in assessing moral desert and the gravity of a criminal act—the other significant factor being the agent's culpability for performing the act. Göran Duus-Otterström states that "criminal seriousness is usually taken to be a function of the harm, or risk of harm, imposed by the offender, and the culpability of his doing so."[31] Antony Duff says that to rank crimes in terms of their seriousness, "we must … identify and rank criminal harms, identify and rank kinds of criminal culpability, and then combine these two rankings into a single scale of criminal seriousness."[32] But we have not done a particularly good job of correctly ordering criminal offenses from worst to least bad in terms of the wrongful harm caused or risked—even if this is what we are roughly, imperfectly, trying to do.

If we have set offender culpability to the side, including facts about the role that the offender played in causing the wrongful harm, it becomes somewhat unclear what properties of actions could matter to proportionality analysis other than wrongful harm. Evilness? I suspect that intuitions about *evilness* of actions are really standing in for something else: *social deviance*. But there remains the question of why it would be permissible to punish actions more simply for being comparatively socially abnormal. A possibility here is that even similarly harmful actions might differ in how strongly a political community wants to punish or deter them, and this might relate to democratic or popular decisions regarding punishment. Assume that two kinds of actions cause the same amount of wrongful harm (always, or on average). Could a democratic polity permissibly decide to punish one of the two twice as severely as the other? Ten times as severely? It might be a matter of reasonable disagreement how much wrongful harm is caused by an offense or type of offense, and democratic politics might be one way of permissibly resolving such disagreements. But even there, wrongful harm caused or risked seems to be the correct anchor for the discussion and disagreement.

30 See Von Hirsch, "Proportionate Sentencing."

31 Duus-Otterström, "Why Retributivists Should Endorse Leniency in Punishment," 469.

32 Duff, *Punishment, Communication, and Community*, 135.

Importantly, although it is natural to think that crimes that target or disproportionately affect certain groups—perhaps those in already marginalized social positions—or that are aimed at sustaining gender or racial hierarchy might be particularly bad and deserving of greater punishment, these broader social effects will be included in the wrongful harm analysis, as suggested in the previous section.

Wrongful Harm and Equality

Proportionality identifies a limit on how much an individual can be punished. Many retributivist theorists see this as setting the exact appropriate amount: a person should not be punished *more* than this, but they also should not be punished *less* than this.[33] Hybrid theorists might see this as setting a ceiling: you cannot punish a person *more* than this, given the severity of what they have done. As a result, one question that emerges for hybrid theorists is the question of fairness of punishment across a range of cases, involving different individuals. For retributivists—at least of the "mandatory" variety—if proportionality is being respected, and if people are being punished exactly as much as they deserve, then equality across cases will be assured.

But for others, the question emerges: When, and on what grounds, is it morally permissible to punish two people convicted of the same offense differently? If two agents, Smith and Jones, are equally culpable for offending, it is permissible to punish Smith and Jones different amounts only if their offenses were different in some morally significant way. The claim I assert in 5 is that the only morally significant difference between offenses that might license differential punishment is the wrongful harmfulness caused or risked.

As in the case with proportionality, it is hard to imagine what other properties of offenses (evilness?) might be morally relevant in terms of licensing greater punishment. Unlike in the case of proportionality, however, here there might be a temptation to consider factors beyond either wrongful harmfulness or culpability: factors that are agent focused, rather than offense focused. These considerations might not affect the culpability of the agent, but might seem to permit differences in punishment. Consider, for example, the possibility that one of the two people is substantially more likely to reoffend, and this is known by the sentencing authority. This kind of forward-looking consideration veers closely into troubling pre-punishment territory. But there is much that might be said about

33 "Mandatory" retributivists believe that we are required to punish exactly as much as the person deserves; "permissive" retributivists maintain that we are permitted but not required to punish people in line with what they deserve (we may punish them less than that). See Braithwaite and Pettit, *Not Just Deserts*, 34–35.

this, and a full defense of 5 would require saying more. For our purposes, we can obviate the need for that discussion by simply noting the empirical fact—discussed in section VI—that those who commit violent crimes are no more likely to recidivate than those who commit nonviolent crimes. If agent-focused factors do end up being appropriately considered, we could modify 5 to incorporate that fact, and then add the empirical claim regarding recidivism to the argument.

In the next sections, I will consider the implications for accepting these claims regarding the morality of punishment and the significance of wrongful harm for how or whether "violent" criminal action should be treated as a distinct category. Importantly, one need not accept these claims to consider the implications of accepting them. And considering these claims also motivates the question: If one does not embrace these claims about the morality of punishment, what are the other claims that one does accept that justify treating "violent" criminal action as a significant category within law?

V

Consider the following claim:

6. *Wide Variation in Harmfulness of Violence*: Violent criminal action is not a uniform category such that all or most actions in that category cause or risk causing a similar amount of wrongful harm.

This claim should be uncontroversial. All of the following count as violent crimes in US jurisdictions: murder (in different degrees), manslaughter (voluntary and involuntary), rape and other forms of sexual assault, assault with a deadly weapon, kidnapping, robbery, aggravated assault, reckless endangerment, and simple assault (attacks or attempted attacks without a weapon resulting in either no injury or minor injury). Jurisdictions differ with respect to how these crimes are defined and the precise terminology used to describe them. Still, in every case, there is wide variation in how wrongfully harmful actions in this category are—either taken on a case-by-case, act-token basis, or looking at the act types. Punching someone is much less wrongfully harmful than murdering someone.

Furthermore, although many violent offenses are very wrongfully harmful, it is not true that, as a class, they are more harmful than nonviolent action. That is, we should also accept:

7. *Violent Action Not Systematically More Harmful than Nonviolent Action*: It is not true that all or almost all violent criminal actions are more wrongfully harmful than nonviolent criminal actions.

Even if violent offenses varied significantly in their wrongful harmfulness, it might still be reasonable for them to be treated as a legally significant category if violent offenses were all or almost all worse than nonviolent offenses in terms of wrongful harmfulness caused or risked. But that is not so.

First, consider the range of nonviolent offenses: fraud, tax crime, bribery, forgery, racketeering, theft, burglary, embezzlement, cybercrime, identity theft, illegal drug manufacturing and distribution, possession and distribution of child pornography, and criminal damage to property—just to name some of the more central examples. Now, consider the categorical claim that all violent criminal action is more wrongfully harmful or risks more wrongful harm than any nonviolent criminal action. This is clearly false. Irreparably defrauding a person of their life savings causes more wrongful harm than stealing that person's car at knifepoint. Embezzlement that causes a company's financial ruin, and the attendant loss of employment of fifty people, causes more wrongful harm than the poorly thrown punches of two drunk people in a bar fight. Nonviolent crimes like those committed by former judges Michael Conahan and Mark Ciavarella, who were convicted of fraud and racketeering for accepting money in return for imposing arbitrary and excessively harsh judgments on more than five thousand juveniles in order to increase occupancy at for-profit detention centers, can cause nearly unimaginable amounts of wrongful harm—more than all but the most horrific violent crimes.[34]

It is hard to see how the categorical claim could be defended. One way might be to try to argue that *physical* harm (such as might be caused directly by violent actions) is always more significant than *nonphysical* harm (such as might be caused by nonviolent actions). But physical harm is not always worse than nonphysical harm. We can run a simple Millian argument to show this. Many of us have experienced both physical and nonphysical harms. It is not the case that all those who have experienced both kinds of harms feel the physical ones always to be the worst of the two. Indeed, many would happily exchange nonphysical harm for physical harm, if given the choice. I would rather be punched or have my arm broken by someone pushing me down than to be defrauded out of my life savings.

More to the point, both physical and nonphysical harm can be caused by nonviolent criminal actions. It is often straightforward to determine the harm caused by a violent criminal action: a person was shot in the arm or had his jaw broken. There are often also nonphysical harms that follow from those physical ones. Physical harms caused by nonviolent crimes may be more diffuse (although they may not be—consider driving under the influence, which is classified as non-

34 See Urbina, "Despite Red Flags about Judges, a Kickback Scheme Flourished."

violent in the United States post-*Begay*).[35] There may be hard questions about exactly which harms that were in some sense caused by the action are going to count. I steal tens of thousands of dollars from you. This causes you financial devastation, you end up temporarily homeless, and this in turn contributes to serious health problems. Or I illegally operate a "pain management" clinic that is really just a supplier of illegal prescriptions for OxyContin, causing physical harm both to those addicted but also to their children (through malnourishment and neglect) and the broader community as chaos and disrepair takes over. In these cases, the physical harm is clear, although the full accounting of the wrongful harm caused by the criminal action may be complicated by the fact of intervening agency (to some degree) of those who knowingly use the drugs.

It is worth taking a brief detour to address the question regarding the extent of wrongful harm caused by an action that should count for proportionality and equality analysis. Nonviolent crime might generate more questions in this regard, as it can be unclear how to delimit the full scope of harm that should be included in cases of fraud, bribery, money laundering, drug trafficking, and so on. Many views regarding the metaphysics of causation (particularly views focused on counterfactual or "but for" causation) include more as caused by an action than would be counted by ordinary reflection. (There are many views in which, for example, your birth is a cause of your death.) This issue has been notoriously tricky in tort and criminal law, leading to the not unproblematic use of so-called proximate cause analysis. One question has been whether to see the correct causation standard as one that comes from metaphysics or as one that comes from normative considerations regarding moral responsibility.[36] One thing seems clear: not all consequences that might in some sense be caused by a person's criminal action should count.

There are at least two kinds of potential limitations. First, there are cases in which what happened was not reasonably foreseeable as a result of an action of this kind. Second, there are cases in which the intervening agency of another person is substantial enough to render the previous person "causally innocent," even if it is true that their action was a "causally relevant condition" or a but-for cause of what transpired.

Questions regarding foreseeability and intervening agency make it difficult to provide an exact accounting of the wrongful harm caused by a particular action. Additionally, the consequences of actions and the wrongful harm they cause are ongoing, and questions of punishment have to be answered at particular moments of time, with approximations and estimations made of what wrong-

35 See discussion in note 11.

36 See Honoré, "Causation in the Law."

ful harm appropriately attached to the action still to come. There are different views one might adopt regarding how these questions should be resolved. Still, the very significant wrongful harms that result from nonviolent crimes like fraud, theft, identity theft, and illegal manufacture and trafficking of highly addictive and destructive drugs like heroin and fentanyl are foreseeable and predictable. And in the nondrug cases, there is no question of intervening agency.

If we focus on wrongful harm caused by criminal actions, we will not see a simple, categorical sorting with all violent criminal actions rating worse than all nonviolent criminal actions in terms of wrongful harm caused and risked. But consider a rejection of 7 that maintains that violent criminal action is almost always, although not uniformly, more wrongfully harmful than nonviolent criminal action. What should we make of this weaker claim? There are at least two different ways of trying to evaluate this claim: at the level of act types or at the level of act tokens.

If we focus on act types, we might generate a list of, say, the main one hundred types of criminal actions—the fifty violent ones (murder, manslaughter, rape, robbery, simple assault, etc.) and the fifty nonviolent ones (fraud, espionage, theft, burglary, driving under the influence, etc.). We would then ask: Do the fifty violent crimes generally cause or risk more wrongful harm than the fifty nonviolent ones? To answer this, we could imagine ordering the one hundred types of criminal action from most wrongfully harmful to least wrongfully harmful. One question would be how to do this. Would we somehow have in mind a "normal" or "prototypical" example of each of these actions?[37] Or the most wrongfully harmful instance of each act type? There are difficulties to both these ways of doing things. One worry is that our effort to remain at the level of types will just collapse into a token-level or token-derivative assessment. This might worry us if our motivation for going type level was to avoid collecting jurisdiction-specific statistics or from having our claims be highly relativized to specific places and times.

Let us assume we find some way of fixing on a generic prototype for all one hundred types of actions, and then ordering all one hundred from least wrongfully harmful to most wrongfully harmful. We do something like this when it comes to sentencing. Crimes are sorted in a criminal code with sentencing

37 This is basically what is required under the ACCA, which instructs courts to engage in "ordinary case" analysis. See Begay v. United States, 553 US 137 (2008). To determine whether a given crime is a "violent felony," a court is supposed to disregard the specific facts of the case it is addressing, and (again, somehow!) consider only an "ordinary" case of the crime of which the person was convicted. Because courts do not have empirical data to guide their assessment of what happens in an "ordinary case" of a given crime, this has produced unpredictable, confusing results.

guidelines, with possible sentence ranges indexed to each categorized crime. If we have these one hundred types of criminal actions lined up from least wrongfully harmful to most wrongfully harmful, do we expect that the violent criminal act types are *almost always* more wrongfully harmful than nonviolent criminal act types? We would see something like this:

Least Wrongfully Harmful — Most Wrongfully Harmful

Here "N" stands for a nonviolent criminal action type and "V" stands for a violent criminal action type. But given the wide range of violent and nonviolent criminal action types, that claim also seems implausible. Violent crimes like simple assault and crimes committed while in possession of weapons that are not in fact used are often not all that harmful. And even harmful or potentially harmful violent crimes like aggravated assault and criminal endangerment typically cause or risk physical harm to just one person. Nonviolent crimes like fraud, espionage, embezzlement, racketeering, bribery, and theft can cause great and widespread wrongful harm to many people. Indeed, we may already implicitly acknowledge this. There are many examples in which Crime Type *A* (though violent) is given a shorter sentence range than Crime Type *B* (which is nonviolent). Given that, why should we treat the commission of the violent crime as appropriately triggering the additional penalties and limitations set out earlier? Why should those who are accused or convicted of Crime Type *B* still be eligible for various diversion programs, early parole, probation, and various benefits, while those accused or convicted of Crime Type *A* are not? The answer cannot be that Crime Type *B* is less wrongfully harmful, because it is not.

Another way of interpreting the claim about violent crimes being generally more wrongfully harmful than nonviolent ones is as a claim concerning act tokens, rather than act types. On this interpretation, we would need to look at all the violent crime and nonviolent crime that took place in some jurisdiction over some period of time and sort these individual criminal actions from most wrongfully harmful to least wrongfully harmful. The claim would then be true if the violent token crimes were almost all on the more wrongfully harmful end of the spectrum. And, if true, this claim about act tokens might support a categorical difference in treatment, since definitions of crimes and attachment of sentences might well result in some necessary but imperfect categorization.

There are various problems with this route. One is that it would mean that the claim, and thus the normative support for the practice of treating violent crime in a categorially different way, would be sensitive to these contingent, time-and-

place-relative facts about the wrongful harmfulness of the (say) fifty thousand violent and nonviolent criminal actions that took place over the relevant period of time. But it does not seem as if the way in which existing law draws the categorical distinction is at all responsive to contingent empirical facts about a jurisdiction.

A more significant problem for this way of trying to reject 7: there is no reason to think that a claim like "violent criminal actions are always or almost always more wrongfully harmful than nonviolent criminal actions" will be true when interpreted as about act tokens. The most serious violent crimes are, thankfully, relatively rare—particularly when compared with relatively less serious, much more prevalent violent crimes like simple assault. By far the most common kind of violent crime is simple assault—attacks or attempted attacks without a weapon resulting in either no injury or minor injury.[38] And simple assault will often be less harmful than common nonviolent offenses such as burglary, fraud, tax evasion, money laundering, identity theft, and drug trafficking.

VI

The claims so far have focused on the suggestion that violent criminal actions are more wrongfully harmful than nonviolent crime. It is hard to find a claim that is both (a) strong enough to support the actual categorical violent/nonviolent distinction drawn in law and (b) true.

But perhaps, although violent crime is not always or almost always more wrongfully harmful than nonviolent crime, those who engage in it are categorically *more culpable* than those who engage in nonviolent crime. If so, treating violent criminal action categorically differently than nonviolent criminal action is morally appropriate, not because the acts are more wrongfully harmful, but because those performing them are more culpable. This view would reject the following claim:

8. *No Positive Correlation between Violence and Culpability*: Those who commit violent offenses are no more likely to be fully culpable for offending, nor are they likely to be relatively more culpable, than those who commit nonviolent offenses.

In order to consider the plausibility of the claim that those who engage in violent action are categorically or typically more culpable than those who engage

38 In the United States, of the almost 6.5 million violent crimes in 2018, there were 16,214 homicides, 734,630 rapes/sexual assaults, 573,100 robberies, 1,058,040 aggravated assaults, and 4,019,750 simple assaults. See Morgan and Oudekerk, "Criminal Victimization, 2018"; and Federal Bureau of Investigation, "Uniform Crime Report."

in nonviolent crime, it will be useful to have two broad pictures of culpability or moral responsibility for an action.

The first picture holds that an agent is morally responsible—and correspondingly praiseworthy or blameworthy—only if, or only to the degree that, the agent's actions are under her control. Some who hold such a view do so in an incompatibilist way, maintaining that control is incompatible with determinism.[39] But one might do so in a compatibilist way as well. Or it might be that some forces from the outside impinge upon us, but that these do not fully determine what we do. This would leave us less than perfectly responsible, but still responsible to some degree. More must be said about when an agent's actions are under her control in order to fill out the picture. Call this the *agential control* view.

The second picture is concerned not with *control*, but with what the agent's actions *reveal* about her moral beliefs, attitudes, and values. These are often described as "quality of will" views.[40] Pamela Hieronymi provides a nice statement of this kind of view: "We are fundamentally responsible for a thing . . . because it reveals our take on the world and our place within it—it reveals what we find true or valuable or important."[41] Call this the *agential revelation* view: we are morally responsible for—and correspondingly potentially punishable and blameworthy for—those things that reveal who we are, morally speaking, or what our moral attitudes are like.

On either control or revelation views, culpability will come in degrees, as both *how much control an agent has* in performing an action and *how revealing an action is of who an agent is* are factors that plausibly come in degrees.[42]

We can now ask: On either the agential control or agential revelation views, are those who engage in violent criminal action categorically or almost always *more culpable* than those who engage in nonviolent criminal action? Answering this question might seem to require answering questions regarding the etiology and psychology of violent and nonviolent criminal action. That is a project spanning several disciplines—criminology, sociology, law, psychology—and it is not possible to say anything comprehensive here. I will, however, discuss what I take to be two widespread beliefs—I will call them dogmas—about violent crime that are relevant to the assessment of 8. I will suggest that the available evidence should undermine or at least weaken confidence in them.

39 See, e.g., Strawson, "The Impossibility of Moral Responsibility."

40 For examples of views in this category, see Smith, "Identification and Responsibility"; Hieronymi, "Reflection and Responsibility"; Arpaly, *Unprincipled Virtue*.

41 Hieronymi, "Reflection and Responsibility."

42 For discussion, see Nelkin, "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness."

Here are two common beliefs about violence and those who commit violent actions, even if they are not always formulated quite this explicitly:

> *Dogma of Depravity*: Perpetrators of violence are morally bad people—even evil, depraved. Violent crime is perpetrated by people who have very bad moral characters, people who have disturbed, depraved moral worldviews.
>
> *Dogma of Difference*: Perpetrators of violence are unusually and distinctively bad. They are different from the rest of us. Most of us might be such that we would engage in nonviolent criminal action—if the circumstances were right, if we happened to be around the wrong people at the wrong time. But that is not true of violent criminal action. Only distinctively bad people engage in violent criminal action.

These dogmas do not seem particularly relevant for rejecting 8 if one embraces the agential control model. Indeed, violent action often seems less under an agent's control than nonviolent criminal action; it is hard to see how the agential control picture would lend support to rejecting 8.

But if culpability and moral responsibility are construed on an agential revelation model, these dogmas, if true, might lead us to reject 8. If true, there would be a significant correlation between violent criminal action and greater culpability, relative to nonviolent criminal action. On this view, the commission of violent crimes can be used as evidence connected to a characterological assessment: those who commit violent crimes are somehow in a different, and worse, category of people—they are bad; they have violent *natures*. That does not mean that violent crimes are more wrongfully harmful, but it would mean that those who commit violent crimes are in some important sense more blameworthy, more culpable, and perhaps more justifiably excluded from our broader political and social communities—because of what their violent actions reveal about who they are. We should ask, though, whether the inference from engaging in violent crime to a differentially worse characterological assessment is a good one, and whether these two dogmas are consistent with the available evidence.

There is general reason to be suspicious of characterological assessments, particularly those based on a single action or a few actions. Psychological evidence suggests that our actions are more the product of our situation and environment than we typically believe, and that we are too quick to explain actions as emanating from characterological dispositions. Psychologists have called this the "fundamental attribution error."[43] This might push against the agential rev-

43 For extensive discussion, see Doris, *Lack of Character*.

elation view in general, although there are debates about the psychological evidence here.[44]

Less generally, we might ask whether people who commit violent crime somehow have different and morally worse characters, such that (for example) they are more likely than others to engage in other violent crime. There is little evidence for this. The eminent sociologist and criminologist Randall Collins argues that "it is a false lead to look for types of violent individuals, constant across situations."[45] He goes on: "I want to underline the conclusion: even people that we think of as very violent—because they have been violent in more than one situation, or spectacularly violent on some occasion—are violent only in very particular situations."[46] He argues, backed by extensive empirical evidence, that many instances of individuals who engage in what to an outsider might look like particularly heinous crimes—violent elder abuse, child abuse, spousal abuse—are really quite ordinary people located in particularly difficult, emotional, isolated, stressful situations.[47] The suggestion is that most of us, whatever we think of our characters, might have ended up acting similarly under those conditions. This is not to cast doubt on the culpability of people in these situations (though others might push in that direction), but it is to cast doubt on the view that general character-focused considerations will single out those who have been convicted of violent crimes as particularly bad or particularly culpable. This evidence, at least, suggests that we should reject both the dogma of depravity and the dogma of difference.

Other evidence that inclines against these dogmas comes from the success of alternatives to incarceration for those convicted of violent crimes. If these two dogmas were true, we might expect that little would work to "rehabilitate" or to prevent recidivism of those who have committed violent criminal actions. But that is not what the evidence suggests. Studies have found that participants who were charged with violent crimes or had histories of violence performed as well or better in drug courts and diversion programs than those who were charged with nonviolent crimes or had no such histories of violence.[48] Similarly with mental health diversion programs. Mental health diversion programs that

44 Clarke, "Appealing to the Fundamental Attribution Error."

45 See Collins, *Violence*, 1.

46 Collins, *Violence*, 3.

47 Collins, *Violence*, 137–41.

48 Carey, Mackin, and Finigan, "What Works?"; Carey, Finigan, and Pukstas, "Exploring the Key Components of Drug Courts"; Saum and Hiller, "Should Violent Offenders Be Excluded from Drug Court Participation?"; Saum, Scarpitti, and Robbins, "Violent Offenders in Drug Court."

accept violent offenders have proven to be successful.[49] There is also significant evidence that people tend to "age out" of violence, based on a host of social, developmental, and neurobiological factors.[50] This evidence is hard to reconcile with either of the two dogmas.

Note, too, that to defend a categorical difference in treatment of violent as opposed to nonviolent crime, the dogmas would have to apply categorically—to all violent criminal action, not just the very worst instances of violent criminal action.[51] So, even if the dogmas were true with respect to a certain kind of violent criminal action (e.g., serial rape or serial murder), they might well not be plausible when simple assault is brought into the picture.[52] Related to this, it is plausible that some nonviolent crime is such that it seems to reveal a character that is as evil (if one wants to speak in those terms) as the perpetrators of even particularly heinous violent crime. Think of someone like Bernie Madoff, callously indifferent to the harm he causes or risks. Or think of the former emergency managers and water plant officials in Flint, Michigan, who have been charged with the nonviolent crimes of false pretenses, willful neglect of duty, and conspiracy for their role in misuse of public funds leading to widespread contamination of drinking water, lead poisoning of a generation of children, and at least twelve deaths from Legionnaire's disease. Again, as in the previous section, it starts to look implausible that there will be a categorical difference here that is captured by the violent/nonviolent distinction.

49 See, e.g., Treatment Advocacy Center, "Assisted Outpatient Treatment Laws"; Fisler, "Building Trust and Managing Risk," 587, 602.

50 For discussion, see Ulmer and Steffensmeier, "The Age and Crime Relationship"; Goldstein, "Too Old to Commit Crime?"

51 A similar point can be made, too, against the suggestion that because violent crime can usually not be committed via negligence, whereas other kinds of crimes can be, this would license differential treatment of those convicted of violent crimes across the board. At most this would suggest that some nonviolent offenders, those who offend via negligence, might be less culpable, but that does not line up with the violent/nonviolent categorization generally. And it is, at any rate, controversial whether those who do things *negligently* are less culpable (in this sense of culpable as morally responsible) than those who do things *recklessly* or *intentionally*. That is not uncontroversial on either a control or agential revelation model.

52 Similarly, although in some cases—think of violence or stalking targeted at a particular individual—early release or diversion programs might pose distinct concerns, that is not a reason to see these concerns as presented by every instance of violent crime. The point is not that such concerns can never be appropriately considered; to the contrary, they should be appropriately considered when present, whether the crimes involve violence or not.

VII

The argument I have offered suggests that we should jettison the category of "violent" crime in the criminal law and the law more generally—replacing it with an analysis that orders crimes based on the wrongful harm they cause or risk, rather than on whether they are violent, at least for the purposes of broad sentencing categories and practices of punishment more generally. But this goes against a pretty broad sensibility, which says that violent crime is worse than nonviolent crime and is appropriately treated differently. Here, I want to say a few things that might explain why this sensibility is present, but in a way that suggests that it is in error.

What We Think of When We Think of Violence

Here is a simple explanation for why people think violent crime should be treated differently: the worst violent crimes are truly horrifying and terrifying, in addition to being very wrongfully harmful, and these are what people think of when they think of "violent crime"—even though these are hardly a representative sample of everything that falls under the heading of "violent crime."

Some violent crime is incredibly, terribly wrongfully harmful. For example, violent actions where the person or persons doing the violence is considerably more powerful than the victim(s) of the violence can be harmful not only in the instant physical ways that violence is harmful, but also in structuring relationships of terror and domination, so that the person, the family, or even a whole community is entirely shaped and constrained by violence and the threat of violence in a host of deeply harmful ways.

For example, those who study domestic violence highlight that there are two significantly different forms of intimate partner violence—"situational couple violence" and "patriarchal terrorism"/"intimate terrorism."[53] The first of these is "fairly frequent, not very severe, and practiced rather equally (in modern America) by both males and females."[54] This kind of violence stays within certain parameters, is often symmetrical, rarely escalates over time, and results in injuries in around 3 percent of cases.[55] The second of these, patriarchal terrorism, "is violence used for purposes of control … involving serious physical injury or an ongoing atmosphere of threats; perpetrators are chiefly males, their victims chiefly

53 For the canonical study on this, see Johnson, "Patriarchal Terrorism and Common Couple Violence." For more recent discussion, see Eckstein, "Intimate Terrorism and Situational Couple Violence."

54 Collins, *Violence*, 141.

55 Stets and Straus, "Gender Differences in Reporting Marital Violence."

females."[56] This second kind of domestic violence is rarer than the first, but is generally more salient when many people think of domestic violence. These are the cases that are dramatized in movies and television, and the ones that are covered in the news. Obviously, it should go without saying that all domestic violence is serious and deserves a significant social and legal response; the suggestion here is only that our response ought to be more nuanced and responsive to the facts in particular instances.

Or think of the widespread portrayal of terroristic violence perpetrated by organized crime and criminal gangs. Many of the great works of film and television focus on this kind of violence—think *The Sopranos*, *The Wire*, and so on. Or think of the violent crimes depicted on the many law and crime television shows that are routinely among the highest-rated shows on television (*NCIS*, *NCIS: Los Angeles*, *Law & Order: SVU*). These shows depict a lot of violent crime, but almost always on the "most harmful, most horrific" end of the spectrum of violent crime. So, too, with the violent crime that makes the local or national news.

On the other side, nonviolent crime is only rarely the subject of films or television, and, when it is, the focus is almost always on the wizardry involved in perpetrating the crime rather than on the harm to victims.

This gives us a deeply misleading sense of what *most* violent crime is like, particularly in terms of how harmful it is. Instead of thinking of armed robbery or simple assault that results in little or no physical harm, we think of Jeffrey Dahmer, Tony Soprano, or the horrifying evening news report.

Related to this, Randall Collins details the ways in which we have false beliefs about what violence looks like. He notes that people are "not good at violence," and that most of our beliefs about what violence is like are false. He writes that

> we have been exposed to so much mythical violence. That we actually see it unfolding before our eyes in films and on television makes us feel that this is what real violence is like. Contemporary film style of grabbing the viewers' attention with bloody injuries and brutal aggressiveness may give many people the sense that entertainment violence is, if anything, too realistic. Nothing could be further from the truth. The conventions of portraying violence almost always miss the most important dynamics of violence: that it starts from confrontational tension and fear, that most of the time it is bluster.[57]

Collins notes that most violence is brief, incompetent, and leads to little injury of consequence. (But even this violence still results in assault convictions and

56 Stets and Straus, "Gender Differences in Reporting Marital Violence."

57 Collins, *Violence*, 10.

violent criminal records.) This is not what we expect, having been raised on a steady diet of serial killer stories, NCISes, and *Game of Thrones* violent fantasy stories. The suggestion: as a result of the portrayals of violence that we encounter, we come to have false beliefs about what most violence and violent crime is like, about how harmful it is, and, consequently, about how appropriate it is to treat it categorically differently than nonviolent crime.

The Harm We Can See

Another possible (and non-rival) explanation for why we may feel that violent crime is different than nonviolent crime: the harms from the most salient examples of violent crime are easy to see and to quantify. It is easy for us to understand the exact harm of violent crime, certainly the most proximate harms. With many kinds of nonviolent crime—financial crime, cybercrime, fraud, embezzlement—it may be hard to even understand what the crime was, let alone the harms that it caused. We should not infer from this, however, that these crimes are harmless. Nothing could be further from the truth.

If one wanted to consider a psychological or evolutionary story here, one could also note that violence is one of the oldest and most intimately familiar ways in which we can harm each other. We might well expect to have more deeply ingrained attitudes about violently caused harms than about harms caused in nonviolent ways. In the same way that it is plausible that our ideas about morality did not originally develop to take into account the ways in which we might *help* or *save* those physically very distant from us, so, too, it is plausible that our ideas about morality did not originally form to take into account ways in which we might badly *harm* those physically very distant from us, or those whom we may never see or meet.

These attitudes and intuitions about morality and harm might not have adequately updated to the modern, globally interconnected world in which we live. We may pay more attention to the local harms that might be caused through, say, physical violence, and not enough to the distant harms we can cause through, say, destroying the pensions of thousands of people through fraud and illegal market manipulation.[58]

58 A related possibility, which might push back against the diagnosis that this is always a sign of error: it might be that, for some violent crimes and some nonviolent crimes, the violent crimes are such that the wrongful harms are less diffuse (more concentrated on a few individuals) and the nonviolent crimes have wrongful harms that are more diffuse (spread out in relatively small increments across many individuals). There are theories on which this kind of relative diffuseness might appropriately make a moral difference, even in cases in which the total wrongful harm might be equivalent or comparable. This would suggest that an additional dimension to the wrongful harm analysis would be appropriate, but it would

Class, Race, and Violence

A final thought about why we might see violence differently. Here we must ask who "we" are. It is plausible that many of the attitudes about how bad violence or violent crime is, or how bad those who engage in violence are, have a class and possibly racial dimension to them.

First, if one rarely encounters violence, then the myths about violence and the Hollywood portrayal of violence will more dramatically affect one's view about what most violence and violent crime are like. And one directly encounters less violence as one moves up the socioeconomic ladder (which is not to say that it disappears). If we accept Randall Collins's explanation that much violence is the product of situational factors like stress, powerlessness, and isolation, we should expect that those in certain socioeconomic environments may more often engage in and witness violence, without this meaning that those people are morally worse than those who, say, engage in nonviolent crime. And this will be more familiar to those who have some personal experience at levels of lower socioeconomic standing.

Second, if popular views about crime and particularly violent crime are biased and warped by presentations in media and background racism and classism, they can also contribute to how violence is understood and in particular the extent to which the two dogmas discussed above are accepted. As one of the leading experts on violence and law suggests, "the racialization of violent crime has likely had more than a little to do with the increasing tendency to understand criminal violence as a product of offenders' characters, not of the situations in which they find themselves."[59]

Third, use of violence, even amounting to violent crime, can be defensive or protective, particularly for those who do not expect reliable police protection. Elijah Anderson discusses the "code of the street" that arises because of a lack of reliable police protection, along with the view that the police are prejudiced against everyone in particular neighborhoods and of particular races, so that a person calling the police is as likely to end up arrested as the perpetrator.[60] Under these circumstances, it becomes rational for individuals to demonstrate their ability to defend themselves by displaying a willingness to use violence if necessary. Engaging in violence on occasion may even be necessary, but those familiar

---

not vindicate sorting along violence/nonviolence, as diffuseness of harm caused does not line up with violent or nonviolent crimes, particularly once secondary harms are factored in.

59 Sklansky, *A Pattern of Violence*, 62.

60 Anderson, *Code of the Street*. Gruen, Meikle, and Pierce develop complementary ideas and arguments in "Destabilizing Conceptions of Violence."

with these environments will not see this use of violence as supporting either of the two aforementioned dogmas.

Fourth, and amplifying the first three points, if we have an elite political class of legislators (as most electoral democracies in fact have), we will have people making decisions about violent crime who are themselves largely unfamiliar with violence, and whose sense of it comes from film, television, sensationalistic news stories, and possibly racist and classist biases. They—and the prosecutors and judges who also comprise this elite political class—also have political incentives to sensationalize the danger and violent crime that exists.[61] And this elite political class, supported by the socioeconomic elite, will also sometimes have incentives not to want attention turned toward so-called white-collar crimes like tax fraud, securities fraud, and other potentially very wrongfully harmful but nonviolent crime.

VII

Although it is difficult to offer precise quanta of the wrongful harm caused by particular crimes or by typical crimes in various categories, this is the kind of inquiry we should be engaged in—just as those involved in public health and the allocation of medical resources and interventions have to think about the harms and benefits that are likely to result from various actions and options. Rather than letting sentencing ranges be set by political whims or manipulated emotional responses, we should be having serious, evidence-based conversations about the wrongful harmfulness of crime and the morally appropriate responses to crime.[62]

As noted above, the felony/misdemeanor classifications, as well as intricate and complex criminal codes and sentencing guidelines, already try to categorize and distinguish tiers and categories of criminal actions. A focus on wrongful harm allows this to be done in a more principled way, allowing intelligible comparisons across kinds of crimes, including crimes across the violent/nonviolent divide. A simple system might have five different categories, Category One to Category Five (like hurricanes), corresponding with how much wrongful harm was caused or risked by the particular criminal action, or, alternatively, by a typical criminal action defined by these specific elements. Things will inevitably

61 For classic discussion of these issues, see Stuntz, "The Pathological Politics of Criminal Law," 505, 510.

62 Part of this conversation is already under way, as the public good/public health justification of criminal law is offered to supplant more punitive or retributive justifications. See Chiao, *Criminal Law in the Age of the Administrative State*.

become more complicated and there are many questions to be addressed. But it seems that we should endorse:

9. *Better Categories Possible*: There are usable categorizations of actions that do a better job sorting actions by their wrongful harmfulness than the violent/nonviolent categorization.

If the argument of the article is successful, then we should embrace these categories—categories structured around the idea of wrongful harm—and jettison our misplaced focus on violence. Doing this might also result in us addressing some of the true deep roots of the problem of mass incarceration and enable a more effective, less devastating response to the problems of crime and wrongful harm. And doing so would not be heading out into uncharted territory; indeed, references in law to "violent crime" or "violence" are actually a recent development, beginning in the late 1960s.[63] Categories in law are important and useful, but they should track what matters, morally. We can do better without "violence" as a central category in law.[64]

Rutgers University–New Brunswick
alex.guerrero@rutgers.edu

REFERENCES

Alexander, Michelle. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. New York: The New Press, 2010.

Anderson, Elijah. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York: W. W. Norton and Company, 1999.

63 Sklansky, *A Pattern of Violence*, 45–55. He notes that "the sharp distinction between violent and nonviolent crimes, and the great weight placed on that distinction, are modern developments, roughly half a century old.… References to 'violent crime' did not become common in American discourse until the 1970s. Before the late 1960s, in fact, references to 'violent crime' were less common than references to 'infamous crime'—a legal category that … was never terribly important, and that in no way tracked the line now drawn between violent and nonviolent offenses" (45).

64 Thanks to the many people who have helped me think through these issues and provided comments on versions of this paper over the years, including Kristen Bell, Elizabeth Harman, Thomas Hurka, Douglas Husak, Adam Kolber, Jennifer Lackey, Christopher Lewis, David Plunkett, Alice Ristroph, Patrick Tomlin, and Gary Watson; audiences at Brooklyn Law School, the 2017 New Directions in Philosophy of Law Conference at Oxford, the Princeton Workshop in Normative Philosophy, the USC Conceptual Foundations of Conflict Project, and the University of Toronto; and several anonymous referees.

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

Braithwaite, John, and Philip Pettit. *Not Just Deserts: A Republican Theory of Criminal Justice*. Oxford: Oxford University Press, 1990.

Carey Shannon M., Michael W. Finigan, and Kimberly Pukstas. "Exploring the Key Components of Drug Courts: A Comparative Study of 18 Adult Drug Courts on Practices, Outcomes and Costs." NPC Research, August 2008. https://npcresearch.com/publication/exploring-the-key-components-of-drug-courts-a-comparative-study-of-18-adult-drug-courts-on-practices-outcomes-and-costs-3.

Carey Shannon M., Juliette R. Mackin, and Michael W. Finigan. "What Works? The Ten Key Components of Drug Court: Research-Based Best Practices." *Drug Court Review* 8, no. 1 (2012): 6–42.

Chiao, Vincent. *Criminal Law in the Age of the Administrative State*. Oxford: Oxford University Press, 2018.

Clarke, Steve. "Appealing to the Fundamental Attribution Error: Was It All a Big Mistake?" In *Conspiracy Theories: The Philosophical Debate*, edited by David Coady, 129–32. London: Routledge, 2018.

Collins, Randall. *Violence: A Micro-sociological Theory*. Princeton: Princeton University Press, 2008.

Ditton, Paula M., and Doris James Wilson. "Truth in Sentencing in State Prisons." Bureau of Justice Statistics, January 1999. https://bjs.gov/content/pub/pdf/tssp.pdf.

Doris, John. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press, 2002.

Duff, Antony. *Punishment, Communication, and Community*. Oxford: Oxford University Press, 2001.

Duus-Otterström, Göran. "Why Retributivists Should Endorse Leniency in Punishment." *Law and Philosophy* 32, no. 4 (July 2013): 459–83.

Eckstein, Jessica J. "Intimate Terrorism and Situational Couple Violence: Classification Variability Across Five Methods to Distinguish Johnson's Violent Relationship Types." *Violence and Victims* 32, no. 6 (December 2017): 955–76.

Edwards, James. "Theories of Criminal Law." *Stanford Encyclopedia of Philosophy* (Summer 2013). http://plato.stanford.edu/entries/criminal-law/.

Federal Bureau of Investigation. "Uniform Crime Report: Crime in the United States, 2018—Murder." Fall 2019. https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/murder.

Fisler, Carol. "Building Trust and Managing Risk: A Look at a Felony Mental

Health Court." *Psychology, Public Policy, and Law* 11, no. 4 (December 2005): 587–604.

Goldstein, Dana. "Too Old to Commit Crime?" *New York Times*, March 20, 2015. https://www.nytimes.com/2015/03/22/sunday-review/too-old-to-commit-crime.html.

Gruen, Lori, Clyde Meikle, and Andre Pierce. "Destabilizing Conceptions of Violence." In *The Ethics of Policing and Imprisonment*, edited by Molly Gardner and Michael Weber, 169–86. London: Palgrave Macmillan, 2018.

Harman, Elizabeth. "Harming as Causing Harm." In *Harming Future Persons: Ethics, Genetics, and the Nonidentity Problem*, edited by Melinda A. Roberts and David T. Wasserman, 137–54. Berlin: Springer, 2009.

Hieronymi, Pamela. "Reflection and Responsibility." *Philosophy and Public Affairs* 42, no. 1 (Winter 2014): 3–41.

Honoré, Antony. "Causation in the Law." *Stanford Encyclopedia of Philosophy* (Fall 2010). https://plato.stanford.edu/archives/win2010/entries/causation-law.

Howard, Marc Morjé. *Unusually Cruel: Prisons, Punishment, and the Real American Exceptionalism*. Oxford: Oxford University Press, 2017.

Johnson, Michael P. "Patriarchal Terrorism and Common Couple Violence: Two Forms of Violence Against Women." *Journal of Marriage and the Family* 57, no. 2 (May 1995): 283–94.

Kim, Catherine. "Why People Are Being Released from Jails and Prisons during the Pandemic." Vox, April 3, 2020. https://www.vox.com/2020/4/3/21200832/jail-prison-early-release-coronavirus-covid-19-incarcerated.

Morgan, Rachel, and Barbara Oudekerk. "Criminal Victimization, 2018." *Bureau of Justice Statistics Bulletin*, September 2019. https://www.bjs.gov/content/pub/pdf/cv18.pdf.

Nelkin, Dana Kay. "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness." *Noûs* 50, no. 2 (2016): 356–78.

Outlaw, Lucius III. "Time for a Divorce: Uncoupling Drug Offenses from Violent Offenses in Federal Sentencing Law, Policy, and Practice." *American Journal of Criminal Law* 44, no. 1 (2016): 49–70.

Pfaff, John. *Locked In: The True Causes of Mass Incarceration—and How to Achieve Real Reform*. New York: Basic Books, 2017.

Prieto, Luis, and Jose Sacristán. "Problems and Solutions in Calculating Quality-Adjusted Life Years (QALYs)." *Health and Quality of Life Outcomes* 1, no. 80 (December 19, 2003).

Ristroph, Alice. "Criminal Law in the Shadow of Violence." *Alabama Law Review* 62, no. 3 (2011): 571–622.

Saum, Chistine A., and Matthew L. Hiller. "Should Violent Offenders Be Excluded from Drug Court Participation? An Examination of the Recidivism of Violent and Nonviolent Drug Court Participants." *Criminal Justice Review* 33, no. 3 (September 2008): 291–307

Saum, Christine A., Frank R. Scarpitti, and Cynthia A. Robbins. "Violent Offenders in Drug Court." *Journal of Drug Issues* 31, no. 1 (January 2001): 107–28.

Shiffrin, Seana Valentine. "Harm and Its Moral Significance." *Legal Theory* 18, no. 3 (September 2012): 357–98.

Silva, Lanhy R. "Clean Slate: Expanding Expungements and Pardons for Non-Violent Federal Offenders." *University of Cincinnati Law Review* 79, no. 1 (October 2011): 155–205.

Sklansky, David Alan. *A Pattern of Violence: How the Law Classifies Crimes and What It Means for Justice*. Cambridge, MA: The Belknap Press, 2021.

Smith, Angela. "Identification and Responsibility." In *Moral Responsibility and Ontology*, edited by Ton van den Beld, 233–46. Dordrecht, Netherlands: Kluwer Academic Publishers, 2000.

State of New Jersey. "Governor Murphy Signs Major Criminal Justice Reform Legislation." December 18, 2019. https://nj.gov/governor/news/news/562019/approved/20191218a.shtml.

Stets, Jan E., and Murray A. Straus. "Gender Differences in Reporting Marital Violence." In *Physical Violence in American Families*, edited by Murray Straus and Richard Gelles. London: Routledge, 1990.

Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75, nos. 1–2 (August 1994): 5–24.

Stuntz, William J. "The Pathological Politics of Criminal Law." *Michigan Law Review* 100, no. 3 (December 2001): 506–600.

Treatment Advocacy Center. "Assisted Outpatient Treatment Laws." 2017. https://www.treatmentadvocacycenter.org/component/content/article/39.

Ulmer, Jeffrey T., and Darrell Steffensmeier. "The Age and Crime Relationship: Social Variation, Social Explanations." In *The Nurture Versus Biosocial Debate in Criminology: On the Origins of Criminal Behavior and Criminality*, edited by Kevin M. Beaver, J. C. Barnes, and Brian B. Boutwell, 377–96. Los Angeles: SAGE Publications, 2014.

United States Sentencing Commission. *Report to the Congress: Career Offender Sentencing Enhancements*. August 2016. http://www.ussc.gov/sites/default/files/pdf/news/congressional-testimony-and-reports/criminal-history/201607_RtC-Career-Offenders.pdf.

Urbina, Ian. "Despite Red Flags about Judges, a Kickback Scheme Flourished."

New York Times, March 27, 2009. https://www.nytimes.com/2009/03/28/us/28judges.html.

Von Hirsch, Andrew. "Proportionate Sentencing: A Desert Perspective." In *Principled Sentencing*, 3rd ed., edited by Andrew von Hirsch, Andrew Ashworth, and Julian Roberts, 115–25. Oxford: Hart, 2009.

Wines, Michael. "Kentucky Gives Voting Rights to Some 140,000 Former Felons." *New York Times*, December 12, 2019. https://www.nytimes.com/2019/12/12/us/kentucky-felons-voting-rights.html.

Wringe, William. *An Expressive Theory of Punishment*. London: Palgrave Macmillan, 2016.

Journal of Ethics and Social Philosophy
VOL. 22, NO. 1 · MAY 2022
https://doi.org/10.26556/jesp.v22i1.1550

MORAL DECISION GUIDES

COUNSELS OF MORALITY OR COUNSELS OF RATIONALITY?

Holly M. Smith

MORAL AGENTS, wishing to use their moral codes to guide their decisions, are often impeded by lack of information about the circumstances and consequences of their actions. Mayor Katya's moral code directs her, in cases of financial retrenchment, to reduce the city's budget in the least damaging way. But what if she is uncertain whether it would be less damaging to cut the education budget or the public transportation budget? Many moral philosophers, contemplating questions of this sort, have concluded that the best moral theories (sometimes called "dual ought" theories) should have two tiers: a top tier stating what is *objectively* right and wrong, and a lower tier consisting of one or more decision guides designed to provide advice about what is *subjectively* best for agents to do in light of their uncertainty about what is objectively best to do. Such an agent's blameworthiness, if she does what is objectively wrong, depends in part on whether she also does what she believes to be subjectively wrong. Decision guides, then, have a strong link to blameworthiness.

In *Making Morality Work*, I recently described the kind of structure that such a two-tier moral code should exhibit.[1] This structure requires a large hierarchically organized set of decision guides to accommodate the many different kinds of uncertainty. The decision guides themselves recommend (or proscribe) acts as morally choice worthy, choice mandated, or choice prohibited. The subjectively right act is, in the simplest cases, the act recommended by the most highly ranked decision guide that is usable by the agent and suitable to her objective theory. These decision guides might include such user-friendly principles as "Do what is most likely to be objectively right," "Do what will maximize expected value," or "Do what your predecessor in office did in similar circumstances." My argument for this proposal starts from what I call the "Usability Demand," which requires that any acceptable moral theory must be usable by every agent on each occasion for decision making. In *Making Morality Work*, I argue that an

1 Smith, *Making Morality Work*.

acceptable moral theory can satisfy this demand, even if it is not always *directly* usable, so long as it is *indirectly* usable by means of a suite of appropriate decision guides that together would provide any moral agent with guidance for what to do when she wants to apply her moral theory.[2]

The issue before us in this paper is what the nature is of these decision guides: Are they *moral* principles of a certain sort, or are they principles of *rationality*, used here in the context of moral decision making? In my book I left this question open, and aim to resolve it in this paper. I will start by examining a recent attempt to show that these decision guides are prescriptions of rationality, not of morality.

1. Peter Graham's View That Subjective Oughts Are Rational Oughts, Not Moral Oughts

Peter Graham is currently the most energetic proponent of the view that decision guides are generic pragmatic principles—or, as he puts it in later writing, are principles of rationality.[3] Graham is a staunch advocate of objectivism, the view that, roughly speaking, a person's moral obligations depend on all the facts about her situation except the facts concerning her beliefs or evidence about her situation.[4] Given his commitment to objectivism, he must explain why (as he puts it) a morally conscientious person ought to perform an act that she believes is objectively wrong. This occurs in the famous Dr. Jill case (presented in table 1), in which John is afflicted with a minor but not trivial skin complaint. Treatment with Drug *A* would completely cure John, treatment with Drug *B* would partially cure him, treatment with Drug *C* would kill him, and giving him no treatment at all (*D*) would leave him permanently incurable. Dr. Jill must choose which treatment to use. Unfortunately, although Jill knows *B* would partially cure John and not treating him (*D*) would leave him permanently incurable, her evidence indicates only that Drug *A* and Drug *C* would have opposite effects, and that for each of Drug *A* and Drug *C* there is a 50 percent chance that the drug would completely cure John and a 50 percent chance that the drug would kill him.[5]

2 Smith, *Making Morality Work*.

3 Graham is far from the only adherent to this view. A recent example is offered by Muñoz and Spencer, who say, "For . . . uncertain agents, the subjective 'ought' is the proper guide to action; it is the 'ought' of rationality" ("Knowledge of Objective 'Oughts,'" 77).

4 Graham, "In Defense of Objectivism about Moral Obligation," 88–89. This is only roughly correct, but will do for purposes of this paper.

5 This version of the case, including the values in table 1, is from Zimmerman, *Living with Uncertainty*, 17–20. Zimmerman follows the case description in Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 462–63. This type of case originated earlier, in Regan, *Utilitarianism and Co-operation*, 264–65.

Virtually everyone, including Graham, agrees that Jill, if she is a conscientious person, would choose Drug *B* for her patient John.

Table 1. Dr. Jill

Act	Value in Situation 1 ($p = 0.5$)	Value in Situation 2 ($p = 0.5$)	Actual Value	Expected Value
A	50*	−100	50*	−25
B	40	40	40	40*
C	−100	50*	−100	−25
D	0	0	0	0

Note: Asterisks indicate the best outcome in each column.

One way to explain this—my preferred way—is to say that while Drug *A* would be *objectively* right, Drug *B* would be *subjectively* right, and Jill ought subjectively to treat John with Drug *B*. Graham is highly allergic to the idea that there can be dual competing oughts—the objective and the subjective ought—partly because he believes that a decision maker needs an unequivocal answer to the questions of what she should do. It is no help to tell her that she ought to do *A* but in another sense she ought to do *B* instead.[6] Graham argues instead that objective wrongs can be more or less serious, that killing John is a much more serious wrong than merely partially curing him, and that objectivism recommends to a conscientious agent like Jill that she not risk the more serious wrongs that Drugs *A* and *C* might incur, but instead prescribe the objectively wrong but less risky Drug *B*.

Despite Graham's argument, objectivism itself actually has nothing to say about what agents ought to do about risk in situations of uncertainty. It confines itself to prescriptions based on the actual facts, not on the agent's credences, probability estimates, or evidence concerning those actual facts. So how does Graham conclude that objectivism prescribes the less risky Drug *B* to Dr. Jill? He answers this question by arguing that in saying "Jill ought to use Drug *B*" we are not using the "ought" of moral obligation, but rather a different kind of "ought"—the *pragmatic* ought that is associated with ends and means.[7]

Graham holds that there are two such pragmatic "oughts"—a subjective one and an objective one. Clarifying this, he says:

> According to the objective pragmatic "ought" ($\text{ought}_{\text{pragmatic (objective)}}$), a person ought to do something just in case doing it will bring about the

6 Graham, "In Defense of Objectivism about Moral Obligation," 95.

7 Graham, "In Defense of Objectivism about Moral Obligation," 103.

> outcome, among the various outcomes from which she is choosing, she most prefers relative to her goals in acting.... According to the subjective pragmatic "ought" ($\text{ought}_{\text{pragmatic (subjective)}}$), a person ought to do something just in case, roughly, doing so is the output that results from inputting into ... the correct decision theory ... the agent's preference, and subjective probability, functions.[8]

So Graham concludes that if we say Jill ought to prescribe Drug *B*, we are not asserting anything about what Jill ought *morally* to have done, but instead are expressing a view about what she *pragmatically* (subjectively) ought to do.[9] It is the very same ought, he says, that the devil might employ in saying to himself "I ought to cause a plague instead of an earthquake because that will cause more pain and suffering."[10] With a little imagination we can derive from this view a claim that all decision guides are really principles prescribing what it is pragmatically—or rationally—subjectively obligatory for an agent to do. They are not moral principles, but instead more general normative principles available to be used in connection with any type of decision, moral or otherwise, in which the agent attempts to decide what to do in light of her personal goals and epistemic limitations.

In the context of Graham's discussion this seems to be an unhappy proposal. He has rejected the dual-oughts view about moral obligation, claiming that there cannot be both an objective and a subjective moral ought, since no agent can make a decision if faced with two conflicting types of "oughts."[11] But why then should we accept the dual-oughts view about *pragmatic* oughts? On this view agents will still be faced with two conflicting types of pragmatic oughts (the pragmatic objective ought and the pragmatic subjective ought), and will be just as much at sea as an agent faced with two conflicting moral oughts.[12] Even

8 Graham, "In Defense of Objectivism about Moral Obligation," 103.

9 Graham actually sets this up as something Jill says about herself, but for brevity I have put the words in our mouths.

10 Graham, "Moral Conscientiousness and the Subjectivism/Objectivism Debate about Moral Wrongness," 28.

11 See also Graham, "Avoidable Harm," 191n31.

12 Perhaps this could be resolved by Graham's switching to a single subjective pragmatic theory that consists of one principle, "One subjectively ought to maximize expected value," that generates recommendations both when the agent faces uncertainty and when the agent believes there is a 1.0 chance that a certain act would maximize value. But if Graham were willing to take this tack for pragmatic theories, why not take it for moral theories? Moreover, such a theory runs into another problem, which is that it is hardly clear that this theory would provide guidance for every agent, since relatively few are in a position to estimate which act would maximize expected value. Moral theories need to be supplemented by many decision guides, not just one.

worse, agents like Dr. Jill will be faced with yet a further conflict between the *pragmatic* subjective ought and the *moral* objective ought.[13] By Graham's own lights, this should not qualify as an acceptable resolution to the agent's problem about what it is best to do.

There is another problem as well. Graham considers a case in which Jill has a momentary lapse in moral conscientiousness and prescribes Drug *A*. Before learning the outcome, she regains her conscientiousness, and says to herself, "I ought to have prescribed Drug *B* instead."[14] Graham maintains that this is a pragmatic ought, not a moral ought. But he envisions an objector protesting to his appeal to pragmatic oughts as follows: "The 'ought' in Jill's thought about [what she ought to have done] is *not* a *pragmatic* 'ought.' It's clearly a moral 'ought.' It has a distinctly moral cast to it."[15] Graham responds as follows: "The 'ought' in Jill's thought indeed has a moral cast to it. But that is certainly consistent with its being a pragmatic 'ought.' If the pragmatic 'ought' is, as I have indicated, an 'ought' relative to the goals of the agent in question, then it is natural that that 'ought' take on the cast of the goals to which it is relativized. The moral flavor of the 'ought' in Jill's thought, then, is easily explained by the fact that the goals Jill has . . . are the . . . thoroughly moral goals of the morally conscientious person."[16]

There is a major difficulty with this response. Graham initially defines his "pragmatic ought" in terms of what a person most prefers *relative to her goals* in acting.[17] However, as observers we can identify what a person ought morally to do—either objectively or subjectively—even if acting morally is *not* one of her goals. Even if Jill were not a conscientious agent, we would say that, in light of her uncertainty about the effects of the drugs, she subjectively ought morally to choose Drug *B*.[18] Graham's pragmatic oughts fail to reflect the essential feature of moral oughts that they are not hostage to the agent's own preferences. His proposal thus fails to capture what we mean when we say "Jill ought to use Drug *B*."[19]

In later work Graham apparently realizes that relativization of his pragmatic ought to the agent's actual goals leads to trouble. Accordingly, he revises his view.

13 Since Dr. Jill cannot know which particular action she objectively ought to do, a more accurate statement is that she will be faced by a conflict between her pragmatic subjective ought and her moral objective prohibition.

14 Graham, "In Defense of Objectivism about Moral Obligation," 103.

15 Graham, "In Defense of Objectivism about Moral Obligation," 104.

16 Graham, "In Defense of Objectivism about Moral Obligation," 104.

17 Graham, "In Defense of Objectivism about Moral Obligation," 103.

18 Furthermore, we can identify what would be morally wisest for an agent to do, even though we, as observers, have no goal that she act morally.

19 Some of the material in the foregoing paragraphs of this section appeared originally in Smith, "The Zimmerman-Graham Debate on Objectivism versus Prospectivism."

He now says that in saying Jill ought to use Drug *B*, we are saying that giving Drug *B* is what she *rationally* ought to do *if* she had the set of goals "we think ideally she ought to have, i.e., the set of goals of the morally conscientious person."[20] He continues: "Given that decision theory is a theory of what it is rational to do, then it follows that [the rational] 'ought' is governed by the rules of some acceptable decision theory."

But this revised proposal is also problematic. Graham claims that when we say Jill ought to give her patient Drug *B*, we are saying she rationally ought to do so, relative to the set of goals we think she ideally ought to have. The first problem here is that this new notion of "rationality" is a *substantive* normative notion, not a pure concept of rationality. What goals are those that Jill ought to have? There is no single set of ideal goals. Instead, there are many different ideal goals. There is the ideal of perfect financial achievement, the ideal of perfect athletic performance, the ideal of perfect altruism, the ideal of perfect prudence, the ideal of a perfect balance between morality and prudence, and so forth. As we normally evaluate Jill's case, we consider what she ought *morally* to do, not what would increase her financial worth or improve her athletic performance. If Graham is right that we have an ideal in mind, it is a moral ideal. For our statement "Jill ought to give Drug *B*" to accurately convey this ideal, the ideal must be overtly (or contextually) expressed as part of what we mean. This means that our statement "Jill ought to give Drug *B*" is not a statement of a *purely* rational ought, as Graham claims. Instead, if it incorporates an ideal, it expresses a *moral* ideal, just as when her financial advisor says that "Jill ought to rebalance her portfolio to include more bonds" he means to express a financial ideal. Claiming that we are merely expressing some all-purpose ideal through the notion of what an agent rationally ought to do fails to capture what we actually mean, which could only be expressed by saying something like, "Morally speaking, Jill ought to give Drug *B*."

My conclusion is that Graham has not argued successfully that we can interpret moral decision guides as principles spelling out what it is pragmatically or rationally obligatory to do.

2. THE CASE FOR DECISION GUIDES AS PRINCIPLES OF MORALITY

My own theory, as I mentioned at the start, says that an acceptable moral theory

20 Graham, "Moral Conscientiousness and the Subjectivism/Objectivism Debate about Moral Wrongness," 26. He actually states this as "given that she has the set of goals we think ideally she ought to have, i.e., the set of goals of the morally conscientious person." This still suggests we think she has these goals, whereas he needs a hypothetical to deal with the case in which the agent has alternative goals.

must have two tiers, a top tier consisting of the principles of objective rightness, and a lower tier consisting of a set of decision guides that are designed to enable agents to make decisions by reference to their objective moral theory via applying it indirectly through one of its decision guides. The idea is that if the moral theory's principle of objective rightness says, for example, "It is obligatory to maximize value," then an agent who applies this by following a decision guide that says, "It is choice mandated to maximize *expected* value" may be applying her moral code indirectly through use of this decision guide. But what actually counts as applying one's moral code indirectly is somewhat more complicated than this suggests.

For an agent to apply her moral code indirectly through use of a decision guide, the agent must employ the guide *because* she believes that it has an appropriate relationship to that objective account of right or wrong. Consider Liz, who believes act utilitarianism to be the correct account of objective right and wrong. She believes the expected utility rule to be the highest appropriate decision guide relative to act utilitarianism that she can presently use, and hence derives a prescription, via the expected utility rule, to perform act *A* as subjectively right relative to act utilitarianism. Liz counts as someone who "indirectly" applies act utilitarianism in deciding what to do, in part because she appropriately connects her governing moral theory to her decision guide and thence to her choice of action.

By contrast, consider Ned, who believes some act *A* would have greater expected utility than any other option, and derives from this a prescription to perform act *A*. Given only this information we cannot conclude that Ned has indirectly applied act utilitarianism as a theory of objective moral status. Ned might believe, for example, that the expected utility rule just *is* the sole account of right and wrong—he might be following Zimmerman's "Prospectivism" rather than act utilitarianism.[21] In this case he would correctly understand himself to be *directly* applying his moral code, not indirectly applying act utilitarianism.

Or consider Katya, the aforementioned mayor who needs to decide whether to reduce city expenditures by cutting the education or the transportation budget. She knows that her predecessor, facing a similar issue, cut the transportation budget and was nonetheless reelected. Katya derives a prescription to cut the transportation budget from a decision guide recommending doing what one's predecessor did in similar circumstances. Katya believes this decision guide is appropriate for ethical egoism. But she *also* believes it is appropriate for someone trying to do what is prudentially best. Given only this information, we cannot determine whether Katya is indirectly applying ethical egoism or prudence.

21 Zimmerman, *Living with Uncertainty*.

For her to count as indirectly applying ethical egoism, she must use this decision guide *because* she believes it is appropriate to ethical egoism.

Part of an agent's employing a guide because she believes that it has an appropriate relationship to a given objective account of right or wrong is her believing that the guide delivers the *kind* of recommendation appropriate to that account. This is trickier than it may at first appear. An appropriate guide must specify not only the right kind of value to be fostered (say, utility, or honoring rights), but also the right kind of normative recommendation. Decision guides, as I see them, recommend (or proscribe) acts as choice worthy, choice mandated, or choice prohibited. The "valence" of the decision guide is clearly important. It would obviously be a mistake to try to indirectly apply the objective principle, "One *ought* to maximize utility" by following a decision guide stating, "It is choice *prohibited* to maximize expected utility."[22] And the normative type of the choice mandate (or prohibition) must be appropriate as well. Consider an artist, Raul, who is uncertain whether to color a certain portion of his painting orange or blue. He wants to apply the color that will maximize aesthetic value. Not being sure which color is aesthetically best, he tries to indirectly apply his objective rule by following a decision guide that says "It is choice mandated, financially, to apply the color that will maximize expected aesthetic value." Raul might believe that he is indirectly applying his aesthetic principle, but in fact he has gone astray, since he is using a decision guide incorporating a *financial* choice mandate in order to indirectly apply an objective principle requiring him to maximize *aesthetic* value. There are different kinds of objective "oughts"—moral, legal, prudential, epistemological, financial, aesthetic, and so on—and the different objective "oughts" call for different kinds of choice mandates and different kinds of subjective "oughts." For a comprehensive normative theory to provide legitimate guidance to its users, it must include decision guides whose choice mandates and subjective oughts are properly tied to the theory's type of objective oughts. This means that the theory's decision guides must specify whether an act is morally choice mandated, legally choice mandated, epistemologically choice mandated, prudentially choice mandated, and so forth. Only if the decision guides of a *moral* theory provide *morally* choice-mandated recommendations will they form the proper basis for an agent, in trying to apply the objective moral principle indirectly, to be able to derive a recommendation for an action as subjectively morally obligatory, which is what the agent needs.

Such considerations suggest that we should adopt something like the following definition for what it is to be able to use a moral theory indirectly in making a decision:

22 This can be subtle. A decision guide stating, "It is choice prohibited to minimize expected utility" might be appropriate.

Definition 1: Ability to indirectly use a moral principle in the core sense to decide what to do: An agent *S* who is uncertain at t_i which of the acts she could perform (in the epistemic sense) at t_j is prescribed by *P* (a principle of objective moral obligation or rightness) is nonetheless able at t_i to *indirectly* use *P* in the core sense to decide at t_i what to do at t_j *if and only if*

a. *S* believes at t_i of some act *A* that *S* could perform *A* (in the epistemic sense) at t_j;
b. at t_i *S* believes of act *A* that it is prescribed as morally choice mandated or choice worthy for performance at t_j by the highest-ranked decision guide (relative to *P*) usable by her at t_i; and
c. if and because *S* wanted, all things considered, to use principle *P* for guidance at t_i for an act performable at t_i, then her beliefs together with this desire would lead S to derive a prescription for *A* as subjectively morally obligatory (or as subjectively morally right) for her relative to *P*.[23]

Definition 1 defines an agent's ability to indirectly use a moral principle in making a decision. Important features of this definition are its crucial stipulations (1) that the agent believes of some act *A* that it is prescribed as *morally* choice mandated (or choice worthy) by the highest-ranked decision guide, relative to her governing moral theory, that she can use, and (2) that she would derive a prescription for the act as subjectively *morally* obligatory (or right) for her, relative to *P*. If her belief or derived prescription would be phrased in terms of some other type of normativity, such as legal or prudential choice worthiness or subjective rightness, then she has gone astray.

Our conclusion should be that a comprehensive moral theory must include decision guides that enable agents to indirectly derive prescriptions from the theory. But these decision guides must be phrased in normative terms that correlate with the normative nature of the moral theory: they must be phrased in terms of moral choice worthiness, not choice worthiness of some other type of normativity. It follows that decision guides are counsels of morality, not counsels of prudence or legality or even rationality. They qualify as *moral* principles serving as fully fledged components of a comprehensive moral theory, not mere principles of rationality.

23 This definition is a simplified version, with slight changes, of one proposed in Smith, *Making Morality Work*, 280, definition 12.1. The changes involve inserting the word "morally" at key points, a necessity I overlooked in *Making Morality Work*. Subsequent definitions in that work address the ability of an agent to use a moral principle indirectly when she is uncertain about which of several decision guides is ranked highest.

3. OTHER TYPES OF DECISION GUIDES

Of course, some (but perhaps not all) decision guides may seem obviously appropriate, not only within many *moral* theories, but also within a wide variety of other normative theories that require decision guides to assist agents facing uncertainty. Such theories would include theories about what it is prudent to do, what it is legally appropriate to do, what a code of etiquette requires, what duties of religious observance require, and so forth. For example, the decision guide recommending (roughly) that one do what is most likely to be best is a decision guide that could be appropriate within any one of these alternative normative theories. These considerations may suggest that at least some of the moral decision guides are not specifically *moral* guides, but rather have a more general normative character.

But if, as I have argued, the decision guides used within a moral theory must issue prescriptions that recommend certain conduct as "morally choice worthy," then clearly those decision guides are *not* appropriate for use within one of these other normative spheres. Someone trying to make a prudential decision is not helped by being told what it would be *morally* choice worthy for her to do. Each normative sphere requires its own type of prescriptions: what is morally choice worthy, what is prudentially choice worthy, what is epistemically choice worthy, and so forth. Nonetheless the more formal decision guides from distinct normative domains may share a general form and a good deal of their content. The general form of such a decision guide could be, "Acts of type *X* are [fill in with type of normativity] choice worthy." Specific versions would state, "Acts of type *X* are *morally* choice worthy," or "Acts of type *Y* are *prudentially* choice worthy," and so forth. These decision guides differ in the kinds of normativity they recommend, although the same "type *X*" (such as "most likely to be right") might appear in the decision guides of many domains. Note that these specific versions often differ from each other in what things they view as being valuable when they recommend, for example, that the agent maximize expected value. But they may not. Two decision guides from very different normative spheres might recognize the same things as valuable, but nonetheless issue normatively distinct recommendations. Thus a decision guide for ethical egoism may say that it is *morally* choice mandated to maximize expected personal well-being, while a decision guide for the prudential sphere may say that it is *prudentially* choice mandated to maximize expected personal well-being. These are very different recommendations, despite their common focus on the value of personal well-being.

The various normative spheres may have comprehensive theories with similar structures, requiring decision guides as well as objective principles. And

there may well be abstract templates for the content of certain, more formal decision guides that, with suitable substitutions, will serve in many such normative spheres. We might call these templates "principles of rational decision making," but they must be adapted appropriately within each sphere in order to issue the prescriptions that are suitable for that sphere, and that can enable a decision maker to indirectly apply the relevant objective principle for the sphere in which she is operating.

4. CONCLUSION

Our question has been whether decision guides usable for indirectly applying objective moral theories should be considered as counsels of morality or counsels of rationality. I have argued, *contra* Peter Graham, that they cannot carry out their job unless it is part of their content that they recommend actions as *morally* choice mandated or choice worthy. This clearly places them in the category of moral principles rather than more neutral principles of rationality.[24]

Rutgers University
hsmith@philosophy.rutgers.edu

REFERENCES

Graham, Peter A. "Avoidable Harm." *Philosophy and Phenomenological Research* 101, no. 1 (July 2020): 175–99.

———. "In Defense of Objectivism about Moral Obligation." *Ethics* 121, no. 1 (October 2010): 88–115.

———. "Moral Conscientiousness and the Subjectivism/Objectivism Debate about Moral Wrongness." Presented at the Workshop in Normative Ethics, University of Arizona Center for the Philosophy of Freedom, Tucson, January 16–18, 2020.

Jackson, Frank. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3 (April 1991): 461–82.

Muñoz, Daniel, and Jack Spencer. "Knowledge of Objective 'Oughts': Mono-

24 I am grateful for comments on previous versions of this material by the audiences at the Center for Ethics and Public Affairs at the Murphy Institute of Tulane University, the Roots of Responsibility Workshop at University College, London, and the Moral and Political Philosophy Seminar at Helsinki University, and especially to the referees for this journal.

tonicity and the New Miners Puzzle." *Philosophy and Phenomenological Research* 103, no. 1 (July 2021): 77–91.

Regan, Donald. *Utilitarianism and Co-operation*. Oxford: Clarendon Press, 1980.

Smith, Holly M. *Making Morality Work*. Oxford: Oxford University Press, 2018.

———. "The Zimmerman-Graham Debate on Objectivism versus Prospectivism." *Journal of Moral Philosophy* 15, no. 4 (2018): 401–14.

Zimmerman, Michael. *Living with Uncertainty*. Cambridge: Cambridge University Press, 2008.

Journal of Ethics and Social Philosophy
VOL. 21, NO. 1 · MAY 2022
https://doi.org/10.26556/jesp.v22i1.1181

WHAT IS THE INCOHERENCE OBJECTION TO LEGAL ENTRAPMENT?

Daniel J. Hill, Stephen K. McLeod, and Attila Tanyi

ENTRAPMENT is deservedly the topic of much philosophical attention.[1] The attention is deserved not merely because the topic is in itself of philosophical interest but also because the relationship between entrapment and the guilt, or liability to punishment, of the victim or target of entrapment is important, controversial, and treated differently in different jurisdictions.[2] Entrapment in the sense that we later expound also seems to be widespread, in some jurisdictions, as a method of policing. While it is hard to get figures for entrapment operations *per se*, there were 1,229 undercover police officers operating in England and Wales in 2014, and there were 3,466 authorized undercover operations in England and Wales.[3] It may be speculated that quite a few of these operations involved entrapment in the sense that we later expound.[4] Since the early days of the modern philosophical discussion of entrapment, the incoherence objection—the objection that in some way it is incoherent for an agent of the state to use entrapment, and that this incoherence has negative moral implications for the practice—has often been endorsed.[5] Before we weigh in on this debate, we

1 Two early papers in the modern philosophical discussion of entrapment are Stitt and James, "Entrapment and the Entrapment Defense"; and Dworkin, "The Serpent Beguiled Me and I Did Eat."

2 For example, in the United States, entrapment is a defense to a criminal charge in every state and in federal courts, though not always on the same basis: see Marcus, *The Entrapment Defense*, for details. (Tennessee became the last state to adopt entrapment as a defense, in 1980: see Department of the Army, *The Army Lawyer*, 40n8.) In England and Wales, it is not a defense, but a judge has discretion to exclude evidence gained through entrapment or to stay proceedings: see R v. Looseley [2001] UKHL 53, [2001] 1 WLR 2060.

3 HMIC, *An Inspection of Undercover Policing in England and Wales*, para. 22, 26.

4 In addition, there have been various high-profile cases of journalistic entrapment in many countries, including the United Kingdom. See, for example, O'Neill, "Straw and Rifkind Brought Down by Sting Journalism, but What It Revealed Still Stinks," *Conversation*, February 24, 2015, https://theconversation.com/straw-and-rifkind-brought-down-by-sting-journalism-but-what-it-revealed-still-stinks-37991.

5 This begins with Dworkin, "The Serpent Beguiled Me and I Did Eat."

set out, drawing on earlier work of ours, the basics of our philosophical understanding of the concept of entrapment.[6]

I. LEGAL ENTRAPMENT TO COMMIT A CRIME

Cases of entrapment involve an entrapping party, whom we call the "agent," and an entrapped party, whom we call the "target." Let the terms "party," "agent," and "target" encompass both individuals and groups.

We draw two distinctions, which cut across each other, concerning acts of entrapment. The first concerns the status of the agent; the second concerns the act that the target performs and that the agent procures.

Legal entrapment occurs when the agent is *either* a law enforcement officer acting (lawfully or otherwise) in their official capacity as a law enforcement officer *or* a party acting on behalf of a law enforcement officer as their deputy. When, on the other hand, the agent is neither a law enforcement officer acting in that capacity nor the deputy of such an officer acting in their capacity as deputy, we have *civil entrapment*.[7]

We distinguish between procured acts of criminal and of noncriminal types. An investigative journalist might entrap a politician into performing a morally compromising act that is not a crime in order that the journalist might expose the politician for having performed the act. When the act is noncriminal but is morally compromising (whether by being immoral, embarrassing, or socially frowned on in some way), we are dealing with *moral entrapment* (using the word "moral" in a wide sense). When the act is of a criminal type, we have *criminal entrapment*.

Thus, four types of entrapment can be distinguished: legal criminal entrapment (e.g., the police entrap someone into selling illegal drugs), civil criminal entrapment (e.g., a journalist entraps someone into selling illegal drugs), civil moral entrapment (e.g., a journalist entraps a politician into making an embarrassing boast), and legal moral entrapment (e.g., when law enforcement agents, in their capacities as law enforcement agents, entrap someone into performing a morally compromising act that is not a crime). The incoherence objection to entrapment applies only to legal entrapment to commit a crime. Henceforth, we use "legal entrapment" as an abbreviation for "legal criminal entrapment," and we focus exclusively on this type of entrapment. When legal entrapment occurs, we take it, the following conditions are all met:

6 See Hill, McLeod, and Tanyi, "The Concept of Entrapment."

7 For details of alternative terminologies for the legal/civil distinction, see Hill, McLeod, and Tanyi, "The Concept of Entrapment."

1. A law enforcement agent (or the agent's deputy) acting in an official capacity as (or as a deputy of) a law enforcement agent plans that the target perform an act.[8]
2. The act is of a type that is criminal.
3. The agent procures the act (using solicitation, persuasion, or incitement).
4. The agent intends that the act should, in principle, be traceable to the target either by being detectable (by a party other than the target) or via testimony (including the target's confession)—that is, by evidence that would link the target to the act.
5. In procuring the act, the agent intends to be enabled, or intends that a third party be enabled, to prosecute (or threaten to prosecute) the target for having performed the act.[9]

Condition 2 states that the entrapped act is of a type that is criminal. We are not here concerned with whether the target's token act is one for which the target is *criminally liable*. In our experience, to say that the target's token act is a crime suggests to some readers that the act is one for which the target is criminally liable. We seek to avoid this mistaken impression and to provide a definition of entrapment that prejudges neither the question of the target's criminal liability nor that of the permissibility of entrapment.

8 It has been suggested to us that this condition may not be necessary because it would still count as entrapment if the agent were merely going through the motions and did not plan or intend that the target actually be entrapped. We bite the bullet here: we deny that an agent merely going through the motions actually does entrap the target, even though no doubt the target would feel just as if they had been entrapped.

9 We intend condition 5 to include blackmail cases in which the agent intends not that the target will be prosecuted but that the target will be placed under threat of prosecution. We are aware that this is controversial but do not defend the inclusion here. Many writers hold (which we do not) that entrapment necessarily involves deception. These include Dworkin, "The Serpent Beguiled Me and I Did Eat"; Skolnick, "Deception by Police," 81; Kleinig, *The Ethics of Policing*, 153; Miller and Blackler, *Ethical Issues in Policing*, 104; Miller, Blackler, and Alexandra, *Police Ethics*, 263; and Ho, "State Entrapment," 74. Condition 3, our procurement condition, slightly adapts the wording of Dworkin, "The Serpent Beguiled Me and I Did Eat," 21. In section 2, we provide a philosophical account of procurement itself. Our account there appeals to the agent's influence on the target's will: we intend our account to be independent of any conditions, such as those with which US courts have grappled, regarding whether the target was predisposed to perform the procured act or acts of its type (Sorrells v. United States 287 US 435 [1932], Sherman v. United States 356 US 369 [1958], United States v. Russell 411 US 423 [1973], and Jacobson v. United States 503 US 540 [1992]). For detailed discussion and defense of our conditions, and of our omission of any deception condition, see Hill, McLeod, and Tanyi, "The Concept of Entrapment."

When defining entrapment, some theorists include a counterfactual (or "but for") condition according to which the target has been entrapped only if the target would not have committed the crime but for the agent's actions. For reasons we have explained elsewhere, we do not consider it necessary or desirable to include such a counterfactual condition.[10]

2. PROCUREMENT AND THE CREATION OF CRIME

We turn now to the contention that legal entrapment is objectionable because it *creates* crime.[11] In the context of entrapment, creation is to be understood, we take it, in terms of the creation of *token crimes*. Since a type of act can be illegal even if no one in fact ever happens to commit it, *type crimes* are neither created by nor dependent on token crimes. For example, there is such a type of crime as murder, even if nobody ever in fact commits murder, as long as a legislature outlaws it.[12]

If the incoherence objection to entrapment appeals to the creation of crime (which, we will argue, it does in its most plausible form), then the objection must rest on the contention that it is incoherent for law enforcement agents (or their deputies) in their official capacities to aim to create token crimes. In the literature, the counterfactual account of the creation of crime is popular.[13] According to it, agents create token crimes if the token crimes would not have occurred but for their actions. This does not seem to us to be the right way in which to understand the creation of a token crime. The analysis of creation in terms of the "but for" counterfactual drains the notion of the creation of token crimes of the applicability it is presumably intended to have when it is invoked in an attempt to advance the incoherence objection. That without which an act could not have occurred (even if itself an act) is not to be confused with the thing that happens to have brought it about.[14] (Token crimes would not have occurred but for all manner of things: but for the existence of the criminal, but for the existence of

10 See Hill, McLeod, and Tanyi, "The Concept of Entrapment."

11 Some theorists endorse this objection and add that to entrap is to create, *rather than to detect*, crime. We shall shortly explain their view and why we disagree with that element of it.

12 A natural law theorist would adopt the stronger position that actions like murder are still crimes even if no legislature actually outlaws them.

13 E.g., Dworkin, "The Serpent Beguiled Me and I Did Eat," 21; Stitt and James, "Entrapment and the Entrapment Defense," 114; and Ho, "State Entrapment," 74. Counterfactual accounts of the creation of crime appear (as in the case of Ho) to be localized versions of the more general strategy of attempting to account for causation in counterfactual terms.

14 See further Hill, McLeod, and Tanyi, "The Concept of Entrapment."

the criminal's parents, but for the existence of the victim, but for the meeting of the victim's parents, and so on.)

The notion of creation as we understand it must also be distinguished from that of having acted in a manner that, even if not *necessary* to the target's commission of the token crime, made the target's act *more likely* than would otherwise have been the case.[15] In a decoy operation, the actions of the law enforcement agents make more likely the target's commission of the token act, and thus the situation meets the condition just mentioned. The actions of an agent posing during a decoy operation as a potential victim of a type of crime do not thereby amount to actions that, if a token crime is in fact committed against that agent, mean that the agent *created* the crime. If creation were to be understood so broadly, then the incoherence objection would not be to entrapment *per se*. Instead, it would be a wider objection to all forms of proactive law enforcement that involve the active presentation to the target of an opportunity to commit a crime. We hold that creation goes beyond the mere presentation of an opportunity. On our account, to have created a crime is to have procured it. Entrapment involves the procurement of the actual commission of a crime, rather than mere presentation of the opportunity to commit that crime.[16]

The third condition of entrapment, on our view, is that the agent *procures* the act (using solicitation, persuasion, or incitement). To advance our argument here, we need to explain this condition in more detail. For an agent to procure a target's act is, we stipulate, for the agent to influence the target's will through responsiveness on the target's part to the content of a communicative act (or series of such acts) on the part of the agent. These communicative acts need not be spoken or written: they can, for example, be gestural. What matters is that the communicative acts persuade, solicit, or incite the target.[17]

15 By "more likely," we intend to suggest an act that raises the probability to something less than 1 but greater than 0.5.

16 For a different view, on which both entrapment and creation are conceived of more loosely, see Miller and Blackler, *Ethical Issues in Policing*, 107. On their conception, the mere presentation of an opportunity, such as leaving cash somewhere in the hope that the target will steal it, can count both as an act of creation and as one of entrapment. In our view, the intentional presentation of an opportunity does not count as entrapment if it is not done with the intention that the target actually commit the crime. Andrew Ashworth seems to have a similar view: he writes, "If test purchases are acceptable, they should be excluded from the definition of entrapment" ("What Is Wrong with Entrapment?" 297).

17 The account of procurement in the law of England and Wales is somewhat broader. In Attorney General's Reference (No. 1 of 1975) [1975] EWCA Crim 1, [1975] QB 773, Lord Widgery defined procurement as follows: "To procure means to produce by endeavour. You procure a thing by setting out to see that it happens and taking the appropriate steps to produce that happening" (at 779F). He allows the surreptitious lacing of a drink without the driv-

The considerations in this section, along with our account of procurement, lead us to the following conclusion about how the notion of creation pertinent to the incoherence objection is to be understood. For the agent to *create* a crime is for the agent to procure an act, on the part of the target, that constitutes a token crime. In *procuring* an act of a criminal type, the agent influences the target's will (via the agent's communicative act or acts) in order to bring about that act.

3. INTERPRETING THE INCOHERENCE OBJECTION

The exact nature of the alleged incoherence that legal entrapment to commit a crime is thought, by supporters of the incoherence objection, to involve is, based on the literature so far, difficult to grasp. Moreover, the objection is formulated by those that advocate or mention it in various ways that are apparently not all equivalent to each other. In this and the next section, we demonstrate that existing accounts of the incoherence objection are diverse and that, particularly over the question of the nature of the purported incoherence, they are far too imprecise. We aim to render more precise the various versions of the objection that are in the literature as well as some versions that, while absent from the literature, are interesting theoretical possibilities. To do so, we begin by considering Gerald Dworkin's advocacy of the objection. Probing the objection as it appears in his work enables us eventually to settle on a new and relatively precise specification of the objection. We then argue that from among the various interpretations of the objection we canvass, this specification best maximizes the objection's plausibility.

Dworkin's version of the incoherence objection appeals to the notion of the creation of crime and, more specifically, to that of criminal procurement.[18] His initial statement of the objection appears to be relatively clear:

> The law is set up to forbid people to engage in certain kinds of behavior. In effect it is commanding "Do not do this."…
>
> But for a law enforcement official to encourage, suggest, or invite crime is to, in effect, be saying "Do this." It is certainly unfair to the citizen to be invited to do that which the law forbids him to do. But it is more than unfair; it is conceptually incoherent.[19]

er's knowledge to qualify as procurement of the offense known as "drink driving." On our account, this would not qualify as an example of procurement, unless the agent were encouraging the driver to drink the laced liquid. For more on our view that procurement and causation are distinct, see Hill, McLeod, and Tanyi, "The Concept of Entrapment."

18 Dworkin, "The Serpent Beguiled Me and I Did Eat," 30–34.

19 Dworkin, "The Serpent Beguiled Me and I Did Eat," 32.

This passage gives the impression that the incoherence is a case of *utterance contradiction*. Two utterances are contradictories when one is the negation of the other. Among utterance contradictions, we may distinguish between *statement* (or *assertion*) *contradiction* and (unconventionally, but usefully in the context) *command contradiction*. A contradictory pair of statements (or assertions) cannot be true together and cannot be false together. If two statements (or assertions) are in contradiction, then exactly one of them is true. If a pair of commands, requests, or bans is contradictory, then an agent cannot be in compliance with or out of compliance with both of them at the same time. If two commands are in contradiction, then for any given agent at a given time, the agent is compliant with exactly one of them.[20] While Dworkin appears to depict the incoherence at issue as a form of command contradiction, his suggestion readily lends itself to being construed, as follows, as involving a deontic-logical statement contradiction. On this construction, when the agent entraps, the agent suggests that the entrapped act is permissible. Given that the law debars acts of that type, the law logically implies the impermissibility of the entrapped act. Thus, what the agent suggests about the permissibility of the type of act contradicts what the law implies about that permissibility. This statement-contradiction interpretation, however, suffers from the flaw that the attempted (or successful) procurement of a token act of a criminal type need not (and typically will not) involve any communicative act (or series of such acts) on the agent's part such that its content implies the legal permissibility of the entrapped act. The entrapping agent will typically not be concerned about conveying any message, or impression, to the target that the entrapped act is not illegal. Would a command-contradiction interpretation fare better? It would not, and for a similar reason. The procurement of a token act of a criminal type involves having a certain kind of influence, as explained in section 2 above, on the will of the target. To command the target to commit the act is only one of many ways in which to attempt (or to achieve) this, and we have no reason to believe that most attempts at entrapment use this method. So, to attain a plausible conception of the sort of incoherence involved in Dworkin's version of the incoherence objection, we require a notion weaker than command contradiction.

In any case, insofar as our concern is with understanding wherein, precisely, the supposed incoherence of legal entrapment to commit a crime lies on Dworkin's account, the above quotation sets us off, according to Dworkin's own sub-

20 It does not follow that the agent is *obedient* to exactly one of them. Obedience involves complying *for the right reason*. As Robert Paul Wolff puts it, "Obedience is not a matter of [merely] doing what someone tells you to do. It is a matter of doing what he tells you to do *because he tells you to do it*" (*In Defense of Anarchism*, 9, italics in original).

sequent remarks in the piece, on the wrong path. While the quotation suggests an utterance-contradiction account of the alleged incoherence, Dworkin almost immediately announces that the incoherence objection is not to be construed this way. The piece, however, then characterizes the incoherence that is supposedly involved only in negative terms, leaving us none the wiser as to wherein, exactly, the supposed incoherence lies.[21] A possible escape route from this situation emerges from a little more reflection on the command-contradiction interpretation. Given that the act that the agent intends the target to perform is of a criminal type, it is an act of a type that is legally prohibited. Thus, the law commands that it not be performed. The agent's communicative act (or series of communicative acts) of procurement is intended to encourage the target to perform the act. It expresses an intention, on the agent's part, that the target break the law. It is the agent's *intention* and the law's *requirement* that fail to cohere with each other, for the target cannot simultaneously satisfy them.[22]

We offer this observation as a way of trying to convert Dworkin's incomplete, and wholly negative, characterization of the relevant form of incoherence into something more precise. We believe, and argue over the course of this article, that the best prospects for the incoherence objection lie in the appeal to the notion of *practical* incoherence. In order to cast the objection in its best light, proponents of the incoherence objection ought to allude not to a formal or utterance contradiction or contrariety, but rather to the notion, recognized by Aristotle and within the Aristotelian philosophical tradition, of *contrariety of ends*.[23] Two ends, such as enforcing a party's observance of a law and encouraging that same party to disobey that law, are contraries when the attainment of one of them by an agent necessarily precludes the simultaneous attainment by the agent of the other.

This is, however, still not precise enough. In particular, we still need to get a grip on exactly wherein the aforementioned contrariety of ends consists. What exactly are the entrapping agent's contrary ends? We can get to an answer by noticing, first, that Dworkin's ultimate position does not seem to be that entrap-

21 Dworkin, "The Serpent Beguiled Me and I Did Eat," 32–33. Dworkin's negative characterization of the incoherence consists in the denial that the incoherence involves either a "literal" or a "pragmatic" contradiction.

22 Dworkin remarks that "it is not the purpose of officers of the law to encourage crime," and he holds, further, that it is contrary to their purpose for them to do so ("The Serpent Beguiled Me and I Did Eat," 32). Thus, we take his to be what we call a "functional" version of the incoherence objection.

23 "Utterance contrariety" describes the situation when the two utterances cannot both be true (whether or not they can both be false), while "utterance contradiction" describes the situation when they cannot both be true and also cannot both be false.

ment is always incoherent. Instead, he appears to hold that incoherence enters the picture when (but only when) law enforcement agents attempt to entrap an individual that they do not have good reason to believe is already engaging in acts of the same type as the intended token criminal act.[24]

Dworkin claims that random entrapment involves creating, rather than detecting, crime.[25] On his account, to entrap a target that has not already been committing (or intending to commit) crimes of a given type is to create a token crime that manifests neither prior nor ongoing criminal conduct (or intended conduct) of the same type. Dworkin appears to hold that to entrap into committing an act of a certain type a target who is already engaged in (or already intends to engage in) criminal conduct of that type counts as genuine detection (rather than creation) of crime. It seems, then, that for Dworkin creation of crime occurs when a target is entrapped into committing a crime of a type none of whose tokens the target was already engaged in committing (or intending to commit). Dworkin's ultimate position is that it is the use of entrapment against people not already suspected of committing crimes (or of intending to commit crimes) of the relevant type that is incoherent: for on Dworkin's view, creation is inconsistent with detection, and detection of crime, but not its creation, is a legitimate aspect of law enforcement. In short, the contrary ends for which we have been looking on the part of the entrapping agent are those of detection (of crime) on the one hand, and of creation (of crime) on the other. There is a question, however, whether such cases of the creation of crimes are to be held inconsistent with *detection*.

Recall that, on our account, when an agent procures a crime, the agent has influence of a certain sort on the target's will. To procure a crime is, we stipulated, to bring it about through solicitation, persuasion, or incitement that another commits that crime. Acts of solicitation, persuasion, and incitement are communicative acts: these include, but are not restricted to, speech acts.[26] To have

24 Dworkin, "The Serpent Beguiled Me and I Did Eat," 33. We use hesitant language in making this statement about Dworkin because our interpretation of what he says relies on connecting incoherence to impermissibility. We reason that if Dworkin regarded entrapment as always incoherent, then plausibly he would think it impermissible in all circumstances too. Instead, his position appears to be that entrapment is impermissible only in the cases he calls "virtue testing." Ultimately, though, whether we are right in our interpretation of Dworkin makes no difference to the cogency of our argument.

25 Dworkin, "The Serpent Beguiled Me and I Did Eat," 33. Cf. Sherman v. United States, 356 US 369, 384 (1958), concurring judgment.

26 Flagging down a taxi, for example, is a communicative act that is not a speech act. In this respect, it differs from a gesture of a sign language such as British Sign Language. The gestures of BSL are part of an overall system of communication that possesses both a syntax and

procured a crime that a target has committed is to have inclined, via the content of such a communicative act (or series of such acts), the target's will toward committing that token crime. Now, even when a target is already inclined to commit a crime of a given type, it is nevertheless possible for an agent to entrap that target: a will that is generally disposed to committing crimes of a certain type need not always be inclined, whenever an opportunity to commit such a crime with an apparently low risk of being caught is presented, to take up that opportunity. In fact, even a record of convictions for crimes of a given type is strong evidence only of predisposition to commit crimes of that type: it is not the case that for every relevant token of that type, such a record is strong evidence of a predisposition to commit *that token*. It is therefore unclear whether the incoherence objection can really be restricted, as Dworkin seeks to have it, to cases where the target was innocent of the relevant type of crime prior to the entrapment scenario.

Let us clarify this further by providing a more formal representation of Dworkin's position. Dworkin seems to appeal to the following principles:

1. When legal entrapment occurs, either the target is already reasonably suspected of engagement, or of intending engagement, in crimes of the same type as the token entrapped crime or the target is not so suspected.
2. If the target is not so suspected, then the agent is creating, or attempting to create, the token crime (whether or not it is traced to the target).[27]
3. If the target is so suspected, then the agent is detecting the token crime (on the assumption that it is traced to the target).
4. The agent cannot both detect and create (or even *attempt to create*) one and the same token criminal act.
5. Creation (and attempted creation) and detection are contrary functions: thus, the creation (or attempted creation) of a crime by law enforcement agents is inconsistent with their role of detecting crimes.

The fundamental problem we see here is that the fourth principle is false: it *is* possible to detect and create one and the same token criminal act. When agents entrap, they help to create a token crime. They may also find evidence that links

a semantics. It seems to us that this cannot be said of such gestures as flagging down a taxi, waving, or giving the thumbs up, at least when these gestures are not parts of an overall system of communication in which the symbols involved are type homogeneous (e.g., they are all inscriptions or phonemes or gestures) and in which there are formation rules for strings of them.

27 It is possible that the target, unknown to the agent, is engaging in crimes of the same type as the token entrapped crime. In this case, the agent is not, according to this principle, creating the crime but attempting to create it.

the target to the crime, in which case they detect it too. If creation and detection are contrary functions, then this is not because it is impossible both to create and to detect the same token act of a criminal type. On the contrary, doing this is clearly possible. We conclude, therefore, that the incoherence objection cannot succeed if it appeals to the alleged incompatibility of creation and detection at the level of the target's token act.

To make the incoherence objection work, we need, then, to find some other contrariety in the agent's ends. Let us go back to our original idea: it is the agent's intention that the target should perform an illegal act and the law's injunction against that act that are incompatible, for the satisfaction of one necessarily precludes the simultaneous satisfaction of the other. Dworkin's incoherence objection, when interpreted in the most plausible way, consists, we take it, in the assertion that the function of law enforcement is incompatible with, and therefore subverted by, satisfaction of the entrapping agent's intention. Since, as we understand the concept of entrapment, it is impossible to entrap without having that intention, entrapment itself is functionally incompatible with law enforcement.

The underlying incompatibility, we suggest, is not between creation and detection but between creation and *prevention*. Since it is impossible for an agent both to prevent and to create a given token crime, but possible for that agent to do neither, we are dealing with a form of contrariety but not a form of contradiction. Agents that procure a token act of a criminal type create it and have intended to create it. They have not intended to prevent it. The law expresses the intention that the act should not occur, while the act of entrapment expresses the intention that it should.

Nevertheless, the incoherence objection is too strong if interpreted like this. The objection relies on the premise that law enforcement agents have a duty to prevent the crime that they procure in entrapment. Law enforcement agents, however, do not have a duty to prevent *every* crime that they possibly can: there will certainly be occasions when they must choose between preventing two crimes, with the result that there is a preventable crime that they do not prevent. We therefore need an argument for the premise that law enforcement agents have a duty to prevent crimes *that is breached in cases of entrapment*.

It would beg the question to assume that law enforcement agents are engaging in incoherent conduct, or even are guilty of dereliction of duty, if they intentionally allow a minor crime to occur in order to prevent a major crime. It is not obviously incoherent for law enforcement agents to allow a minor crime for the sake of the *possibility* of preventing a major crime, as when they allow the boss's minion to get away with something small in order that they might find out who the boss is.

How can the proponents of the incoherence objection respond? One way is to accept the above and try to argue that law enforcement agents have a duty to prevent all crimes that they can where preventing the crime will not frustrate the aim of preventing another crime that is equally bad or worse. We think, however, that most proponents of the objection intend it to apply to all cases of entrapment; besides, the objection is more interesting and powerful if its scope is not restricted. What can its proponents say, then?

They could take a Kantian-style position that law enforcement agents simply have a duty never to create crimes and that this duty can never be suspended for any higher purpose. Although this means that there may be more crime in a state than there otherwise would be, blame for this regrettable fact is not to be laid at the feet of law enforcement agents. Rather, it is a potential side effect of any theory that denies that an action can always be justified if the consequences are good enough.[28]

We are not here attempting to answer the moral question of whether entrapment is ever permissible. We are merely seeking to show what must be believed in order for the incoherence objection to work. If the objection is to encompass all cases of legal entrapment, even cases of entrapment into minor offenses, then we believe that it must involve the assertion that law enforcement agents have an absolute duty never to create crimes.

During our discussion of Dworkin, we have come across the following forms of incoherence and weighed each of them up as an interpretation of the alleged incoherence.[29]

Statement contradiction. According to this interpretation, when agents entrap they declare that a type of action that is legally debarred is, in fact, legally permissible. This is not a charitable interpretation of Dworkin's objection because it is untrue that entrapping agents must make, or even suggest, any such declaration.

28 For further illustration of the strictness of the duty, note that its demands clearly spill over to undercover work. Take the case of an undercover officer witnessing or even contributing to crimes, but doing so in order to avoid blowing their cover. The Kantian duty, it seems, would also not allow this behavior, and since undercover work is likely to involve instances of permitting or even helping others to commit crimes, the Kantian duty would (severely) restrict (if not eliminate) undercover work.

29 We do not claim that these interpretations exhaust the possibilities. For example, it might be claimed that the entrapping agents' utterances are contrary to one of the agents' law enforcement functions, or that it is the utterances of the judge or the prosecution, if the case gets to court, that are contrary to those of the agents. We admit that these possibilities, among others that we have not discussed, are in principle available. We have concentrated on what we take to be the more plausible candidate interpretations of the incoherence objection.

Command contradiction. This interpretation has it that when entrapping, agents enjoin the targets to commit acts that are of a type the criminal justice system enjoins people not to commit. While perhaps more plausible than statement contradiction, this objection is also based on an exaggerated generalization. Entrapping agents need not go so far as to *enjoin* the targets to commit the acts. If the targets' acts have been procured by solicitation, persuasion, or incitement on the agents' part, then it does not follow that the agents have specifically enjoined the targets to commit them: even if *incitement* involves enjoining the targets to commit the acts, *solicitation* and *persuasion* can be subtle forms of encouragement that need not involve going so far as enjoining the targets to commit the acts. For example, a communicative act that is intended to "nudge" the target, and succeeds in doing this, can procure the act.

The two forms of contradiction listed so far are both cases of *utterance contradiction*. This provides what is the strongest form of the incoherence objection from a logical point of view, but which is consequently the weakest in terms of philosophical credibility. When two utterances contradict each other, this situation cannot be changed by the addition of further utterances. It could easily be written in statute that while it is a criminal offense for civilians to abet or encourage someone in committing a crime, it would not necessarily be criminal for the police to do so in the context of attempting to bring someone to justice. If the incoherence objection had to be interpreted as involving an allegation of utterance contradiction, then it would be utterly implausible. Moreover, more plausible interpretations are available. Thus, no utterance-contradiction interpretation should be adopted.

Functional contrariety/contrariety of ends. When agents entrap, they pursue an end (the encouragement of targets to commit crimes) that cannot be pursued (by the same agents) at the same time as their end of enforcing the law. The agents create token acts of a criminal type, and this is contrary to the end of preventing such acts. The latter end, in turn, is one that the agents have, whether it is present to their minds or recognized in their actions and intentions, in virtue of their offices as law enforcement agents. It is part of the functional role of law enforcement to prevent, and so not to create, acts of a criminal type. This is the interpretation of Dworkin's version of the incoherence objection that we have suggested is the most plausible. Unlike earlier candidates, this interpretation does not appear to rest on a false empirical generalization about the behavior of entrapping agents. In order for the objection to apply to *all* cases of legal entrapment, however, it has to be supplemented by a Kantian-style thesis that this role can never be suspended in the short term for the sake of a long-term gain in crime prevention.

4. OTHER FORMULATIONS OF THE INCOHERENCE OBJECTION

We have argued that the most plausible understanding of Dworkin's version of the incoherence objection involves the idea that law enforcement agents engaging in entrapment thereby lapse into a form of practical incoherence involving contrariety of ends. The objection rests on, we have suggested, the proposition that law enforcement agents have an absolute duty never to create crimes. In this section, we survey formulations of the incoherence objection in the work of writers other than Dworkin. Our purpose now is to assess whether any of these fare better than the version of the objection that we specified, via our probing of Dworkin's account, in the previous section. We argue that none do. Each such formulation either does not give us a readily workable version of the objection or is best interpreted as a less precise way of stating the objection in the form given in the previous section.

We begin with Andrew Ashworth's formulation. He is another prominent supporter of the incoherence objection, though he uses the word "inconsistent" rather than "incoherent." In one article, he writes:

> It would compromise the integrity of the courts if they were to act on the fruits of manifestly unacceptable practices by law enforcement officers; or, to put it another way, . . . criminal justice would lose its moral authority if courts did not insist that those who enforce the law should also obey the law. It is therefore, at root, a principle of consistency—that it would be inconsistent for the courts, as guardians of human rights and the rule of law, to act on evidence obtained by methods which violate human rights and/or the rule of law.[30]

In an earlier article, he argues as follows:

> [When entrapment occurs] the entrapping officer has breached the internal rules of the police or other law enforcement agency, and may well have committed a crime. Entrapment will usually involve the inchoate offence of incitement, and may make the entrapper an accomplice to the substantive offence as a counsellor or even a procurer. The English Law Commission went so far as to suggest that there should be a specific

30 Ashworth, "Re-drawing the Boundaries of Entrapment," 163. For more on Ashworth's starting point in this quotation, namely the "integrity principle," see Ashworth, "Exploring the Integrity Principle in Evidence and Procedure"; and Hunter et al., *The Integrity of Criminal Process*. (So as to concentrate on what we take to be common to different formulations of the incoherence objection, we limit our engagement with the integrity principle to some passing remarks in footnotes.)

> crime of entrapment, which an officer would commit if he incited the commission of an offence and even if he intended that the completion of that offence would be prevented or nullified.[31]

There are several suggestions in play here. We rephrase two of the most salient in our own language and by reference to the account of legal entrapment given in section 1:

> *Rule Breach*: A law enforcement officer that engages in entrapment breaches the internal rules of the officer's law enforcement agency.

> *Criminality through Complicity*: To procure a crime involves being complicit as an accomplice to the crime; entrapment involves procurement; so, entrapment involves criminal complicity.

Each suggestion can be construed as providing a reason why legal entrapment might be considered, at least under certain circumstances, incoherent.

If Rule Breach is intended as an empirical generalization, then it is easily seen to be false. There are law enforcement officers in certain jurisdictions, such as China, in which neither the law enforcement agency itself nor the law proscribes entrapment as being against the rules or a form of misconduct.[32] Moreover, this goes not just for formal rules but also for informal rules that are matters of "custom and practice" or "ethos" without being formally codified or documented.[33]

Rule Breach appears more plausible when interpreted, rather than as an empirical generalization, as making the same essential point as Dworkin's "functional" version of the incoherence objection. On this understanding, it is a rule internal to the practice of law enforcement that law enforcement does not involve entrapment.[34] This is for a subsidiary reason that underlies the above statement of Rule Breach. To entrap is to procure a token crime, the procurement of which is incompatible with law enforcement's function of preventing, not creating, (token) criminal acts. Since Rule Breach is intended as a general injunction that

31 Ashworth, "What Is Wrong with Entrapment?" 310–11. Although Ashworth does not explicitly appeal to the notion of incoherence in this quotation, it seems to us that he is in the same general territory. See also Ashworth, "Re-drawing the Boundaries of Entrapment": notes 36–38 focus in particular on the contention that legal entrapment involves criminality on the part of the agent. For more on this, see also Williams, *Criminal Law*, 781–82.

32 See Zhou, "Research on Entrapment in China."

33 Cf., for the distinction, D'Agostino, "The Ethos of Games."

34 One might be tempted to construe this rule as a "practice rule" in John Rawls's sense (in his "Two Concepts of Rules"). As we note below, however, it is perfectly possible to conceive of a law enforcement agency with the sole function of investigating crime. Hence, this rule cannot be taken to constitute the practice of law enforcement as Rawls would have it.

debars all acts of legal entrapment on the grounds of their alleged incoherence, it must appeal to a factor that is common to all cases of legal entrapment. We have already argued that it is not the breach of rules, whether formal or informal, that is this common factor. In identifying procurement as the common factor, we are able to advocate, to some extent, on Ashworth's behalf.

There is another drawback, however, with Rule Breach as Ashworth states it. It is too narrow to construe breach of the rules, as he does, as happening when an entrapping law enforcement agent's conduct is inconsistent with the rules of the law enforcement agency to which the agent belongs, or, as we prefer, to construe it as inconsistent with a principle internal to the practice of law enforcement. To see this, note that a law enforcement agency could be established whose sole function was to investigate crime, and perhaps also prosecute the perpetrators, with law enforcement's other functions being carried out by other agencies.[35] There seem to be no rules internal to the practice of investigating crime that debar the creation of token crimes. This drawback can be remedied by widening the sort of rules involved. A very wide way of doing this would be to include all those rules that are internal to those functions had by law enforcement in general, rather than by any particular branch of it or agency responsible for it.

As a result of the above discussion, a full argument can now be reconstructed based on considerations inspired by the above quotation from Ashworth:

1. It is a rule internal to the practice of law enforcement (as a whole) that law enforcement agents do not create token crimes. (Premise)
2. Whenever law enforcement agents entrap those not intending to commit the crime in question, they create token crimes. (Premise)
3. Whenever law enforcement agents entrap those not intending to commit the crime in question, they breach a rule that is internal to the practice of law enforcement. (From 1, 2)
4. To breach a rule that is internal to a practice in which one is involved is to engage in conduct that is incoherent. (Premise)
5. Whenever law enforcement agents entrap those not intending to commit the crime in question, they engage in conduct that is incoherent. (From 3, 4)

The main flaw in this argument seems to be premise 4. Let us give an example different from entrapment. Can law enforcement agents exceed speed limits and go through red lights at junctions when in pursuit of a dangerous criminal? While some jurisdictions actually write exceptions for emergency services in statute, it

35 Perhaps the Serious Fraud Office in the United Kingdom is an example of such a law enforcement agency.

seems to us that in the absence of such exceptions, it is not necessarily incoherent for a law enforcement agent to commit the minor offense of breaking traffic laws in order to prevent a major crime from taking place.[36] Although this is quite a different case from entrapment, premise 4 is stated as applying quite generally, and so can be maintained only if Ashworth adopts a strong Kantian stance to the effect that the duty of police officers to uphold the law is always and everywhere inviolable.

Let us now turn to the other main suggestion that can be developed from Ashworth's comments. Criminality through Complicity can also be extended into a more substantial argument, as follows:

1. Upholding the law is a general end/function of law enforcement. (Premise)
2. Every act of entrapment is an act that procures a token crime. (Premise)
3. For every token criminal act that one procures, one is an accomplice to that token criminal act. (Premise)
4. To be an accomplice to a token criminal act is to act criminally. (Premise)
5. To act criminally is to fail to uphold the law. (Premise)
6. Whenever a law enforcement agent entraps, the agent is an accomplice to a token criminal act. (From 2, 3)
7. Whenever a law enforcement agent entraps, the agent acts criminally. (From 4, 6)
8. Whenever a law enforcement agent entraps, the agent fails to uphold the law. (From 5, 7)
9. Whenever a law enforcement agent entraps, the agent's conduct is contrary to a general end/function of law enforcement. (From 1, 8)

Again, the defender of entrapment is likely to respond that it is permissible to act contrary to the general end in one way if one ends up serving it (or is likely to serve it) in another way: in consequence, creation of a small crime may be justified in pursuit of prevention of a big crime or more than one crime. Once more, then, it seems that Ashworth must, if his version of the incoherence objection is to hold across all cases, adopt a strong Kantian-style stance that the end of

36 In England and Wales, under the Road Traffic Regulation Act 1984, sec. 87, speed limits do not apply to police vehicles being used for police purposes, and under the Traffic Signs Regulations and General Directions 2002, sec. 36, red lights do not apply to emergency services in the same manner in which they apply to the general public.

upholding the law can never legitimately be breached in the short term in order to be achieved more thoroughly in the long term.

When Rule Breach and Criminality through Complicity are spelled out in their more developed versions above, the differences between them emerge as minimal. Crucially, they both rely on the contention that entrapment is incoherent because it is contrary to a general function/end of law enforcement. The main difference between the two arguments is that Criminality through Complicity goes further than Rule Breach in that it also alleges a form of criminality. This additional aspect of Criminality through Complicity and the correctness of the corresponding grounding premises in the argument are, however, irrelevant to our current dialectical purposes.

The upshot of our discussion of Ashworth's version of the incoherence objection is that it, like Dworkin's objection, is most plausible when interpreted as resting on the appeal to a form of practical incoherence stemming from contrariety of ends. Our interpretations of what these two theorists have to say about the incoherence of entrapment are thus in a relationship of mutual support.

Another writer on entrapment, Jeffrey Howard, states the incoherence objection (without endorsing it) as follows: "Entrapment is incoherent; the state acts inconsistently when it insists that citizens adhere to the law, but then takes measures to induce them to break it."[37] On this understanding, the alleged incoherence appears to be between, on the one hand, *pronouncements* or *utterances* of the state that citizens must adhere to the law and, on the other hand, *actions* on the part of some of its agents that are designed to encourage some citizens in some circumstances to break the law.

Let us survey three ways to interpret Howard's statement of the objection. First, it is familiar that the pronouncements of individual agents may be at odds with their own behavior. A television evangelist, for example, might condemn adultery in public but commit it in private. This sort of incoherence is hypocrisy. Suppose that the state can act, in virtue of its agents. The analogy with the television evangelist is straightforward only if whenever the state induces someone to break the law, the state thereby *does* what it itself condemns. In discussing Ashworth's Criminality through Complicity, we saw that this might be the case, but it is not necessarily so: the special responsibilities of law enforcement agents come with a certain amount of special license that they have in virtue of their offices as law enforcement agents.

Howard's exact formulation of the charge speaks not of the state's breaking the law but of the state's *encouraging* its citizens to break the law. This sug-

37 Howard, "Moral Subversion and Structural Entrapment," 26.

gests—leading to our second interpretation of Howard's formulation—that the incoherence, if any, of legal entrapment is not like that of the television evangelist mentioned earlier. Rather, it is more like that of television evangelists that preach against adultery but, without committing it, intentionally tempt others to do so, with the aim that they will succumb to the temptation. Now, intentionally tempting someone in this manner to do something that one declares to be wrong (in our case, criminal) might be criminal, as well as morally wrong. Whether it is *incoherent*, though, is less obvious.[38]

Both of the two interpretations of Howard's formulation of the incoherence objection that we have discussed so far assume that when legal entrapment occurs, it is the *state* (in virtue of its agents) that is acting. How about giving up this assumption? Doing so leads us to our third interpretation of Howard's formulation. In this case, rather than understanding the incoherence objection in terms of the state's doing something that is inconsistent with its pronouncements, which would involve entanglement in the issue of whether the state is itself an agent, it is perhaps better to view it in the following terms. When law enforcement agents entrap someone, one group of state agents—namely, the law enforcement officers (who are part of the executive)—encourages the target to do something that another group of state agents, constitutive of the legislature, has deemed (in statute) to be legally impermissible or that a third group of state agents, consisting of the judiciary, has deemed (e.g., on the basis of case law) to be impermissible under the law. Read in this way, the incoherence involved in legal entrapment would not be one of hypocrisy focusing on one agent only but would appear at the level of the system of criminal justice. It still seems, however, that a strong Kantian-style absolute prohibition of encouragement to break the law would be necessary to sustain this argument—there is nothing obviously incoherent about encouraging someone to break a minor law in order to prevent the breach of a major law (or of several laws).

Yet another formulation of the incoherence objection comes from Jonathan C. Carlson. He asserts that "for the government itself to encourage acts that could actually cause injury to the interests it wished to protect would be the height of absurdity": the idea here is that if, for example, someone were selling illegal drugs, then this would cause "injury to the interests that the law seeks to protect."[39] In consequence, it would be incoherent for the government to encour-

38 For a discussion of the relationship between entrapment and temptation, and its ethical implications, see Hill, McLeod, and Tanyi, "Entrapment, Temptation and Virtue Testing."

39 Carlson, "The Act Requirement and the Foundations of the Entrapment Defense," 1061. Although Carlson uses the word "absurdity" here, rather than "incoherence," it does not seem to us that there is any relevant difference in meaning between the two.

age this injury by having its agent request illegal drugs from the target. Carlson makes this assertion, however, only to point out that it does not apply to most cases of entrapment: in many cases (e.g., when the agent pretends to be an assassin and encourages the target to place an order for someone to be eliminated), no injurious act in fact takes place, and the target is arrested for the offense of attempting to procure an injurious act; and in other cases (e.g., when the agent purchases illegal drugs from the target), the harm that would tend to result from token crimes of the same type is neutralized (because the drugs are destroyed, rather than consumed, by the agent).

Nevertheless, there are some cases of entrapment in which the critique mentioned by Carlson does apply. For example, if an undercover agent encourages some bank robbers to rob a particular bank in which the police will lie waiting, the agent may well know that the robbers will cause some harm (physical damage and shock to innocent bystanders) before they are apprehended. (This critique would extend to cases of proactive policing as well, in which police officers might watch an area notorious for assaults in the hope of catching an assailant in the act, while knowing that they will not be able to stop the assailant before harm has been caused to the victim.) Although there is a *prima facie* case here for incoherence ("absurdity," to use Carlson's word), once again it seems to us that the argument requires a strong Kantian premise to the effect that it is never permissible to encourage a small injury to the interests one wishes to protect in order to prevent a bigger injury to them. Absent such a premise, it seems to us that the existence of incoherence or absurdity is not made out.

5. A MORE EXACT FORMULATION OF THE INCOHERENCE OBJECTION

We have argued that the incoherence objection is best formulated using the distinction between the prevention and the creation of crime. According to the objection, legal entrapment gives rise to a contrariety of ends and thus to a *prima facie* form of practical incoherence. We can now provide a more exact formulation of the incoherence objection as follows:

1. The prevention of crimes is a general function of law enforcement. (Premise)
2. If the prevention of crimes is a general function of law enforcement, then law enforcement agents, on pain of incoherence, must not intentionally bring about or intentionally help to bring about token crimes. (Premise)

3. When an agent entraps a target, that agent intentionally procures a token crime. (Premise)
4. If an agent intentionally procures a token crime, then the agent intentionally brings about or intentionally helps to bring about that token crime. (Premise)
5. When an agent entraps a target, the agent intentionally brings about or intentionally helps to bring about a token crime. (From 3, 4)
6. Given the general functions of law enforcement, on pain of incoherence, law enforcement agents must not intentionally bring about or intentionally help to bring about token crimes. (From 1, 2)
7. Given the general functions of law enforcement, on pain of incoherence, law enforcement agents must not entrap anyone. (From 5, 6)

On the assumptions both that the definition of entrapment on which this argument draws is correct (and hence that premises 3 and 4 are defensible) and that it is correct that the prevention of crime is a general function of law enforcement (premise 1), the controversy is likely to center on premise 2. Note that the consequent of premise 2 is normative, for it states what law enforcement agents *must not* do. The key questions now concern how the normativity is to be construed and whether it is absolute.

Is there something normative to say about the contrariety of the goal of crime prevention and the creation of crime by entrapment? It is tempting to hold that the normativity in question arises simply from the contrariety involved: it is just wrong to be incoherent. What kind of wrongness is this? Why is it wrong to have contrary ends? It would make the incoherence objection stronger if we answered these questions by telling a story about why incoherence of this kind is problematic. What is wrong, then, with practical incoherence, understood as a contrariety of ends, in the case of legal entrapment?[40]

We think the following story might be told on behalf of proponents of the incoherence objection. The wrongness of practical incoherence, it might be suggested, consists in the fact that it involves breaching a requirement of practical reason. Now "requirement of practical reason" can be interpreted in two different

40 There is an interesting parallel here with one way in which the incoherence objection might be related to the integrity principle. (We say this while remaining neutral about whether any supporters of the integrity principle would actually see things this way.) It is one thing to hold that entrapment introduces incoherence into the legal system and that this is wrong. It is another to say *why*. One such reason may be because legal entrapment would damage the integrity of the criminal justice system. It is "extra ammunition" of this kind for which we are looking here and an explanation, moreover, that is broader than the appeal to integrity (which would only cover some instances of entrapment and their incoherence).

ways: first as requiring *structural rationality* (i.e., structural requirements on our attitudes), and second as requiring *reasons for action*.[41] To take the first of these, is there *structural* practical irrationality involved in entrapment? There is an argument for that conclusion. Consider the following remarks from Thomas E. Hill Jr. on different forms of irrational practical incoherence:

> If certain means are necessary to an end, one must choose the means or else give up the end; to hold on to an end while refusing to take the necessary steps to achieve it is a form of practical incoherence. . . . Similarly, it is generally a mark of incoherent (though possible) practical thinking to pursue goals that undermine one's other goals or to employ means that violate the values that were the basis for choosing one's goals.[42]

The first requirement is given by what is called the instrumental principle. This can easily be met, however, by cases of entrapment: there is no reason to suppose that entrapment is not a means, and not chosen as a means, to an end, such as long-term crime prevention.

The other two phenomena that Hill enlists do seem, at first sight, better candidates for the proponent of the incoherence objection. If instead of writing, we choose to go walking, we should be able to return to our writing and take it up where we left off; we shall still be writing, no damage having been done (except, perhaps, to our schedule if there is a deadline). When law enforcement officers entrap, however, they seem to go against one of the very values (crime prevention), and an associated rule (not to create crimes), central to their roles as law enforcement officers. In this way, the kind of practical incoherence involved in entrapment threatens to turn into something more damaging: practical irrationality.[43]

Still, this apparent threat is not real, for the charge that this amounts to practical irrationality seems to fall to the response that there is no irrationality in sacrificing short-term crime prevention for greater crime prevention in the long term. The only way to get around this response would be to appeal to the strong, Kantian-style duty that law enforcement agents must never create crime. Once this Kantian-style prohibition is in place, the agents' goal of crime prevention is in effect restricted to the short term, barring them from the pursuit of long-term crime prevention through entrapment. In fact, this appears to give us a natural way to adapt Hill's point to entrapment: the entrapping agents' actions, so a

41 See Wallace, "Practical Reason," esp. sec. 4; and Kolodny and Brunero, "Instrumental Rationality," esp. sec. 1.

42 Hill, "Reasonable Self-Interest," 68n27.

43 Notice the interesting parallel here with the appeal to the integrity principle. The irrationality we describe can also be seen as a loss of *intra*personal integrity.

supporter of the incoherence objection can argue, go against the foundational values of their service by violating the Kantian duty in question. However, while this move can be made, it would also mean that the normativity we find when interpreting the "must not" in premise 2 is, at base, not (broadly) practical but (narrowly) moral. This is not what we set out to look for when we began our investigation in this section.

Finally, then, let us consider an approach based on the idea that practical reason consists in requiring reasons for one's conduct. Consider the following remarks from Thomas Scanlon:

> Being a good teacher, or a good member of a search committee, or even a good guide to a person who has asked you for directions, all involve bracketing the reason-giving force of some of your own interests which might otherwise be quite relevant and legitimate reasons for acting in one way rather than another. So the reasons we have for living up to the standards associated with such roles are reasons for reordering the reason-giving force of other considerations: reasons for bracketing some of our own concerns and giving the interests of certain people or institutions a special place.[44]

Scanlon's ideas could be applied to form an incoherence objection as follows. Good law enforcement officers are like good committee members—in virtue of their role, they have reason to do what prevents crime from happening, and in virtue of the same role, any considerations that might otherwise have counted in favor of creating crime do not so count (they are bracketed).[45] We could then say that entrapment involves a significant practical failing on the agent's part since the agent is no longer responding properly to the balance of reasons in the agent's case.[46]

Still, it seems to us that this Scanlon-inspired theory meets the same fate as the previous attempt to use Hill's account of practical irrationality. Namely, it fails in the face of the response that the considerations in favor of creating

44 Scanlon, *What We Owe to Each Other*, 53.

45 Considerations that are bracketed do not constitute reasons for a particular course of action since, in virtue of their being bracketed, they do not count in favor of adopting that course of action.

46 We speak of "practical failing" because entrapment might not be construed, if we adopt Scanlon's theory, as involving a form of irrationality since the theory then requires understanding rationality as responsiveness to the balance of reasons and no longer as a structural requirement on the agent's attitudes. Not everyone, however, would accept construing rationality in this way. For a critical treatment of the debate, see González de Prado Salas, "Rationality, Appearances, and Apparent Facts."

crime do not cease to count for the agent just because the agent occupies a law enforcement role (since, again, creation of crime is, in the envisaged situation, instrumental to preventing crime, and this is in line with the agent's role). They may count for less for the agent than they do for someone who does not occupy a law enforcement role, but they do still count for something, and, it seems to us, they could count strongly enough to outweigh the considerations against creating crime. The only counter to this response, we think, is the Kantian-style reply that the considerations against creating crime are insuperable. That is, the point to be made by advocates of the objection would have to be that the Kantian duty is integral to the function of law enforcement and thus either brackets (i.e., renders inapplicable) or trumps law enforcement agents' reason to do what, in their view, will best promote the long-term goal of reducing crime. This would mean, however, that the normativity of the "must not" in premise 2 derives, at base, from a moral duty that proponents of the incoherence objection must assume law enforcement officers have and not from some broader form of practical failure. Again, this was not what we set out to find here.

There is, however, one last option to consider. One could point out that throughout this brief discussion we have assumed that the entrapping agent's intention is to serve the *long-term* goal of crime prevention (albeit by violating it in the short term). But what if this is not the case? We agree that if the agent's aim is not long-term crime prevention but something else, both Hill's second (and third) form of practical irrationality and the practical failure described by Scanlon might be used to criticize the agent without making use of the Kantian-style duty. (This, we presume, is easy to see since our above discussion relied on the move that the entrapping agent creates crime only in order to prevent more crime in the long term.)

However, whether this last suggestion works depends on whether there are cases of entrapment where the agent's end is not long-term crime prevention, and entrapment still appears to be a phenomenon to be reckoned with. For example, the agent might entrap just to get promoted, or just so as not to get demoted, or just to achieve a set number of arrests in a set period—but we do not think that anyone would want to defend entrapment of this kind. How about cases where the agent's aim is that of preventing civil unrest or of safeguarding life and limb (where these are shown not to be construed as instances of long-term crime prevention)? Here the question becomes whether one can even invoke the incoherence objection in the first place (since the end in question might not be contrary to the end of crime prevention). Thus, these are interesting cases to consider, but they are also, arguably, rather marginal. If they were to

turn out to be the only cases to invoke in this context, there would remain not much (if any) ground on which the incoherence objection could operate.

6. CONCLUSION

We have considered in depth various formulations of the incoherence objection and have reconstructed them in detail and with considerably more rigor than we have come across in the literature so far. We have found an interesting commonality between the versions of the objection proposed by Dworkin and by Ashworth, namely that they both (in their most plausible form) depend on the contention that entrapment serves an end contrary to that of law enforcement. We have, however, also pointed out that obtaining the conclusion that entrapment is always incoherent would require a strong Kantian-style premise to the effect that the end of preventing crime can never be suspended in the short term for the sake of greater realization in the long term. We have also tried to embed the incoherence objection into the broader context of practical normativity. Here, too, however, we have found that in almost every relevant case of entrapment, the invoking of the Kantian-style duty ultimately takes center stage. The need to add this Kantian-style premise means that the incoherence objection cannot stand on its own, unaided by further arguments and assumptions, as an objection to all cases of legal entrapment to commit a crime.[47]

University of Liverpool
djhill@liv.ac.uk
skmcleod@liv.ac.uk

UiT: The Arctic University of Norway
attila.tanyi@uit.no

47 We formed the aspiration to write about the incoherence objection at a conference entitled "Public Standards, Ethics and Entrapment" (Liverpool, May 19, 2016), supported by the University of Liverpool's Interdisciplinary Networking Fund. We are grateful to our fellow conference participants, especially those from the University of Liverpool's School of Law and Social Justice, for having provided us with intellectual stimulation. For their comments on earlier versions and presentations, we thank Richard Gaskin, Laura Gow, Matthew Hart, Miroslav Imbrišević, Christopher Nathan, Fredrik Nyseth, Thomas Schramme, Findlay Stark, two anonymous referees, and the members of audiences in Copenhagen, Liverpool, and Tromsø.

REFERENCES

Ashworth, Andrew. "Exploring the Integrity Principle in Evidence and Procedure." In *Essays for Colin Tapper*, edited by Peter Mirfield and Roger Smith, 107–25. London: LexisNexis UK, 2003.

———. "Re-drawing the Boundaries of Entrapment." *Criminal Law Review* (March 2002): 161–79.

———. "What Is Wrong with Entrapment?" *Singapore Journal of Legal Studies* (December 1999): 293–317.

Carlson, Jonathan C. "The Act Requirement and the Foundations of the Entrapment Defense." *Virginia Law Review* 73, no. 6 (September 1987): 1011–1108.

D'Agostino, Fred. "The Ethos of Games." *Journal of the Philosophy of Sport* 8, no. 1 (October 1981): 7–18.

Department of the Army. *The Army Lawyer*. Pamphlet 27–50–193. Judge Advocate General's School, January 1989. https://www.loc.gov/rr/frd/Military_Law/pdf/01-1989.pdf.

Dworkin, Gerald. "The Serpent Beguiled Me and I Did Eat: Entrapment and the Creation of Crime." *Law and Philosophy* 4, no. 1 (April 1985): 17–39.

González de Prado Salas, Javier. "Rationality, Appearances, and Apparent Facts." *Journal of Ethics and Social Philosophy* 14, no. 2 (December 2018): 83–111.

Hill, Daniel J., Stephen K. McLeod, and Attila Tanyi. "The Concept of Entrapment." *Criminal Law and Philosophy* 12, no. 4 (December 2018): 539–54.

———. "Entrapment, Temptation and Virtue Testing." *Philosophical Studies* (forthcoming). Published ahead of print, January 6, 2022. https://doi.org/10.1007/s11098-021-01772-4.

Hill, Thomas E., Jr. "Reasonable Self-Interest." *Social Philosophy and Policy* 14, no. 1 (Winter 1997): 52–85.

HMIC. *An Inspection of Undercover Policing in England and Wales*. HMIC, 2014. http://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/an-inspection-of-undercover-policing-in-england-and-wales.pdf.

Ho, Hock Lai. "State Entrapment." *Legal Studies* 31, no. 1 (March 2011): 71–95.

Howard, Jeffrey W. "Moral Subversion and Structural Entrapment." *Journal of Political Philosophy* 24, no. 1 (March 2016): 24–46.

Hunter, Jill B., Paul Roberts, Simon N. M. Young, and David Nixon, eds. *The Integrity of Criminal Process: From Theory into Practice*. Oxford: Hart, 2016.

Kleinig, John. *The Ethics of Policing*. Cambridge: Cambridge University Press, 1996.

Kolodny, Niko, and John Brunero. "Instrumental Rationality." *Stanford Ency-*

clopedia of Philosophy (Winter 2016). https://plato.stanford.edu/archives/win2016/entries/rationality-instrumental/.

Marcus, Paul. *The Entrapment Defense*. New Providence, NJ: LexisNexis, 2016.

Miller, Seumas, and John Blackler. *Ethical Issues in Policing*. Aldershot: Ashgate, 2005.

Miller, Seumas, John Blackler, and Andrew Alexandra. *Police Ethics*. 2nd ed. Winchester: Waterside Press, 2006.

Rawls, John. "Two Concepts of Rules." *Philosophical Review* 64, no. 1 (January 1955): 3–32.

Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

Skolnick, Jerome H. "Deception by Police." In *Moral Issues in Police Work*, edited by Frederick A. Elliston and Michael Feldberg, 75–98. Savage, MD: Rowman & Littlefield, 1985.

Stitt, B. Grant, and Gene G. James. "Entrapment and the Entrapment Defense: Dilemmas for a Democratic Society." *Law and Philosophy* 3, no. 1 (April 1984): 111–31.

Wallace, R. Jay. "Practical Reason." *Stanford Encyclopedia of Philosophy* (Spring 2018). https://plato.stanford.edu/archives/spr2018/entries/practical-reason/.

Williams, Glanville. *Criminal Law: The General Part*. 2nd ed. London: Stevens, 1961.

Wolff, Robert Paul. *In Defense of Anarchism*. 2nd ed. Berkeley: University of California Press, 1998.

Zhou, Sijia. "Research on Entrapment in China—With Reference to the Experience in Canada." LLM thesis, McGill University, 2013. https://escholarship.mcgill.ca/concern/theses/8910jx77z.

Journal of Ethics and Social Philosophy
VOL. 22, NO. 1 · MAY 2022
https://doi.org/10.26556/jesp.v22i1.1502

THE EQUIVALENCE OF EGALITARIANISM AND PRIORITARIANISM

Karin Enflo

EVER SINCE PARFIT distinguished prioritarianism from egalitarianism, there has been a debate concerning the significance of the distinction.[1] While everyone agrees that egalitarianism and prioritarianism are different theories of social welfare, it is controversial what the distinction implies. Will the theories evaluate and rank populations differently? Or do their differences disappear when they are used for evaluations?

Both Temkin and Broome argue that egalitarianism and prioritarianism will evaluate populations differently, whereas Fleurbaey disagrees and is supported (in part) by Tungodden, McCarthy, and Jensen.[2] In this essay I will side with Fleurbaey and argue that, although egalitarianism and prioritarianism are different theories of social welfare, they can always evaluate populations in the same way. They can, in other words, use the same *social welfare measures*.

This proposal runs counter to a common practice of representing egalitarianism and prioritarianism by different social welfare measures. Egalitarianism is often represented by a derived measure that includes a measure of equality, whereas prioritarianism is usually represented by an additively separable concave function on individual welfare values. These choices of measures are meant to reflect the egalitarian view that equality affects social welfare directly, and the prioritarian view that welfare changes for worse-faring people affect social welfare more. I will argue that this practice is unwarranted. More specifically, I will present six different arguments for the thesis that there is no (or little) reason to distinguish between egalitarian and prioritarian measures.

1 Parfit had distinguished between the two views at least by 1989, as noted by Temkin, "Equality, Priority, or What?" 8.

2 See Temkin, "Equality, Priority, or What?" sec. 9.1, and *Inequality*, sec. 1.E; Broome, "Equality versus Priority," secs. 1–3; Fleurbaey, "Equality versus Priority," secs. 1–4; Tungodden, "The Value of Equality," sec. 5; Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" sec. 6; and McCarthy, "Risk-Free Approaches to the Priority View," 441.

The first argument is based on conceptual connections between inequality and worse faring. I argue that a measure that is sensitive to inequality is necessarily more sensitive to welfare changes for the worse-faring people, and vice versa. Thus, any measure that works for egalitarianism will work for prioritarianism, and any measure that works for prioritarianism will work for egalitarianism as well.

The second argument is based on the equivalence of two minimal conditions that egalitarian or prioritarian measures must satisfy. I argue that satisfying a certain egalitarian condition is both necessary and sufficient for a social welfare measure to qualify as egalitarian. The condition states that if everything is equal between two populations, except for the welfare of one pair of persons, the population with the more equal-faring pair does better. I also argue that satisfying a certain prioritarian condition is both necessary and sufficient for a social welfare measure to qualify as prioritarian. This condition states that, given the choice between increasing the welfare of either of two persons by the same amount, it is better to increase the welfare of the worse-faring person. However, the two conditions are equivalent. Since the two conditions are equivalent, and both are necessary and sufficient to identify their respective measures, there cannot be an egalitarian measure that is not also a prioritarian measure, and vice versa.

The third argument is based on the potential double uses for a standard egalitarian and a standard prioritarian measure. The standard egalitarian measure is a derived measure that multiplies a measure of equality with a measure of total individual welfare, whereas the standard prioritarian measure is an additively separable concave function on individual welfare values. I argue that both measures can be used for either theory.

The fourth to sixth arguments are based on the ability of both egalitarian and prioritarian measures to incorporate properties that have been proposed as fitting for only one of the two theories. The properties in question are: pareto satisfiability, level sensitivity, and relationality (implying non-separability). The standard egalitarian measure is non-pareto satisfying, level insensitive, relational, and non-separable, while the standard prioritarian measure is pareto satisfying, level sensitive, non-relational, and separable. I argue that there is no reason to insist that egalitarianism should use a non-pareto-satisfying, level-insensitive measure, while prioritarianism should use a pareto-satisfying, level-sensitive measure. There is also no reason for prioritarianism to avoid a relational and non-separable measure, although there may be a reason for egalitarianism to avoid a non-relational and separable measure. This is however only the case if a measure must reflect intrinsic dependence relations between social welfare and equality in its very form, which is doubtful.

The essay is structured as follows: in section 1, I distinguish between egalitar-

ianism and prioritarianism as (partial) theories of social welfare; in section 2, I present some assumptions regarding the measurability of individual and social welfare; in section 3, I present the argument from conceptual connections; in section 4, I present the argument from minimal conditions; in section 5, I present the argument from standard measures; in section 6, I present arguments from non-distinguishing properties, and in section 7, I make some concluding remarks.

1. EGALITARIANISM AND PRIORITARIANISM

A social welfare theory can be either axiological or normative: as an axiological theory it concerns the value of populations; as a normative theory it concerns what we should do with respect to populations. While an axiological theory mainly has to consider the intrinsic properties of populations that make them good, a normative theory also has to consider the extrinsic properties of populations that are relevant for decisions, such as the probability that a possible population is realized given a certain set of acts. Here, I will consider egalitarianism and prioritarianism only as axiological theories and discuss the value of populations only relative to their intrinsic properties. However, one could easily transform the axiological theories into normative theories—for example by adding that we should maximize expected social welfare.

Regarded as axiological theories, egalitarianism and prioritarianism have two functions: one explanatory and one evaluative. The first function is to explain what intrinsically affects the social welfare of a population (and how); the second function is to assess populations in terms of their degrees of social welfare. The second function is fulfilled by a *social welfare measure*.

All social welfare theories claim that social welfare is a function of individual welfare. The goodness or badness of populations depends, in some way, on how their individual members fare. Thus all theories include the following claim:

> *Dependence*: The individual welfare levels of the members of a population intrinsically affect the degree of social welfare of the population.

The idea that social welfare would depend *only* on aggregated individual welfare seems intuitively wrong, however. Individuals are separate and the low welfare of some individuals cannot be wholly compensated by the high welfare of others. Thus, distribution of welfare matters too. But how? Egalitarianism and prioritarianism give two different answers to this question.[3] The core of these answers can be presented as follows:

3 For egalitarian ideas see Sidgwick, *The Methods of Ethics*, 417; Smart and Williams, *Utilitarianism*, 34; and Temkin, "Equality, Priority, or What?" 60. For prioritarian ideas, see Sen, *On*

> *Egalitarianism*: The degree of inequality in individual welfare among the members of a population intrinsically and invariably negatively affects the social welfare of the population in such a way that had the degree of inequality been less, social welfare would have been higher (everything else being equal).[4]
>
> *Prioritarianism*: Individual welfare changes for a population's worse-faring members intrinsically and invariably affect the social welfare of the population more than equally sized changes for its better-faring members, with increases having a larger positive effect and decreases a larger negative effect on social welfare.[5]

As formulated above, egalitarianism is presented as a theory about the contribution to social welfare by a property of populations (inequality), whereas prioritarianism is presented as a theory about the contribution to social welfare by changes in individual welfare. This difference in subject is standard.

Both the egalitarian and the prioritarian presentations contain terms whose interpretation is contested: "inequality" and "worse faring." "Inequality" admits of more interpretations than can be listed here, while "worse faring" admits of at least two: a personal and an impersonal one, yielding two distinct versions of prioritarianism.[6]

> *Personal Worse Faring*: A member p of a population A is *personally worse faring* if and only if p fares worse than at least one other member of A. Furthermore, a member p_i, with welfare level w_i, fares personally worse than a member p_j, with welfare level w_j, *to the degree that* w_i is lower than w_j.
>
> *Impersonal Worse Faring*: A member p of a population A is *impersonally worse faring* if and only if p fares worse than p would with a higher level of welfare. Furthermore, a member p, with welfare level w_i, fares impersonally worse than p would fare at a higher welfare level w_j *to the degree that* w_i is lower than w_j.

The personal version of prioritarianism identifies the worse-faring members of a population relative to members of the same population, whereas the impersonal

---

Economic Inequality, 18; Scheffler, *The Rejection of Consequentialism*, 31; Nagel, *Equality and Partiality*, 70; and Parfit, "Equality and Priority," 213.

4 Similar presentations may be found in McKerlie, "Equality and Priority," 25; and Holtug, *Persons, Interests, and Justice*, 171.

5 A similar presentation may be found in Parfit, "Equality and Priority," 213.

6 Compare Hirose, *Egalitarianism*, 93.

version of prioritarianism identifies the worse-faring members of a population relative to higher levels of welfare. According to the personal version of prioritarianism, the worse-faring members are those whose welfare levels are below at least one other member's welfare level. According to the impersonal version of prioritarianism, the worse-faring members are those whose welfare levels are below some other level of welfare. Everyone who is personally worse faring is impersonally worse faring as well, although the opposite is not always the case. A population of equally faring members does not have worse-faring members in the first sense, but could have them in the second sense.

The distinction between "personal" and "impersonal" prioritarianism is related to two distinctions made by other authors. Persson makes a distinction between "relative" and "absolute" prioritarianism, which captures whether relations between welfare levels or absolute welfare values matter for social welfare. Temkin makes a similar distinction between "comparative" and "non-comparative" prioritarianism.[7] Both these distinctions are potentially misleading, since absolute welfare values matter for any prioritarian, and any type of prioritarianism can be expressed in a relational or comparative form. I will thus only use the distinctions between "personal" and "impersonal" prioritarianism here.

Personal prioritarianism could be exemplified by rank-weighted total utilitarianism, while impersonal prioritarianism could be exemplified by a theory using an additively separable concave function on individual welfare values.[8] The impersonal version of prioritarianism is favored by Parfit and the personal version is favored by Buchak.[9]

Neither of the core ideas of egalitarianism and prioritarianism presents a complete theory of social welfare. It is not sufficient to point out *what* intrinsically affects the social welfare of a population: one must also explain *how*. Egalitarianism must explain how individual welfare and inequality intrinsically affect the social welfare of populations. Prioritarianism must explain how individual welfare changes vary in their effect on social welfare depending on the members' initial degrees of worse faring. For example: Does the egalitarian think that

7 See Persson, "Equality, Priority and Person-Affecting Value," 35; and Temkin, *Inequality*, 165.

8 For a presentation of the first type, see for example Ebert, "Rawls and Bentham Reconciled," 215; and Buchak, "Taking Risks behind the Veil of Ignorance," 643–44. For a presentation of the second type, see for example Rabinowicz, "Prioritarianism for Prospects," 8–9; Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 99; Peterson and Hanson, "Equality and Priority," 301; Brown, "Prioritarianism for Variable Populations," 330; Holtug, *Persons, Interests, and Justice*, 205; Adler, *Well-Being and Fair Distribution*, 307; Broome, "Equality versus Priority," 221; and Hirose, *Egalitarianism*, 89.

9 See Parfit, "Equality or Priority?" 104; and Buchak, "Taking Risks behind the Veil of Ignorance," 610.

individual welfare and inequality affect social welfare directly and separately, or is it rather that both affect social welfare indirectly and jointly, by inequality (adversely) determining the degree to which individual welfare affects social welfare? And does the prioritarian think that individual welfare changes for worse-faring members matter more because lower welfare levels have a larger weight when individual welfare (indirectly) contributes to social welfare, or is it rather that the individual welfare levels of worse-faring members matter lexically to social welfare, as they do assuming *leximin*?[10]

A completely specified theory of what factors intrinsically affect social welfare, and how, includes a measure of social welfare, as the how question is most precisely answered in mathematical form. A measure of social welfare is, however, not sufficient in itself as a theory of social welfare, because its pure mathematical form does not clearly express anything regarding intrinsic dependence relations between social welfare and other factors (such relations can at best be inferred).[11]

Egalitarianism can be understood as a class of completely specified theories that capture the core egalitarian idea, whereas prioritarianism can be understood as a class of completely specified theories that capture the core prioritarian idea. These classes overlap, although they might not overlap completely. Even if they do not overlap, however, the classes of egalitarian and prioritarian *measures* might.[12]

The remainder of this essay will focus on the evaluative function of egalitarianism and prioritarianism, as it is fulfilled by egalitarian and prioritarian social welfare measures. First, however, I need to make some assumptions regarding the measurability of social welfare.

2. ASSUMPTIONS

In general, a measure of social welfare W is a function that assigns real numbers to all possible populations, directly representing their levels of social welfare, and indirectly representing relations between their levels of social welfare. I will not make any assumptions about whether the measure would be ratio, interval, or just ordinal scale. However, the relation *is better than* (in terms of social welfare) would be represented as irreflexive, asymmetric, and transitive, whereas

10 *Leximin* was proposed by Sen, *Collective Choice and Social Welfare*, 138. Compare Rawls, *A Theory of Justice*, 78.

11 Compare Fleurbaey, "Equality versus Priority," 205.

12 Related remarks regarding social welfare rankings have been made by Adler, *Well-Being and Fair Distribution*, 364; and Fleurbaey, "Equality versus Priority," 213.

the relation *is equally good as* (in terms of social welfare) would be represented as reflexive, symmetric, and transitive.

Since social welfare, at least in part, positively depends on individual welfare, a measure of social welfare must, at least in part, positively depend on a measure of individual welfare. This is the case whether the social welfare measure is egalitarian or prioritarian. I will thus assume that there is a measure of individual welfare w, assigning numbers to all individuals, directly representing their levels of welfare, and indirectly representing relations between their levels of welfare. The relation of *worse faring* is represented by the absolute difference between a lower and a higher degree of welfare and is irreflexive, asymmetric, and transitive (whereas the relation of *equal faring* is reflexive, symmetric, and transitive). The measure w is continuous, as well as ratio scale. For simplicity I will assume that it assigns only positive numbers.

In order to qualify as an egalitarian or prioritarian measure, a social welfare measure should assign numbers in a way that reflects the idea that inequality has a negative effect on social welfare or the idea that welfare changes for worse-faring individuals affect social welfare more (in the sense that welfare increases have a larger positive effect and welfare decreases have a larger negative effect). Such measures could take several different forms. I will consider two possibilities here.

One possibility is to use a measure that aggregates individual welfare by an additively separable, strictly concave function that gives lower welfare values larger weight. This type of measure shows social welfare to be a joint function of individual welfare and the diminishing marginal importance of individual welfare. It is the standard measure for prioritarianism since it captures the idea that welfare changes for (impersonally) worse-faring people affect social welfare more. However, it has also been used for egalitarianism since it also captures the idea that inequality has a negative effect on social welfare, at least in the comparative sense that an unequal distribution of a fixed amount of total welfare yields a lower degree of social welfare than an equal distribution does.

Another possibility is to use a derived measure that combines a measure of aggregated individual welfare with another measure that either captures the effect of inequality or the effect of worse faring. If the two measures are multiplied, such a measure would show social welfare to be a function of two interacting factors. In the egalitarian case, inequality would affect the degree to which aggregated individual welfare contributes to social welfare; in the prioritarian case, aggregated worse faring would.

I should add that I will only discuss measures that are wholly egalitarian or prioritarian. By this I mean measures that completely express either of the core ideas, most importantly the idea that inequality *invariably* has a negative effect

on social welfare or that welfare changes for worse-faring members *invariably* matter more. This does not exclude measures that express the idea that inequality or changes for the worse-faring members matter *pro tanto*. However, it does exclude measures that are only responsive to inequality between the best- and worst-faring members, and measures that only prioritize the worse-faring members at the lowest levels of welfare (like *maximin*).

3. The Argument from Conceptual Connections

The first argument for the thesis that egalitarians and prioritarians can use the same measures focuses on how the measures would be responsive to the properties that social welfare intrinsically depends on (according to these theories). Due to conceptual connections between inequality and worse faring, measures that are responsive to one property are necessarily responsive to the other (in the relevant way). Consequently, egalitarian and prioritarian measures cannot be distinguished (at least not extensionally).

The first obvious conceptual connection is between inequality and personal worse faring. An unequal population consists of members who, when paired with other members, for at least one pairing come out as one better-faring and one worse-faring member. The more unequally the pair is faring, the better faring is one member and the worse faring is the other. The second equally obvious conceptual connection is between personal and impersonal worse faring. A population with personally worse-faring members has impersonally worse-faring members as well, although the opposite is not always the case. The more personally worse faring a member p_i is, relative to another member p_j, the more impersonally worse faring the member p_i is as well, relative to the welfare level of p_j.

By virtue of purely conceptual connections, it is the case that if degrees of inequality intrinsically affect social welfare, then same-sized welfare changes for the worse-faring members of a population will instrumentally affect social welfare more, since such changes affect inequality more.[13] This is the case whether the worse-faring members are personally worse faring or impersonally worse faring. Also by virtue of purely conceptual connections, if same-sized welfare changes for the worse-faring members of a population intrinsically affect social welfare more, then same-sized changes that affect the degree of inequality more will instrumentally affect social welfare more, since such changes affect the wel-

13 Similar remarks have been made by Temkin, "Equality, Priority, or What?" 60; Parfit, "Equality or Priority?" 103; Sen and Foster, *On Economic Inequality*, 145; and Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 101.

fare levels of the worse-faring members more.[14] This is also the case whether the worse-faring members are personally worse faring or impersonally worse faring.

According to egalitarians, individual welfare and inequality intrinsically affect social welfare, and according to prioritarians, individual welfare and worse faring do.[15] An egalitarian measure will thus reflect that *inequality* (i) and *individual welfare* (w) affect W_e (egalitarian social welfare), whereas a prioritarian measure will reflect that *worse faring* (f) and *individual welfare* (w) affect W_p (prioritarian social welfare). Now, if f instrumentally and proportionally affects i, and i and w intrinsically affect W_e, then a measure of f and w can be used as a measure of W_e. Likewise, if i instrumentally and proportionally affects f, and f and w affect W_p, then a measure of i and f can be used as a measure of W_p. And since f and i instrumentally and proportionally affect each other, any egalitarian measure works as a prioritarian measure, and vice versa.

One possible objection to this argument is that it does not consider the different types of measures standardly used to represent egalitarianism and prioritarianism. Prioritarianism is usually represented by an additively separable, strictly concave function on individual welfare values, whereas egalitarianism is usually represented by a derived measure containing a measure of individual welfare and a measure of equality. Thus prioritarianism usually does not represent worse faring as a separate factor in the way that egalitarianism usually represents equality as a separate factor. This difference is not brought up in the argument above and might affect the interchangeability of egalitarian and prioritarian measures.

However, the argument above does not presuppose any particular kind of measure. It does not presuppose that either the egalitarian or the prioritarian measure is a derived measure that, for example, conjoins two separate measures, one of individual welfare, and one of either inequality or worse faring. What the argument presupposes is only that the egalitarian or prioritarian measure is appropriately affected by the relevant properties. If a certain method of aggregating individual welfare makes a measure sensitive to inequality or worse faring in an adequate way, it would qualify as a measure of i and w or f and w, no matter what type of measure is used.

Another possible objection to the above argument is that it misrepresents at least the impersonal version of prioritarianism and thus the relation between egalitarian and prioritarian measures. It is not really that the relation of worse faring affects social welfare, a prioritarian might say, but rather that changes of

14 Similar remarks have been made by Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 101; and Fleurbaey, "Equality versus Priority," 207.

15 The egalitarian remark has previously been made by McCarthy, "Distributive Equality," 1047.

people's lower welfare levels affect social welfare more than changes of people's higher welfare levels. This matter can be expressed without the use of any relations, and is thus independent of any relations.

I concede that impersonal prioritarianism can be expressed without reference to the relation of worse faring. However, as long as impersonal prioritarianism *could* be expressed with reference to the relation of worse faring and this relation has the relationship to inequality as described above, the above reasoning still applies.

Yet another possible objection to the above argument is that it misrepresents the relationship between inequality and worse faring and thus the relation between egalitarian and prioritarian measures. More precisely, the objection is this: an unequal population necessarily contains members who are worse faring and better faring. But an egalitarian cannot say that welfare changes for the worse-faring members affect social welfare more, since, with respect to inequality, the worse-faring and the better-faring members cannot be separated as obstacles to social welfare. Supposedly, adding 1 in welfare to the best-faring member increases inequality by as much as adding 1 in welfare to the worse-faring member decreases inequality; and subtracting 1 in welfare from the best-faring member decreases inequality by as much as subtracting 1 in welfare from the worse-faring member increases inequality. So, on this view, an egalitarian should think that welfare changes for the best-faring and the worst-faring members affect social welfare the most, whereas welfare changes for middle-faring members affect social welfare the least. In contrast, the prioritarian should think that welfare changes for the worst-faring members affect social welfare the most, whereas welfare changes for the best-faring members affect social welfare the least. Thus, if a population A has the welfare vector $\mathbf{v}_A = (3, 2, 1)$, an egalitarian should claim that welfare changes for the person with 3 or 1 in welfare affect social welfare the most, whereas a prioritarian should claim that welfare changes for the person with 1 in welfare affect social welfare the most, while welfare changes for the person with 3 in welfare affect social welfare the least (assuming that the changes are equal). This difference should be reflected by egalitarian and prioritarian measures, and so, according to the present objection, the measures are not interchangeable.

However, this objection does not hold up to scrutiny. Even if an egalitarian would claim that welfare changes for the worse-faring and the best-faring members affect *inequality* equally (which many egalitarians would not, by the way), an egalitarian cannot claim that welfare changes for the worse-faring and the best-faring members affect *social welfare* equally. This could be shown in different ways. One way to show it is just to note that an egalitarian cares about in-

dividual welfare in addition to equality. So, when 1 is added to the best-faring member, this is good in a way, and bad in another, whereas when 1 is added to the worst-faring member, this is only good. (Likewise: when 1 is subtracted from the best-faring member, this is bad in a way and good in another, whereas when 1 is subtracted from the worst-faring member, this is only bad.) Thus changes to the worst-faring and the best-faring members cannot affect social welfare equally.

This point can be illustrated with an example. Let us suppose that degrees of inequality are identified with total welfare differences (similar to a proposal by Rabinowicz).[16] It is then correct that subtracting 1 from the best-faring member decreases inequality by as much as subtracting 1 from the worst-faring member increases inequality. In the example with the population with welfare vector (3, 2, 1), the absolute changes in total welfare differences are the same whether 1 is added to the best-faring member or subtracted from the worst-faring member, subtracted from the best-faring member or added to the worst-faring member. The difference is always 2. However, if we then measure social welfare by subtracting total welfare differences from total welfare, we get the following results: starting with $\mathbf{v}_A = (3, 2, 1)$, we get the welfare vectors $\mathbf{v}_{Ab+} = (4, 2, 1)$, $\mathbf{v}_{Ab-} = (2, 2, 1)$, $\mathbf{v}_{Aw+} = (3, 2, 2)$ and $\mathbf{v}_{Aw-} = (3, 2, 0)$ and the values: $W(A) = 6 - 4 = 2$, $W(A^{b+}) = 7 - 6 = 1$, $W(A^{b-}) = 5 - 2 = 3$, $W(A^{w+}) = 7 - 2 = 5$ and $W(A^{w-}) = 5 - 6 = -1$.

Since the absolute difference in social welfare when the best-faring member gains or loses welfare is 1, but the absolute difference in social welfare when the worse-faring member gains or loses welfare is 3, welfare changes for the worse-faring members affect social welfare more, even according to this measure. (If we would divide total welfare with total welfare differences we would get the same result.)

4. THE ARGUMENT FROM MINIMAL CONDITIONS

The second argument for the thesis that egalitarians and prioritarians can use the same measures focuses on how the measures should rank possible populations in order to capture the core egalitarian and prioritarian ideas. The argument is that there is no difference between egalitarian and prioritarian measures in this respect.

In order to assess how the two theories should rank possible populations, I will begin this section by formulating two ranking conditions, one for egalitarian and one for prioritarian measures. Let us look at the egalitarian ranking condition first.

16 See Rabinowicz, "The Size of Inequality and Its Badness," 62.

4.1. The Egalitarian Condition

An egalitarian measure should rank populations in a way that reflects that inequality affects social welfare negatively. For many comparative cases egalitarians would disagree as to which population is most unequal. The following ranking condition, however, should be generally acceptable:

> *Egalitarian Condition*: For a measure of social welfare W and for all possible populations A and B and their members $p_i \in A$ and $q_i \in B$, such that $|A| = |B|$ and $\Sigma w(p_i) = \Sigma w(q_i)$, *if* there is a bijection from A to B, such that each individual $p_i \in A$ could be paired with an individual $q_i \in B$ so that for each pair of individuals (p_i, q_i) it is the case that $w(p_i) = w(q_i)$, except for four individuals: p_1, p_2, q_1, q_2, such that $|w(p_1) - w(p_2)| < |w(q_1) - w(q_2)|$, *then A does better than B*, and thus $W(A) > W(B)$.

Less formally, the condition states that if the total welfare and cardinality are equal between two populations, and all individual welfare values are equal, apart from the welfare of one pair of persons, the population with the more equal-faring pair is better.

The Egalitarian Condition is most similar to the well-known *Pigou-Dalton Condition* (although this condition concerns welfare transfers and outcomes).[17] It is also slightly similar to *Hammond's Equity Condition* (although that condition does not require the same total sum).[18] The first similarity will be relevant later.

The Egalitarian Condition is a restricted condition in the sense that it applies only to comparisons between two populations that are similar in all respects except for the welfare of one pair of persons, where one pair fares more equally than the other. However, assuming that *better-than* is a transitive relation, the condition implies that for comparisons between populations with the same cardinality and total welfare, the population where everyone fares equally well is the best population. The condition also implies that, for the same comparison, the population where one person has all welfare and the others have none is the worst population.

I take it that the Egalitarian Condition is *necessary* for a social welfare measure to qualify as egalitarian. This idea would be entirely uncontroversial if the condition concerned only comparisons between populations of two persons. Since it does not, someone might object that whether A should be regarded as more equal than B depend on the welfare levels of the persons not being com-

17 See Pigou, *Wealth and Welfare*, 27; and Dalton, "The Measurement of the Inequality of Incomes," 351.

18 See Hammond, "Equity, Arrow's Conditions, and Rawls' Difference Principle," 795.

pared. If the other persons in A and B seem to fare more like q_1 and q_2 than like p_1 and p_2, perhaps B should be regarded as more equal than A (for example, when the welfare values in A are $(8, 5, 4, 1)$ and the welfare values in B are $(8, 8, 1, 1)$).

To this objection one may reply that the inequality resulting from the larger difference between q_1 and q_2 simply cannot be compensated for by similar or equal differences between the other members of B. This point is most clearly illustrated by looking at welfare differences. In the above example, the welfare differences between the welfare levels of the members of B are larger than they are between the members of A. Even though the welfare levels 8 and 1 seem to be more similar to the levels 8 and 1 than the levels 5 and 4 seem to be, they are overall more different. This fact does not conclusively show that A is more equal than B, since the relationship between welfare differences and inequality may be more complicated than mere aggregation. However, considering that welfare differences ground inequality, this fact strongly supports the claim that A is more equal than B.

Let us thus proceed to consider whether the Egalitarian Condition is also *sufficient* for a social welfare measure to qualify as egalitarian (assuming that we are only considering measures that could qualify as social welfare measures at all). In order to support the claim that it is sufficient, we could argue that a social welfare measure that satisfies the condition cannot rank populations in an obviously non-egalitarian way. This argument requires, for a start, that we identify all obviously non-egalitarian rankings of populations (including populations that differ from one another in size and total welfare). Since there are many different ways to measure inequality, the only obviously non-egalitarian rankings (besides the ones directly contradicting the condition) are the extreme ones, that is: the maximal and minimal equality cases. Similar comments apply to both, so let us focus on the minimal case.

One might think that an obviously non-egalitarian ranking would be one where a population in which one person has all welfare and the rest have none is ranked above a population in which this is not the case. However, this is too quick. It is not obviously non-egalitarian to make this kind of ranking because an egalitarian may care about factors other than equality, such as total amount of welfare, average level of welfare, or number of well-faring people. Thus, even an egalitarian may rank $(20, 0)$ above $(0, 0)$ or (0), for example.

The only obviously non-egalitarian minimal equality ranking is thus the one where, *everything else being equal* (total welfare and size of population), a population in which one person has all welfare and the rest have none is ranked above a population where this is not the case. Likewise: the only obviously non-egalitarian maximal equality ranking is the one where, everything else being equal, a population where all persons have the same amount of welfare is ranked below

a population where this is not the case. And both of these rankings are excluded by the Egalitarian Condition.

If I am correct that there are no other obviously non-egalitarian social welfare rankings, then the Egalitarian Condition is both sufficient and necessary for identifying a social welfare measure as egalitarian.

4.2. The Prioritarian Condition

A prioritarian measure should rank populations in a way that reflects that welfare changes for the worse-faring individuals affect social welfare more. The following ranking condition should be uncontroversial:

> *Prioritarian Condition*: For a measure of social welfare W and for any possible population C and for any individuals $r_i, s_i \in C$ such that $w(r_i) < w(s_i)$ and $w(r_i) \geq 0$ and $w(s_i) \geq 0$, *if* it is possible to either increase the welfare of r_i by m, resulting in population C^*, or increase the welfare of s_i by m, resulting in population C^{**}, *then* C^* *does better than* C^{**} and thus $W(C^*) > W(C^{**})$.

Less formally, the condition states that given the choice between increasing the welfare of either of two persons by the same amount, it is better to increase the welfare of the worse-off person.

The Prioritarian Condition is a variant of the Pigou-Dalton Condition (mentioned earlier).[19] It is also similar to conditions previously proposed by Sen, Weirich, Parfit, and Vallentyne.[20] Because the Prioritarian Condition only applies to comparisons between two possible populations that result from changes to the same population, it has a rather limited application.

That the Prioritarian Condition is *necessary* for a social welfare measure to qualify as prioritarian seems indisputable.[21] If a measure would not give the result that it would be better to increase the welfare of a worse-off person by m, rather than a better-off person by the same amount m, then it would not be prioritarian. It is less obvious that the Prioritarian Condition is also *sufficient* for a social welfare measure to qualify as prioritarian (even if we, once again, only

19 See Pigou, *Wealth and Welfare*, 27; and Dalton, "The Measurement of the Inequality of Incomes," 351.

20 See Sen, *On Economic Inequality*, 18; Weirich, "Utility Tempered with Equality," 431; Parfit, "Equality and Priority," 213; and Vallentyne, "Equality, Efficiency and the Priority of the Worse-Off," 1.

21 Temkin, Tungodden, Adler, Fleurbaey, and McCarthy agree. See Temkin, *Inequality*, 64; Tungodden, "The Value of Equality," 28, and "Equality and Priority," 424; Adler, *Well-Being and Fair Distribution*, 356; Fleurbaey, "Equality versus Priority," 207; and McCarthy, "Risk-Free Approaches to the Priority View," 432.

consider measures that could qualify as social welfare measures at all). In order to support the claim that it is sufficient, we can use the same type of reasoning as in the egalitarian case: we can argue that a measure that satisfies the condition cannot rank possible population changes in an obviously non-prioritarian way. This argument requires identifying all obviously non-prioritarian rankings of possible changes. There are three candidates for such non-prioritarian rankings.

The first possibly (or rather obviously) non-prioritarian ranking is one where an increase in the welfare of a better-off person by m is preferred over an increase of the welfare of a worse-off person by m. This ranking is directly excluded by the Prioritarian Condition.

The second possibly non-prioritarian ranking is one where an increase in the welfare of a better-off person by a higher amount n is preferred over an increase of the welfare of a worse-off person by a lower amount m. But this ranking is not obviously non-prioritarian. It does not go against prioritarianism generally to regard an increase of total welfare or average welfare as more important than prioritizing the worse-faring person (for example by choosing $(8, 4)$ rather than $(5, 6)$ as a change from $(5, 4)$).

The third possibly non-prioritarian ranking is one where an increase in the welfare of a better-off person by a lower amount m is preferred over an increase of the welfare of a worse-off person by a higher amount n. This type of ranking can be separated into two cases. In the first case, the addition of n to the welfare of a worse-faring person does not make that person better off than the better-faring person. In the second case, the addition of n to the welfare of a worse-faring person does make that person better off than the better-faring person.

When the addition of n to the welfare of a worse-faring person does not make the worse-faring person better off than the better-faring person, the third type of ranking is obviously non-prioritarian. However, this ranking is excluded by the Prioritarian Condition being consecutively applied to hypothetical choices. Choosing between increasing the welfare of a better-faring person by a lower amount m and increasing the welfare of a worse-faring person by a higher amount $n = m + k$, can be described as first hypothetically choosing between increasing the welfare of either a worse-faring or a better-faring person by m, and then hypothetically choosing between increasing the welfare of either a worse-faring or a better-faring person by k. For both choices the condition will reward raising the worse-faring rather than the better-faring person, and thus reward raising the worse-faring person by n. (The choice of $(5, 5)$ over $(6, 3)$, from $(5, 3)$ is thus done, first by choosing $(5, 4)$ over $(6, 3)$ and then by choosing $(5, 5)$ over $(6, 4)$.)

When the addition of n to the welfare of a worse-faring person does make the

worse-faring person better off than the better-faring person, the third type of ranking is not obviously non-prioritarian. This is because it is not clear that we are prioritizing the better-faring person, when during the change, the better-faring person *becomes* the worse-faring person. Thus: while it is obviously non-prioritarian to rank (6, 3) as a better change than (5, 5) from (5, 3), it is not obviously non-prioritarian to rank (6, 3) as a better change than (5, 9) from (5, 3), for example.

Since the Prioritarian Condition excludes two obviously non-prioritarian rankings, and there are no other obvious such rankings, the condition is plausibly sufficient for identifying a social welfare measure as prioritarian as well.

The conclusion of this section is thus that the Egalitarian Condition is both necessary and sufficient for a social welfare measure to qualify as egalitarian, whereas the Prioritarian Condition is both necessary and sufficient for a social welfare measure to qualify as prioritarian. However, as might be obvious, the two conditions are equivalent. (A proof of this is included in the appendix.) Thus, since the Egalitarian Condition is both necessary and sufficient to identify an egalitarian social welfare measure and the Prioritarian Condition is both necessary and sufficient to identify a prioritarian social welfare measure, and the two conditions are equivalent, then any social welfare measure qualifying as egalitarian will also qualify as prioritarian, and vice versa.

5. The Argument from Standard Measures

Someone may object to the above analysis, however, that the minimal conditions present the theories as too abstract. The differences between the two theories and their measures would be more clearly visible if we looked at standard egalitarian and prioritarian measures directly.

This possible objection leads me to the third argument for the thesis that egalitarians and prioritarians can use the same measures: both theories can use the same *standard* egalitarian and prioritarian measures. To show this, I will first present the measures and then argue that they could be used for either theory. (Both measures satisfy the minimal conditions.)

5.1. A Standard Prioritarian Measure

Let us first consider a standard type of prioritarian measure. It is not the only prioritarian measure in the literature, but it is often presented as *the* prioritarian measure.[22] It measures social welfare by aggregating individual welfare through a strictly concave function, as follows:

22 See Rabinowicz, "Prioritarianism for Prospects," 8–9; Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 99; Brown, "Prioritarianism for

$$PW(A, w) = \sum_{i=1}^{n} f(w(p_i)),$$

where A is a population, w is a measure of individual welfare, $n = |A|$, f is a strictly concave function, and p_i is an indexed individual such that $p_i \in A$.

Since the concave function is not specified above, the *PW* measure is, strictly speaking, a class of measures. Let me give an example of how a *PW* measure might work. If the strictly concave function is a root function, and some population A only has four members, with their welfare levels represented by the vector $\mathbf{v}_A = (9, 16, 0, 4)$, then *PW* would assign A the social welfare value $3 + 4 + 0 + 2 = 9$. By comparison, the population B with the vector $\mathbf{v}_A = (0, 25, 0, 4)$ would be assigned the social welfare value 7, and would thus be lower ranked.

The *PW* measure is considered suitable for prioritarianism since it gives lower welfare values larger weight, thus giving welfare changes for the worse-faring people larger weight as well.

5.2. A Standard Egalitarian Measure

Let us next consider a standard type of egalitarian measure. There is no egalitarian measure known as *the* egalitarian measure, so as a standard egalitarian measure I will choose a mixture of several previous proposals. Its general form is similar to measures proposed by Jensen, Fleurbaey, and Peterson and Hansson, and illustrates the common idea that an egalitarian measure should incorporate a measure of equality, multiplied or added to a measure of individual welfare.[23] The form of the measure is as follows:

$$EW(A, w) = \sum_{i=1}^{n} w(p_i)(1 - I(A, w)),$$

where A is a population, w is a measure of individual welfare, $n = |A|$, p_i is an indexed individual such that $p_i \in A$, and I is a measure of inequality such that $0 \leq I(A, w) \leq 1$.

The *EW* measure multiplies the total sum of individual welfare with a measure of equality $(E(A, w) = 1 - I(A, w))$. Because there are several measures of inequality that can be used for I, the *EW* measure is a class of measures as well.[24] For the measure to satisfy the Egalitarian Condition, the inequality measure should take

Variable Populations," 330; Holtug, *Persons, Interests, and Justice*, 205; Adler, *Well-Being and Fair Distribution*, 307; Hirose, *Egalitarianism*, 89; and Broome, "Equality versus Priority," 221.

23 See Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 94; Fleurbaey, "Equality versus Priority," 207–8; and Peterson and Hansson, "Equality and Priority," 307.

24 See for example inequality measures by Gini, "Variabilità e mutabilità"; Pietra, "Delle Relazioni tra gli Indici di Variabilità"; and Theil, *Economics and Information Theory*.

differences between all members into account (as opposed to, for example, just the best- and the worst-faring member). I will use a very simple inequality measure here, based on a measure proposed by Rabinowicz and Arrhenius.[25] The inequality $I(A, w)$ will be defined as the ratio between the total sum of welfare differences between the individuals in A and the total sum of welfare differences of A_U, which is the maximally unequal possible population that has the same cardinality and total welfare as A (more precisely, $|A_U| = |A|$, and if $q_i \in A_U$ and $p_i \in A$, then $\Sigma w(q_i) = \Sigma w(p_i)$, and for one individual q_1, $w(q_1) = \Sigma w(q_i)$). The measure $I(A, w)$ is thus a proportional measure:

$$I(A, w) = \frac{D(A, w)}{D(A_U, w)},$$

where

$$D(A, w) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |w(p_i) - w(p_j)|}{2},$$

where A and A_U are populations (A_U defined above), w is a measure of individual welfare, $n = |A|$, and p is an individual, such that $p \in A$ (indexed twice as p_i and p_j).

One may note that $0 \leq I(A, w) \leq 1$, and that $D(A_U, w) = \Sigma w(p_i)(n - 1)$.

Instead of using $I(A, w)$ as defined above, we could also use the well-known *Gini Inequality Measure* in the egalitarian measure.[26] In terms of construction, this means that we would use something similar to $D(A, w)$, exchanging 2 in the denominator for $2n^2\mu$ (where μ is the mean value of $w(p_i)$, that is $(1/n)\sum_{i=1}^{n} w(p_i)$), for the inequality measure. This version of the *EW* measure would also satisfy the Egalitarian Condition and has similar properties to the presented measure (apart from satisfying the *Pareto Condition*). Since it is less of a contrast to the *PW* measure, I will not focus on this version of the *EW* measure here, however.

5.3. Double Uses for the Two Measures

The *EW* and *PW* measures differ most fundamentally in their structure. The *PW* measure is an additively separable strictly concave function on individual welfare values, whereas the *EW* measure is a product measure of two factors, where one factor is the total sum of individual welfare and the other factor is a measure of equality. As a result, absolute levels of individual welfare determine the degree to

25 See Rabinowicz, "The Size of Inequality and Its Badness," 62; and Arrhenius, "Egalitarian Concerns and Population Change," 79.

26 See Gini, "Variabilità e mutabilità."

which individual welfare contributes to social welfare for the *PW* measure, whereas degrees of inequality determine the degree to which total welfare contributes to social welfare for the *EW* measure (where in cases of maximal equality, social welfare is equal to the total sum and in cases of minimal equality, social welfare is 0).

The two measures seem to capture their respective theory perfectly: the *PW* measure gives changes of lower welfare values larger weights, as is suitable for prioritarianism, while the *EW* measure gives inequality a negative weight, as is suitable for egalitarianism. However, this does not mean that each measure can be used for only one theory. In fact, both measures seem to work for both theories.

The *PW* measure gives lower welfare larger weights and thus gives changes for people with lower welfare larger weights, as is appropriate for prioritarianism. But by giving lower welfare larger weight, inequality is punished in comparison to equality, and thus the measure is appropriate for egalitarianism as well. In fact, the *PW* measure shows up in the literature both as a prioritarian and as an egalitarian measure. Holtug, Hirose, and Broome present the measure as prioritarian, while Sen, Weirich, and (an earlier) Broome present the measure as egalitarian.[27] If I am correct that any measure that can be used by one theory can be used by the other, this double use is entirely appropriate. (Leximin has the same type of double use.)[28]

The *EW* measure is in part a function of a measure of equality, so it is appropriate for egalitarianism. But it may be appropriate for prioritarianism as well—at least the personal version.[29] Let me explain this idea.

The *EW* measure may be regarded as appropriate for egalitarianism since it represents the idea that it matters for social welfare whether individuals fare equally. However, the measure may also be regarded as appropriate for prioritarianism, since it also represents the idea that the worse-off individuals matter more for social welfare. Just like the egalitarian idea can be represented by multiplying an *equality* value with the total sum of welfare, the prioritarian idea can be represented by multiplying an aggregated *lack of worse faring* value with the total sum of welfare. For the *EW* measure to work both as an egalitarian and a prioritarian measure, it thus suffices to show that both equality and aggregated

27 See Holtug, *Persons, Interests, and Justice*, 205; Hirose, *Egalitarianism*, 89; Broome, "Equality versus Priority," 221; Sen, *On Economic Inequality*, 20; Weirich, "Utility Tempered with Equality," 433; and Broome, *Weighing Goods*, 179.

28 For example, leximin has been proposed as a prioritarian measure by Arneson, "Luck Egalitarianism and Prioritarianism," 341; Crisp, "Equality, Priority, and Compassion," 752; and Esposito and Lambert, "Poverty Measurement," 117; and as an egalitarian measure by Hammond, "A Note on Extreme Inequality Aversion," 465–66; Tungodden, "The Value of Equality," 14; and Bosmans, "Extreme Inequality Aversion without Separability," 592.

29 It is, in fact, similar to a prioritarian measure proposed by Fleurbaey, "Equality versus Priority," 207–8.

lack of worse faring can be measured by the measure $1 - I(A, w)$, and thus that both inequality and aggregated worse faring can be measured by $I(A, w)$. It is easy to show that they can. First we may note that we could measure aggregated worse faring by summing the welfare differences between the worse-faring and the better-faring persons. If we do, we would get a measure that is equivalent to $D(A, w)$, since $D(A, w)$ aggregates all welfare differences but divides them by two, and counts the welfare differences between equally well-faring persons as zero. Thus, $D(A, w)$ works just as well as a measure of aggregated worse faring as it works as a measure of inequality. Since $I(A, w)$ is just a function of the measure $D(A, w)$, it works just as well as a measure of proportional worse faring as it works as a measure of proportional inequality. And since $EW(A, w)$ is just a function of $I(A, w)$ and the total sum of welfare, it works just as well as a prioritarian measure as it works as an egalitarian measure. (Had we used the Gini Inequality Measure instead of $I(A, w)$, we could have used the fact that the Gini measure is $D(A, w)$ multiplied by $1/n^2\mu$ to make the same argument.)

Thus, both measures seem to work as egalitarian and as prioritarian measures of social welfare. Both of them capture the ideas that inequality and worse faring have a negative effect on social welfare and both of them reward lack of inequality and improvements for worse-faring persons more than improvements for better-faring persons.

6. Arguments from other conditions and features

However, one may object to the above analysis by pointing out that there are other important differences between the two standard measures that could reflect important differences between egalitarianism and prioritarianism as well. Three such features that could be used to distinguish the two types of measures are: pareto satisfiability, level sensitivity, and relationality (closely related to non-separability).

The first property has to do with the importance of individual welfare increases for social welfare. A measure that satisfies Pareto evaluates all welfare increases as good. The property thus reflects the idea that it is more important that each individual fare as well as possible than that total welfare has a certain distribution (such as everyone faring equally well or the worse-off individuals faring better). The *PW* measure satisfies Pareto, whereas the *EW* measure does not.

The second property has to do with the importance of absolute levels of welfare for the badness of inequality (or worse faring) for social welfare. A level-sensitive measure reflects the idea that it is worse for social welfare with inequality

(or worse faring) at lower levels of welfare. The *PW* measure is level sensitive, whereas the *EW* measure is not.

The third property has to do with the importance of relations between the welfare levels of different individuals to social welfare. A relational measure reflects the idea that welfare differences between individuals by themselves negatively affect social welfare (as a separate factor). The *EW* measure is relational, whereas the *PW* measure is not.

The property of relationality is closely related to a fourth property: non-separability. A separable measure assesses the contribution to social welfare from each member of a population independently of all other members of the population (which excludes relationality).[30] The *PW* measure is separable, whereas the *EW* measure is not.

The first three properties are independent of one another, and thus there are eight different possible combinations of them and their opposites. I will not discuss all combinations here, however. Instead I will ask for each one of the three properties whether both egalitarianism and prioritarianism can incorporate the property in question as well as its opposite. (The fourth property will be discussed together with the third, as they are closely related.)

My final three arguments for the thesis that egalitarians and prioritarians can use the same measures are that it seems reasonable for each of the three properties that both egalitarian and prioritarian measures can incorporate the property as well as its opposite. (These arguments can be regarded as counterarguments to arguments that there is some property that can distinguish between egalitarian and prioritarian measures.)

6.1. The Argument concerning Pareto Satisfaction

A first proposal for distinguishing between prioritarian and egalitarian measures is to suggest that a prioritarian measure should satisfy a Pareto condition and that an egalitarian measure should not. This is similar to a proposal by Parfit who claimed that a prioritarian must accept Pareto, but that an egalitarian need not.[31] Both proposals can be illustrated by the fact that the *PW* measure satisfies the condition and the *EW* measure does not.

30 Sen claims that relationality excludes separability (*On Economic Inequality*, 41). Adler and McCarthy claim that the distinction between relational and non-relational social welfare measures should be understood as the distinction between non-separable and separable social welfare measures; see Adler, *Well-Being and Fair Distribution*, 363; and McCarthy, "Risk-Free Approaches to the Priority View," 431.

31 See Parfit, "Equality or Priority?" 118. Tungodden agrees with the claim about prioritarians ("The Value of Equality," 28).

The Pareto Condition states that increasing individual welfare is invariably good, and thus implies that increasing individual welfare is more important than retaining equality. It may be put as follows:

> *Pareto Condition*: For a measure of social welfare W and for any possible population A and for any individual $p_i \in A$, whose welfare is represented by the individual welfare function w, *if* A^* would result from raising the welfare of at least one p_i, without lowering the welfare of any p_i, *then* A^* *does better than* A, and thus $W(A^*, w) > W(A, w)$.[32]

The *PW* measure would never rank A^* below A since it is a strictly increasing function on individual welfare values. But the *EW* measure might rank A^* below or equal to A, when the degree of inequality is larger in A^* than in A. (For example, when (1, 1) becomes (2, 1), the *EW* measure gives both populations a value of 2.)

The Pareto Condition is related to other conditions that have been discussed in the same context, such as the *Dominance Condition*, which states that a population that dominates another in terms of individual welfare is better. All such related conditions reflect the same idea: that it is more important for social welfare that each individual fare as well as possible than it is that individuals fare equally well. A measure that does not satisfy these conditions faces the leveling-down objection: the critique that a measure should not rank welfare losses as improvements, even when the losses result in everyone faring more equally. (This objection could be put either in terms of overall improvements or in terms of one-aspect improvements. Because a measure only registers overall improvements, only the first critique is relevant for the purpose of measurement.)[33]

The question here is whether egalitarians should reject Pareto, while prioritarians should accept it. One possible answer is that egalitarians should reject Pareto because equality must at some point be more important than raising individual welfare, if it should be of sufficient importance for an egalitarian. Prioritarians, however, can accept Pareto because welfare changes for worse-off people can have a sufficiently large weight without it being a problem that the welfare levels of better-off people are raised.

However, this argument is unconvincing. If accepting Pareto would give equality insufficient weight, there is no reason to think that it would not also give changes for the worse-faring people insufficient weight. And if accepting Pareto would give changes for the worse-faring people sufficient weight, there

32 The Pareto Condition was first proposed by Pareto. See Pareto, *Manuel d'Economie Politique*, 33. The condition is similar to Broome's *Principle of Personal Good* (*Weighing Goods*, 165).

33 A similar objection was first brought up by Nozick, *Anarchy, State, and Utopia*, 229. See also Parfit, "Equality or Priority?" 105.

is no reason to think that it would not also give equality sufficient weight. For both theories, accepting Pareto has consequences. An egalitarian would have to accept that some equality losses make a population better, as when $\mathbf{v}_A = (1, 1, 1, 1)$ becomes $\mathbf{v}_{A^*} = (4, 1, 1, 1)$. A prioritarian, in turn, would have to accept that sometimes it is better to give smaller benefits to better-faring persons than larger to worse-faring persons, as when $\mathbf{v}_{A^*}$ is preferably changed into $\mathbf{v}_{A^{**}} = (9, 4, 1, 1)$ rather than into $\mathbf{v}_{A^{***}} = (4, 8, 1, 1)$.

Another argument to the same effect is that an egalitarian should consider inequality equally bad, no matter its direction (relative to an equal-faring majority), and a prioritarian should not. Thus, an egalitarian should consider a population A with the welfare vector $\mathbf{v}_A = (2, 1, 1, 1)$ to be equally good as a population B with a welfare vector $\mathbf{v}_B = (1, 1, 1, 0)$, whereas a prioritarian should consider B to be worse. Since Pareto requires that A is ranked above B, an egalitarian must reject it.

But this argument is unconvincing as well. It presupposes that an egalitarian would not consider individual welfare important in addition to equality, and there are no egalitarians like that. (Why would an egalitarian consider equality of individual welfare good for social welfare if she did not consider individual welfare good for social welfare in itself?)[34] Even though an egalitarian (let us suppose) would consider A and B equally good in terms of *inequality*, an egalitarian need not consider A and B equally good *overall*. And overall goodness is the only thing that a social welfare measure registers.

Obviously, satisfaction of the Pareto Condition could be proposed as a dividing line between egalitarianism and prioritarianism. However, it has already been used as a dividing line between different kinds of egalitarianism: *strong egalitarianism* that does not satisfy Pareto and *moderate egalitarianism* that does.[35] Since moderate egalitarianism is accepted as a kind of egalitarianism, and is also much more popular than the strong kind, it seems inappropriate to distinguish between egalitarianism and prioritarianism on the basis of Pareto. In fact, considering that the Pareto Condition is often treated as a necessary condition for a plausible social welfare measure, one could even argue that using Pareto as a dividing line between egalitarianism and prioritarianism would give prioritarianism an unfair (and unwarranted) advantage.[36] (Tungodden, Christiano and

34 Compare McCarthy, "Distributive Equality," 1047.

35 See Parfit, "Equality and Priority," 218.

36 For the claim about Pareto being a necessary condition, see Deschamps and Gevers, "Leximin and Utilitarian Rules," 144; Blackorby, Bossert, and Donaldson, "The Axiomatic Approach to Population Ethics," 346; and Tungodden, "The Value of Equality," 18. Compare Broome, *Weighing Goods*, 200.

Brayen, Holtug, Hirose and Broome all think that egalitarians should accept Pareto. Only Nagel and (perhaps) Temkin think that egalitarians need not.)[37]

6.2. The Argument concerning Level Sensitivity

A second proposal for distinguishing between prioritarian and egalitarian measures is to suggest that a prioritarian measure should be sensitive to *absolute levels* of welfare and that an egalitarian measure should not, so that worse faring is represented as worse at lower levels, but inequality is not. This is illustrated by the fact that the *PW* measure gives larger weight to welfare changes for worse-off persons at lower welfare levels, whereas the *EW* measure gives the same weight to inequality at any welfare level. Thus, only the *PW* measure satisfies the following condition:

> *Level-Sensitivity Condition*: For a measure of social welfare W and for any possible populations A, B, C, D with just two members, and their members: $p_i \in A$, $q_i \in B$, $r_i \in C$ and $s_i \in D$, whose welfare is represented by the individual welfare function w, *if* $\Sigma w(p_i) = \Sigma w(q_i)$ and $\Sigma w(r_i) = \Sigma w(s_i)$, and A and C are equal populations because $w(p_1) = w(p_2)$, and $w(r_1) = w(r_2)$, whereas B and D are unequal populations because $w(q_1) > w(q_2)$ and $w(s_1) > w(s_2)$, and $|w(q_1) - w(q_2)| = |w(s_1) - w(s_2)|$, but $w(q_1) > w(s_1)$, *then* $|W(A, w) - W(B, w)| < |W(C, w) - W(D, w)|$.

The distinction between level sensitivity and level insensitivity has previously been brought up by both Temkin and Rabinowicz in a discussion regarding the badness of inequality.[38]

For a measure that satisfies the Level-Sensitivity Condition, the loss of social welfare due to inequality (or worse-off people) is worse at lower levels of welfare. Thus, the loss of social welfare for B, in comparison to D, is worse than the loss of welfare for A, in comparison to C, when the members of B fare worse than the members of A.

The strict concavity of the *PW* measure assures that the difference between $W(A, w)$ and $W(C, w)$ is always smaller than the difference between $W(B, w)$ and $W(D, w)$. However, for the *EW* measure the difference between these values is always the same.

Even though we can distinguish between the *PW* measure and the *EW* mea-

37 See Tungodden, "The Value of Equality," 18; Christiano and Brayen, "Inequality, Injustice, and Levelling Down," 392; Holtug, *Persons, Interests, and Justice*, 171; Hirose, "Reconsidering the Value of Equality," 306; Broome, "Equality versus Priority," 220; Nagel, *Equality and Partiality*, 107; and Temkin, *Inequality*, 78.

38 See Temkin, "Equality, Priority, or What?" 160; and Rabinowicz, "The Size of Inequality and Its Badness," 67.

sure relative to their level sensitivity, it seems unsuitable to distinguish between egalitarianism and prioritarianism in this way. On the one hand, both egalitarians and prioritarians care about how people fare, and obviously think that lower levels of welfare are worse. It is a natural extension of this idea that also inequality or worse faring is worse at lower levels. On the other hand, it is not a necessary extension of this idea, so an egalitarian could also hold that inequality is equally bad no matter how well people fare, and at least a *personal* prioritarian could hold that worse faring is equally bad no matter how well people fare. It is thus perfectly possible to formulate either egalitarianism or prioritarianism as either level-sensitive or level-insensitive theories. In fact, this has already been done. Temkin has proposed a level-sensitive version of egalitarianism, while Hirose has presented a level-insensitive version.[39] As far as prioritarianism is concerned, Parfit's version is level sensitive, whereas Buchak's version is level insensitive.[40]

Considering that both egalitarianism and prioritarianism come in level-sensitive and level-insensitive versions, it would be inappropriate to distinguish between egalitarianism and prioritarianism on the basis of level sensitivity.

6.3. The Argument concerning Relationality and Separability

A third proposal for distinguishing between prioritarian and egalitarian measures is to suggest that an egalitarian measure should be responsive to *relations* between welfare levels of different persons and that a prioritarian measure should not. This is illustrated by the *EW* measure being a function of relations between different welfare levels, which the *PW* measure is not. The *EW* measure is thus *relational* while the *PW* measure is not. Because of this, the *PW* measure represents each individual as contributing *separately* to social welfare, so that the contribution to social welfare from each member's welfare is independent of the welfare of the other members, whereas the *EW* measure represents each individual as contributing non-separately to social welfare, so that the contribution to social welfare from each member's welfare is dependent on the welfare of the other members. It is thus possible to change the welfare of a member p of some population A from w_1 to w_2 without this change affecting the value of $PW(A, w)$ via anything other than the difference between w_1 and w_2. This is not the case for $EW(A, w)$. The *PW* measure is thus *separable* while the *EW* measure is not. Only the *PW* measure satisfies the following condition:

Separability Condition: For a measure of social welfare W and for any

39 See Temkin, "Equality, Priority, or What?" 160; and Hirose, "Reconsidering the Value of Equality," 307.

40 See Parfit, "Equality and Priority," 213–14; and Sen, *Collective Choice and Social Welfare*, 138.

> possible population A and all individuals $p_i \in A$, whose welfare levels are represented by the individual welfare function w, it is the case that if the welfare of an individual $p_i \in A$ changes from $w_1(p_i)$ to $w_2(p_i)$, and there are no other welfare changes for the members of A, then $W_1(A, w) - W_2(A, w) = f(w_1(p_i), w_2(p_i))$.

This condition implies that a change affecting only a subgroup of a population affects social welfare independently of the fixed situation of the rest of the population. The *PW* measure satisfies the condition because it is a function only of absolute welfare values, and not of welfare differences. The *EW* measure, being a function both of absolute welfare values and of welfare differences, fails to satisfy the condition.

To distinguish between egalitarianism and prioritarianism on the basis of relationality and separability is quite common. The idea that egalitarianism is relational while prioritarianism is not is held by Parfit, McKerlie, and Hirose.[41] The related idea that egalitarianism is non-separable while prioritarianism is not is held by Broome.[42]

As far as only egalitarianism is concerned, everyone agrees that egalitarianism is a relational theory (including Temkin, McKerlie, Parfit, and Holtug).[43] But not everyone agrees that egalitarianism must use a non-separable measure. Several philosophers think that egalitarianism could very well use a separable measure (including Tungodden, Fleurbaey, Jensen, and McCarthy).[44]

Concerning prioritarianism, opinions are more divided. Fleurbaey believes that prioritarianism should be regarded as a relational theory ("or it should have a different name"), whereas Parfit believes that it should be regarded as a non-relational theory (to which McKerlie and Holtug agree).[45] Persson proposes that prioritarianism could be regarded either as a relational or non-relational theory and makes a distinction between an *absolute priority view* and a *relational priority view* (as presented previously).[46] There are a number of philosophers who

41 See Parfit, "Equality or Priority?" 104; McKerlie, "Understanding Egalitarianism," 53; and Hirose, *Egalitarianism*, 95.

42 See Broome, "Equality versus Priority," 221.

43 See Temkin, "Equality, Priority, or What?" 138; McKerlie, "Equality and Priority," 25; Parfit, "Equality and Priority," 214; and Holtug, *Persons, Interests, and Justice*, 174.

44 See Tungodden, "The Value of Equality," 15; Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 106; McCarthy, "Risk-Free Approaches to the Priority View," 439–40; Fleurbaey, "Equality versus Priority," 215; and Buchak, "Taking Risks behind the Veil of Ignorance," 642.

45 See Fleurbaey, "Equality versus Priority," 206; Parfit, "Equality and Priority," 214; McKerlie, "Understanding Egalitarianism," 53; and Holtug, *Persons, Interests, and Justice*, 204.

46 See Persson, "Equality, Priority and Person-Affecting Value," 35.

claim that prioritarianism should use a separable measure (for example Jensen, Tungodden, Adler, and Broome).[47] However, Buchak disagrees and suggests that prioritarianism could use a non-separable measure.[48]

What should we make of all of this? The definitions given for prioritarianism and egalitarianism above present both theories as intrinsically dependent on relations, at least in the sense that both *faring unequally well* and *being worse faring* are relational properties. However, both theories have also been represented by non-relational and separable measures. Two questions thus arise. First, is there any sense in which egalitarianism or prioritarianism could be regarded as non-relational theories? Second, if not, is it unsuitable to represent either theory by a non-relational (and separable) measure?

Let us look at the first question. As far as egalitarianism is concerned, we cannot regard it as a non-relational theory. Inequality depends on the relation of *worse faring* and is intrinsically relational. Prioritarianism is different. Even though *worse faring* is a relation, we could regard prioritarianism as a non-relational theory, at least in some sense. Rather than interpreting the expression "being worse faring" personally, as referring to the relational property of being worse off than other people, we could interpret the expression impersonally, as referring to the property of being worse off than one would be at higher levels of welfare. Given this interpretation, prioritarianism would state that welfare changes for worse-faring people matter more for social welfare when the worse-faring people are further from some fixed higher level of welfare. This version is still a relational version of prioritarianism, but it is equivalent to a non-relational version, stating that the importance of individual welfare changes to social welfare depends on the absolute values of the levels changed. Thus, it is possible to formulate a version (or at least the equivalence of a version) of prioritarianism that does not refer to the relation of worse faring and thus can be regarded as a non-relational theory.

If prioritarianism could be regarded as either a relational or non-relational theory, the second question is interesting only in relation to egalitarianism: Considering that egalitarianism must be classified as a relational theory, is it possible for egalitarians to use a non-relational measure, such as the *PW* measure? This

47 See Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 106; Tungodden, "Equality and Priority," 423; Adler, *Well-Being and Fair Distribution*, 311; and Broome, "Equality versus Priority," 221.

48 See Tungodden, "The Value of Equality," 15; Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 106; McCarthy, "Risk-Free Approaches to the Priority View," 439–40; Fleurbaey, "Equality versus Priority," 215; and Buchak, "Taking Risks behind the Veil of Ignorance," 642.

question cannot be answered without a theory concerning what a measure of social welfare should do. Here I will consider two proposals.

The first proposal is that the sole function of a measure of social welfare is to mathematically represent quantitative relations between populations in terms of social welfare. In that case, any measure may be used as an egalitarian measure as long as it assigns appropriate values to populations (according to egalitarianism). To show that a non-relational measure, such as the *PW* measure, is inappropriate for egalitarianism requires finding some assignment of values by such a measure that is not egalitarian.

As far as I know, no one has attempted to show that the *PW* measure itself renders rankings that are inappropriate for egalitarianism. However, some have suggested that egalitarians and prioritarians will rank populations differently. A few of these suggestions do not give any concrete examples, and cannot really be assessed.[49] (If I am correct, no such examples can realistically be given.)[50] The only concrete proposal that has been generally discussed is a proposal by Broome. I will give a simplified version of it here.

Let us assume that we are to compare four different populations, *A*, *B*, *C*, and *D*, with the welfare vectors $\mathbf{v}_A = (2, 2, 2, 2)$, $\mathbf{v}_B = (4, 1, 2, 2)$, $\mathbf{v}_C = (2, 2, 1, 1)$, $\mathbf{v}_D = (4, 1, 1, 1)$. According to Broome, prioritarianism implies that *A* is better than *B* if and only if *C* is better than *D*. The reason is that the only difference between *A* and *B* is the well-being of the first two people and this difference is exactly the same difference as that between *C* and *D*. However, an egalitarian might think that *A* is better than *B* because *A* is more equal than *B*, and that *D* is better than *C* because *D* has a higher total sum of individual welfare.[51]

However, it is far from obvious that a prioritarian and an egalitarian would reason in these diverging ways. The assumption that a prioritarian must rank *A* over *B* if and only if *C* is ranked over *D* presupposes that a prioritarian ranking must be separable, and this is questionable (as we have already seen). It is also questionable (and I think incorrect) that only an egalitarian could consider two properties important for social welfare and also that only an egalitarian could consider one of them more important in one case and less important in another, in what seems to be a rather unprincipled way.[52]

49 See Parfit, "Equality or Priority?" 105; McKerlie, "Equality and Priority," 26; and Hausman, "Equality versus Priority," 230.

50 Compare Fleurbaey, "Equality versus Priority," 209.

51 See Broome, "Equality versus Priority," 222–23. For a similar example, see Sen, *On Economic Inequality*, 41.

52 Peterson and Hansson contend that Broome's version of egalitarianism is too unspecific to be assessed ("Equality and Priority," 303).

To properly assess different rankings, one needs to formulate precise conditions that specify when rankings should be regarded as egalitarian or as prioritarian. As I discussed this earlier, and failed to find a difference in rankings, I will not pursue this topic further here.

Let us thus look at the second proposal. According to this proposal, a measure of social welfare should have a function besides correctly representing quantitative relations between populations. It should also reflect the intrinsic dependence relation between social welfare and inequality, or worse faring, in its very form.

The idea that a measure of social welfare should reflect intrinsic dependence relations needs to be specified. One way to understand it is that the measure of social welfare should be a derived measure, that is: a function of other functions that measure the properties on which social welfare intrinsically depends.[53] The *EW* measure, being a function of a measure of total welfare and a measure of equality, shows social welfare as intrinsically dependent on individual welfare and equality. The *PW* measure, being a function only of individual welfare, shows social welfare as intrinsically dependent only on individual welfare.[54]

This understanding seems too crude, however. The *PW* measure is not just a function of individual welfare. It is a function of weighted individual welfare, where lower welfare values have larger weight. Thus, it does not show social welfare as intrinsically a function only of individual welfare. Rather, it shows social welfare as a function of individual welfare and the diminishing marginal importance of individual welfare (or some property like this). If we consider the diminishing marginal importance of individual welfare to be a prioritarian property, then the *PW* measure could at least be suitable for impersonal prioritarianism (although it would not be suitable for personal prioritarianism or egalitarianism).

There is something odd about the second proposal, however. Why should a measure reflect intrinsic dependency relations in its very form? It can hardly be for purely pedagogical reasons. But then the only explanation seems to be that a measure must reflect intrinsic dependency relations in order to accurately measure social welfare. And this leads us back to the first explanation. If the ability of a measure to reflect intrinsic dependence relations determines its ability to measure social welfare, then the *EW* measure could be used both for egalitari-

53 This idea is presented (but not endorsed) by Jensen as an interpretation of an egalitarian idea of Temkin's. See Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 94.

54 Some philosophers have insisted that an egalitarian social welfare measure should not be an additively separable function on individual welfare values—for example Jensen, "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" 108; and Broome, "Equality versus Priority," 221. Both Tungodden ("The Value of Equality," 16) and Fleurbaey ("Equality versus Priority," 215) disagree, however.

anism and the personal version of prioritarianism, while the *PW* measure could at most be used for the impersonal version of prioritarianism. However, there is little reason to think that the ability of a measure to reflect intrinsic dependence relations determines its ability to represent the social welfare of populations since measures cannot distinguish between intrinsic and instrumental dependence relations. What matters for measurement is not intrinsic dependence, but *necessary covariation*. Despite their structural differences, both the *PW* measure and the *EW* measure are able to rank populations according to both egalitarian and prioritarian ideas. The non-relational *PW* measure is sensitive to inequality, which is a relational property, and the relational *EW* measure is affected more by changes to worse-faring people, even if they are impersonally worse faring.

7. Conclusion

In this essay I have discussed whether egalitarianism and prioritarianism must use different social welfare measures. I have argued that they need not, because: (1) conceptual connections between equality and worse faring are such that any egalitarian measure will work as a prioritarian measure as well, and vice versa; (2) two necessary and sufficient conditions for egalitarian and prioritarian measures, respectively, are equivalent; (3) two standard measures for egalitarianism and prioritarianism have been or might be used for either theory; (4) the fact that a measure satisfies Pareto cannot disqualify it as egalitarian; (5) the fact that a measure is level sensitive cannot disqualify it as egalitarian either; and (6) the fact that a measure is non-relational and separable precludes using it for egalitarianism only if a social welfare measure must reflect intrinsic dependence relations in its very form, which is doubtful.

The equivalence of egalitarianism and prioritarianism implies that for practical purposes there is no reason to choose between the two theories. It also implies that for theoretical purposes a choice between the two theories cannot be guided by differences in evaluation.[55]

Stockholm University
karin.enflo@philosophy.su.se

55 For useful discussion of an earlier version of this paper, I am grateful to participants at the higher seminar in practical philosophy at Uppsala University, especially to Emil Andersson, Erik Carlson, Anna Folland, Jens Johansson, Andrew Reisner, and Olle Risberg. For useful comments, I am also grateful to several anonymous referees. For discussion, comments, and encouragement, I am especially grateful to Per Enflo and Victor Moberger.

APPENDIX

Egalitarian Condition: For a measure of social welfare W and for all possible populations A and B and their members $p_i \in A$ and $q_i \in B$, such that $|A| = |B|$ and $\Sigma w(p_i) = \Sigma w(q_i)$, *if* there is a bijection from A to B, such that each individual $p_i \in A$ could be paired with an individual $q_i \in B$ so that for each pair of individuals (p_i, q_i) it is the case that $w(p_i) = w(q_i)$, except for four individuals: p_1, p_2, q_1, q_2, such that $|w(p_1) - w(p_2)| < |w(q_1) - w(q_2)|$, *then A does better than B*, and thus $W(A) > W(B)$.

Prioritarian Condition: For a measure of social welfare W and for any possible population C and for any individuals $r_i, s_i \in C$ such that $w(r_i) < w(s_i)$ and $w(r_i) \geq 0$ and $w(s_i) \geq 0$, *if* it is possible to either increase the welfare of r_i by m, resulting in population C^*, or increase the welfare of s_i by m, resulting in population C^{**}, *then* C^* *does better than* C^{**} and thus $W(C^*) > W(C^{**})$.

A1. *Proof that the Prioritarian Condition Follows from the Egalitarian Condition*

According to the assumptions in the Egalitarian Condition, if there are populations A and B such that $|A| = |B|$ and for their members $p_i \in A$ and $q_i \in B$, it is the case that $\Sigma w(p_i) = \Sigma w(q_i)$ and there is a bijection from A to B, such that each individual $p_i \in A$ could be paired with an individual $q_i \in B$ so that for each pair of individuals (p_i, q_i) it is the case that $w(p_i) = w(q_i)$, except for four individuals: p_1, p_2, q_1, q_2, such that $|w(p_1) - w(p_2)| < |w(q_1) - w(q_2)|$, then it is the case that $W(A) > W(B)$.

Let us assume that there is some population C, which can be transformed into either C^* or C^{**} by raising either the welfare of a worse-faring individual $r_i \in C$ by m or a better-faring individual $s_i \in C$ by the same amount m. We must then prove that if the Egalitarian Condition holds, then $W(C^*) > W(C^{**})$.

Let us put $C^* = A$ and $C^{**} = B$, since $|C^*| = |C^{**}|$ and for their members $p_i \in C^*$ and $q_i \in C^{**}$ it is the case that $\Sigma w(p_i) = \Sigma w(q_i)$ and there is a bijection from C^* to C^{**}, such that each individual $p_i \in C^*$ could be paired with the same or a counterpart individual $q_i \in C^{**}$ so that for each pair of individuals (p_i, q_i) it is the case that $w(p_i) = w(q_i)$, except for the two individuals p_1 and p_2 and their counterparts q_1 and q_2, where $w(p_1) = w(r_i) + m$, $w(p_1) = w(s_i)$, $w(q_1) = w(r_i)$, and $w(q_2) = w(s_i) + m$. For these four individuals it is the case that:

$$|w(p_2) - w(p_1)| = |w(s_i) - w(r_i) - m| \ldots (1)$$

and the case that:

$$|w(q_2) - w(q_1)| = |w(s_i) - w(r_i) + m| \ldots \text{(II)}.$$

We can now apply the elementary inequality:

$$\text{If } a > 0 \text{ and } b > 0, \text{ then } |a - b| < |a + b| = a + b \ldots \text{(III)}.$$

We let $a = w(s_i) - w(r_i)$ and $b = m$. Since (I), (II) and (III) hold, this gives: $|w(p_2) - w(p_1)| = |w(s_i) - w(r_i) - m| = |a - b| < |a + b| = |w(s_i) - w(r_i) + m| = |w(q_2) - w(q_1)|$. Thus: $|w(p_2) - w(p_1)| < |w(q_2) - w(q_1)|$, and if the Egalitarian Condition holds, then $W(A) > W(B)$, which is the same as $W(C^*) > W(C^{**})$.

Q. E. D.

A2. Proof that the Egalitarian Condition Follows from the Prioritarian Condition

According to the assumptions in the Prioritarian Condition, if there is some population C, which can be transformed into either C^* or C^{**} by raising either the welfare of a worse-faring individual $r_i \in C$ by m or a better-faring individual $s_i \in C$ by the same amount m, then $W(C^*) > W(C^{**})$.

Let us assume that there are populations A and B, such that $|A| = |B|$ and for their members $p_i \in A$ and $q_i \in B$, it is the case that $\Sigma w(p_i) = \Sigma w(q_i)$ and there is a bijection from A to B, such that each individual $p_i \in A$ could be paired with an individual $q_i \in B$ so that for each pair of individuals (p_i, q_i) it is the case that $w(p_i) = w(q_i)$, except for four individuals: p_1, p_2, q_1, q_2, such that $|w(p_1) - w(p_2)| < |w(q_1) - w(q_2)|$. We must then prove that if the Prioritarian Condition holds, then $W(A) > W(B)$.

Without loss of generality, we can assume that $w(q_1) < w(p_1) \leq w(p_2) < w(q_2)$. Consider then a population C, where $|C| = |A| = |B|$ and members r_i, such that $w(r_1) = w(q_1)$ and $w(r_2) = w(p_2)$ and for all other i, $w(r_i) = w(s_i) = w(p_i) = w(q_i)$. Since $w(q_1) + w(q_2) = w(p_1) + w(p_2)$, we get $w(q_2) - w(p_2) = w(p_1) - w(q_1)$, which gives:

$$w(q_2) - w(r_2) = w(p_1) - w(r_1) \ldots \text{(I)}$$

We get A from C by increasing $w(r_1)$ to:

$$w(p_1) = w(r_1) + w(p_1) - w(r_1) \ldots \text{(II)}$$

We also get B from C by increasing $w(r_2) = w(p_2)$ to:

$$w(q_2) = w(r_2) + w(q_2) - w(r_2) \ldots \text{(III)}$$

Since $w(r_1) < w(r_2)$ and $(w(p_1) - w(r_1)) = (w(q_2) - w(r_2))$ and (II) and (III) hold, and if the Prioritarian Condition holds, then $W(A) > W(B)$.

Q. E. D.

REFERENCES

Adler, Matthew. *Well-Being and Fair Distribution*. Oxford: Oxford University Press, 2012.

Arneson, Richard J. "Luck Egalitarianism and Prioritarianism." *Ethics* 110, no. 2 (January 2000): 339–49.

Arrhenius, Gustaf. "Egalitarian Concerns and Population Change." In *Inequalities in Health: Concepts, Measures, and Ethics*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 74–91. Oxford: Oxford University Press, 2013.

Blackorby, Charles, Walter Bossert, and David Donaldson. "The Axiomatic Approach to Population Ethics." *Politics, Philosophy and Economics* 2, no. 3 (October 2003): 342–81.

Bosmans, Kristof. "Extreme Inequality Aversion without Separability." *Economic Theory* 32, no. 3 (September 2007): 589–94.

Broome, John. "Equality versus Priority: A Useful Distinction." *Economics and Philosophy* 31, no. 2 (July 2015): 219–28.

———. *Weighing Goods: Equality, Uncertainty, and Time*. Oxford: Basil Blackwell, 1991.

Brown, Campbell. "Prioritarianism for Variable Populations." *Philosophical Studies* 134, no. 3 (June 2007): 325–61,

Buchak, Lara. "Taking Risks behind the Veil of Ignorance." *Ethics* 127, no. 3 (April 2017): 610–44.

Christiano, Thomas, and Will Braynen. "Inequality, Injustice, and Levelling Down." *Ratio* 11, no. 4 (December 2008): 393–420.

Crisp, Roger. "Equality, Priority, and Compassion." *Ethics* 113, no. 4 (July 2003): 745–63.

Dalton, Hugh. "The Measurement of the Inequality of Incomes." *Economic Journal* 30, no. 119 (September 1920): 348–61.

Deschamps, Robert, and Louis Gevers. "Leximin and Utilitarian Rules: A Joint Characterization." *Journal of Economic Theory* 17, no. 2 (April 1978): 143–63.

Ebert, Udo. "Rawls and Bentham Reconciled." *Theory and Decision* 24, no. 3 (May 1988): 215–23.

Esposito, Lucio, and Peter J. Lambert. "Poverty Measurement: Prioritarianism, Sufficiency and the 'I's of Poverty.'" *Economics and Philosophy* 27, no. 2 (July 2011): 109–21.

Fleurbaey, Marc. "Equality versus Priority: How Relevant Is the Distinction?" *Economics and Philosophy* 31, no. 2 (July 2015): 203–17.

Gini, Corrado. "Variabilità e mutabilità." In *Studi economico-giuridici della Facoltà di Giurisprudenza*. Bologna: Università di Cagliari, 1912.

Hammond, Peter J. "Equity, Arrow's Conditions, and Rawls' Difference Principle." *Econometrica* 44, no. 4 (July 1976): 793–804.

———. "A Note on Extreme Inequality Aversion." *Journal of Economic Theory* 11, no. 3 (December 1975): 465–67.

Hausman, Daniel. "Equality versus Priority: A Misleading Distinction." *Economics and Philosophy* 31, no. 2 (July 2015): 229–38.

Hirose, Iwao. *Egalitarianism*. London: Routledge, 2015.

———. "Reconsidering the Value of Equality." *Australiasian Journal of Philosophy* 87, no. 2 (June 2009): 301–12.

Holtug, Nils. *Persons, Interests, and Justice*. Oxford: Oxford University Press, 2010.

Jensen, Karsten Klint. "What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?" *Economics and Philosophy* 19, no. 1 (April 2003): 89–109.

McCarthy, David. "Distributive Equality." *Mind* 124, no. 496 (October 2015): 1045–1109.

———. "Risk-Free Approaches to the Priority View." *Erkenntnis* 78, no. 2 (April 2013): 421–49.

McKerlie, Dennis. "Equality and Priority." *Utilitas* 6, no. 1 (May 1994): 25–42.

———. "Understanding Egalitarianism." *Economics and Philosophy* 19, no. 1 (April 2003): 45–60.

Nagel, Thomas. *Equality and Partiality*. Oxford: Oxford University Press, 1991.

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

Pareto, Vilfredo. *Manuel d'Economie Politique*. Paris: Giard et. E., 1909.

Parfit, Derek. "Equality and Priority." *Ratio* 10, no. 3 (December 1997): 202–21.

———. "Equality or Priority?" In *The Ideal of Equality*, edited by Matthew Clayton and Andrew Williams, 81–125. Basingstoke: Macmillan, 2000.

Persson, Ingmar. "Equality, Priority and Person-Affecting Value." *Ethical Theory and Moral Practice* 4, no. 1 (March 2001): 23–39.

Peterson, Martin, and Sven Ove Hansson. "Equality and Priority." *Utilitas* 17, no. 3 (November 2005): 299–309.

Pietra, Gaetano. "Delle Relazioni tra gli Indici di Variabilità, Note I e II." *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti* 74, 2 (1915): 775–804.

Pigou, Arthur Cecil. *Wealth and Welfare*. London: Macmillan, 1912.

Rabinowicz, Wlodek. "Prioritarianism for Prospects." *Utilitas* 14, no. 1 (March 2002): 2–21.

———. "The Size of Inequality and Its Badness: Some Reflections around Temkin's *Inequality*. *Theoria* 69, nos. 1–2 (August 2003): 60–84.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
Scheffler, Samuel. *The Rejection of Consequentialism*. Oxford: Oxford University Press, 1982.
Sen, Amartya. *Collective Choice and Social Welfare*. Cambridge, MA: Harvard University Press, 1970.
———. *On Economic Inequality*. Oxford: Clarendon Press, 1973.
Sen, Amartya, and James Foster. *On Economic Inequality*. Oxford: Clarendon Press, 1997.
Sidgwick, Henry. *The Methods of Ethics*. London: Macmillian, 1874.
Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 1973.
Temkin, Larry. "Equality, Priority, or What?" *Economics and Philosophy* 19, no. 1 (April 2003): 61–87.
———. *Inequality*. Oxford: Oxford University Press, 1993.
Theil, Henri. *Economics and Information Theory*. Amsterdam: North Holland, 1967.
Tungodden, Bertil. "Equality and Priority." In *The Handbook of Rational and Social Choice*, edited by Paul Anand, Prasanta Pattanaik, and Clemens Puppe, 411–32. Oxford: Oxford University Press, 2009.
———. "The Value of Equality." *Economics and Philosophy* 19, no. 1 (April 2003): 1–44.
Vallentyne, Peter. "Equality, Efficiency and the Priority of the Worse-Off." *Economics and Philosophy* 16, no. 1 (April 2000): 1–19.
Weirich, Paul. "Utility Tempered with Equality." *Noûs* 17, no. 3 (September 1983): 423–39.

Journal of Ethics and Social Philosophy
VOL. 21, NO. 1 · MAY 2022
https://doi.org/10.26556/jesp.v22i1.1518

THE LIMITS OF INSTRUMENTAL PROCEDURALISM

Jake Monaghan

According to proceduralism in political philosophy, political power is just, legitimate, or authoritative when it is the output of an appropriate procedure.[1] Even if one disagrees with the output of an appropriate political procedure, we must recognize its legitimacy because we endorse, or there is good reason to endorse, the procedure that generated it.[2] On this view, normative properties are transmitted from the procedure to its output. In this

1 I follow Buchanan's use of the term "legitimacy" in this paper, according to which legitimacy refers to the permission to exercise political power ("Political Legitimacy and Democracy," 689). I will also assume that individual political actions can be legitimate or illegitimate independent of a regime's legitimacy. On this usage, legitimacy does not imply duties of obedience. This is importantly different from other uses of the term. Simmons, for instance, distinguishes legitimacy from justification and holds a voluntaristic conception of legitimacy that implies duties to obey ("Justification and Legitimacy," 769). Justification, as Simmons uses the term, is similar to the conception of legitimacy employed by Buchanan. I follow Buchanan rather than Simmons simply because most proponents of the kind of proceduralism I focus on here have not taken on board Simmons's more fine-grained typology of political evaluation. As we will see, proponents of a view like Simmons's, according to which a state or procedure being reliably good or correct is morally independent of its being legitimate (since that requires a special historical relationship between states and subjects), will reject the family of proceduralist views I focus on here. Proceduralists intentionally collapse the distinction Simmons wants to make by inferring legitimacy from the goodness and reliability of political procedures. I will not attempt to adjudicate this dispute, and the modifications to proceduralism I defend are offered as an internal critique of the view.

2 The following philosophers are prominent proponents of proceduralism: Rawls, in particular his discussion of imperfect proceduralism (*A Theory of Justice*, 75), and his discussion of majority rule (*A Theory of Justice*, 313). Estlund's jury/democracy analogy makes clear the nature of his fallibilist proceduralism (*Democratic Authority*, 110). Finally, see Christiano, who distinguishes his view from pure proceduralist and instrumental justifications of democracy and endorses a dualist view that incorporates both elements ("The Authority of Democracy"). Others are concerned with the reliability of our formal decision-making procedures, but do not characterize their views in explicitly instrumental proceduralist terms. Guerrero ("Against Elections"), for example, motivates lottocratic alternatives to democracy on the basis of clear electoral pathologies.

paper I focus on the property of legitimacy, though I will be most concerned with the more general proceduralist form of justification and its relationship to erroneous outputs.

One way of distinguishing different kinds of justifications of the legitimacy of political decisions is what David Estlund has called "correctness" theories and fallibilist theories.[3] According to the former, a political decision is legitimate and authoritative if it "gets things right" as determined by a comprehensive moral theory. The correctness of a political outcome is sufficient for its legitimacy. We find perhaps the most striking example of this view in Plato's *Republic*, where the insights of the philosopher-king license all sorts of behavior, including taking children from their parents so that they can be raised communally.[4] Proceduralist approaches to legitimacy are often, though not always, fallibilist: they allow that so long as the procedure meets certain appropriateness conditions the outcome is legitimate even if it is substantively unjust or incorrect.[5]

My goal here is to articulate a richer account of the fallibilist, proceduralist justification of normative properties relevant to political institutions and their results, especially the legitimacy of political and legal decisions.[6] In particular, I focus on what I call "instrumental proceduralism," which takes one of the appropriateness conditions of a procedure to be that it has a tendency to produce the right result.[7] This is in contrast to both pure proceduralism and correctness justifications of political outcomes, neither of which are fallibilist. Moreover, I am concerned with the actual political procedures that constitute our political and

3 Estlund, *Democratic Authority*, 8, 57. Pure proceduralism and Estlund's own epistemic proceduralism are contrasted with "dogmatic" correctness theories. Correctness theories are dogmatic because they ignore reasonable moral pluralism.

4 At least if we understand talk of sharing children "in common" to mean that children will be raised communally. See *Republic*, 423e6–24a2.

5 Pure proceduralist accounts are not fallibilist. See, for example, Peter, "Pure Epistemic Proceduralism" and *Democratic Legitimacy*. Pure proceduralists reject an external criterion of correctness with which we could evaluate the outcome of a procedure. So, pure and instrumental proceduralism are not dogmatic theories (assuming they respect moral pluralism), but only instrumental proceduralism is fallibilist.

6 Particular versions of proceduralism can set out to justify different normative properties. The details will depend on the conception of the property in question. The structure of justification is what I am interested in here, so it is again worth emphasizing that my discussion shall be concerned with proceduralism in general, using legitimacy typically as an illustration.

7 I owe this term to David Estlund in personal correspondence. Proponents of markets on the grounds that they allocate resources in an appropriate way, for example, also count as instrumental proceduralists.

legal institutions rather than decision procedures aimed at constructing theories or principles of justice.

The strategy is to examine the structural elements of how procedures are thought to confer normative properties on their outputs and then apply these lessons to particular legal and political procedures. I defend three appropriateness conditions for a procedure to confer legitimacy on particular outputs. Procedures must be *highly* reliable, outputs must not have been the result of *predictable* procedural failures, and the failures of a procedure must be *relatively uniformly distributed* in the population.

I begin by characterizing the type of proceduralism I am interested in (section 1). I then argue for a more demanding reliability requirement and introduce and defend the notion of "predictable failure" (section 2). This sets the stage for my argument that barely reliable procedures, predictable procedural failures, and unevenly distributed procedural failures undermine a criminal justice or democratic procedure's ability to confer legitimacy on its outcomes (sections 3 and 4). I conclude by describing how various procedural failures interact with one another and background structural injustices to give a sense of the scope of the problem (section 5). It is not, I suggest, a small or insignificant one, further motivating the requirements set out in what follows.

1. Instrumental Proceduralism as a Kind of Normative Proceduralism

I shall distinguish two forms of proceduralism: *political proceduralism* and *doxastic proceduralism*. Political proceduralism is a general theory of how political decisions earn certain normative properties. Doxastic proceduralism is a general theory of how beliefs earn certain normative properties. Despite their differences, they are normative at bottom: they are concerned with normative properties like legitimacy and authority, justification and knowledge. We can think of these types of proceduralism as versions of *normative proceduralism*. Thinking of the instrumental proceduralism that is endorsed by many contemporary political philosophers as a type of normative proceduralism shall make perspicuous some requirements for the procedure to justify its outcomes.

Political procedures include not only procedures for constructing principles of justice, but democratic decision-making and legal procedures as well. They can be distinguished using the familiar Rawlsian classification of procedural justice. Pure procedures cannot fail to achieve the correct outcome because there are no external, independently specific criteria for success. Perfect and imperfect procedures, both kinds of instrumental procedures, are evaluable in terms of ex-

ternal, independent success criteria. As the names suggest, perfect procedures never fail, whereas imperfect procedures sometimes do.[8]

Doxastic proceduralism is concerned with the extent to which our beliefs track the truth. There is an independently specified standard: reality. Doxastic procedures are evaluated according to how successful they are in generating beliefs that correspond to reality.[9]

Take, for example, Feldman's bird-watcher case.[10] When a bird lands in front of expert and novice bird-watchers, and both form a correct belief about what type of bird it is, only the expert's belief is justified. The reason is twofold: the process that results in the expert's belief is suited to "get it right," and tends to get it right. Our beliefs are not justified when they are the result of wishful thinking, bad reasoning, are luckily true, and the like; they are only justified when they are the output of a procedure that tracks the truth.[11]

The appropriateness conditions for political proceduralism tend to be founded on concerns about public justification. Because our political institutions coerce people, we must be able to justify institutions and their power to those reasonable individuals who are coerced by them.[12] To do otherwise is to disregard one's status as a moral person. Another closely related concern implies that our institutions must satisfy an equality requirement: they must aim to advance our interests equally.[13] These constraints are distinct, but they all aim at justifying

8 The original position and freely consented to gambles are pure procedures; Rawls, *A Theory of Justice*, 74–75, and "Kantian Constructivism in Moral Theory," 523.

9 Nearly everyone agrees that two of the requirements for knowing *p* are that *p* is true and that *S* believes *p*. Gettier cases show that knowledge requires some sort of anti-luck requirement ("Is Justified True Belief Knowledge?"). If you believe *p*, but you just happen to luckily believe something true, it is unlikely that you know *p*. To explain this, some epistemologists have appealed to proceduralism for part of their analysis of knowledge and justification. These epistemologists, *process reliabilists*, claim in some form or another that our beliefs are only justified (and candidates for knowledge) when they are the output of a suitable process or procedure; see Feldman, *Epistemology* and "Reliability and Justification"; Nozick, *Philosophical Explanations*; Dretske, "Conclusive Reasons" and *Knowledge and the Flow of Information*; Goldman, *Epistemology and Cognition* and *Reliabilism and Contemporary Epistemology*.

10 Feldman, *Epistemology*.

11 There are proceduralist accounts of both justification and knowledge, and each has different requirements. Understanding process reliabilist accounts of knowledge and justification as types of normative proceduralism will be useful for understanding the appropriateness conditions for a normative procedure later on.

12 Vallier, "Against Public Reason Liberalism's Accessibility Requirement"; Larmore, "The Moral Basis of Political Liberalism," 607; Estlund, *Democratic Authority*, 40.

13 Christiano argues that only democratic institutions can satisfy this requirement ("The Authority of Democracy").

institutions to those living within them while avoiding the difficulties associated with evaluating the correctness of individual political decisions.

A key feature of instrumental proceduralism is its fallibilism.[14] Rawls thinks some unjust outcomes may be enforced and must be obeyed.[15] Christiano agrees, and argues that democratic institutions must be evaluated holistically, considering both pure procedural and instrumental evaluations.[16]

To defend fallibilism, Estlund draws an analogy between democratic political procedures and the decision of a jury:

> Recall the jury context: the legitimacy and authority of the verdict are not canceled just whenever the jury is mistaken. If they were, then jailers and police officers ought not to carry out the court's judgment, but should rely on their own judgment of the defendant's guilt or innocence. That conclusion would be the striking and heterodox one.[17]

Correctness theories of legitimacy and authority yield this heterodox implication. Instrumental proceduralism explains how just, or correct, outcomes are legitimate and authoritative, but this is not distinctive of the view. Correctness theories do this as well. Only instrumental proceduralism, in its various forms, can avoid the heterodox implication.

If procedures are good enough, they will tend to produce the right results. And if this obtains, all the results will be legitimate. Note that the outputs have the relevant normative properties *not because we maximize good consequences by going along with them*, but rather because of facts about the procedures themselves.[18] Consequentialist considerations may recommend obeying the output of an ineffective or unreliable procedure because that procedure is simply the best we have. This is not the kind of procedural justification that I am concerned with here.

Here is a more precise way to describe this feature: procedures *transmit* properties to their outcomes. Rawls attributes this feature to pure procedures:

> The fairness of the circumstances under which agreement is reached transfers to the principles of justice agreed to; since the original position

14 Estlund, *Democratic Authority*, 8; Rawls, *A Theory of Justice*, 371.

15 Rawls, *A Theory of Justice*, 308.

16 Christiano, "The Authority of Democracy," 280.

17 Estlund, *Democratic Authority*, 110.

18 Estlund, *Democratic Authority*, 164; Christiano, "The Authority of Democracy," 268.

> situates free and equal moral persons fairly with respect to one another, any conception of justice they adopt is likewise fair.[19]

Something similar might hold for imperfect procedures where the procedure is appropriately formed such that it confers legitimacy on all of its results. Though Rawls uses the term "transfer," I shall use the term "transmit" because procedures sometimes generate a new property rather than transferring an existing one.[20] Instrumental proceduralism, as I understand it, thus relies on the Transmission Thesis:

> *Transmission Thesis* (TT): A procedure *P* with properties *q* will transmit normative property *n* to its outputs *O*.

TT applies to both political and doxastic instrumental procedures. But not just any procedure transmits properties. A bribed judge's decision is not legitimate. An unreliable doxastic procedure does not transmit justification. Instrumental proceduralists must offer an account of which properties make up *q*. If *q*, however we ultimately understand it, is not met, then the transmission of properties fails.

Some philosophers take *q* to be made up of entirely instrumental concerns.[21] Others disagree, as discussed above. But for instrumental proceduralists, *q* must include some instrumental requirements. What are they?

In discussing one type of political procedure, majority-rule voting, Rawls takes its justification to depend on it being the "most feasible way to realize certain ends antecedently defined by the principles of justice."[22] Rawls is not explicit on how reliable a procedure must be in order for it to successfully transmit normative properties, but his comparison to ideal political procedure indicates a concern for reliability and an allowance for some fallibility. If a political procedure always yielded results quite different from what we imagine the ideal procedure would result, we are entitled to think that a particular result is unjust.[23] Christiano's holistic evaluation of democratic procedures explicitly takes on board an instrumental element, thus indicating a concern for reliability. Only Estlund offers a specific standard of reliability: he requires only that a political

19 Rawls, "Kantian Constructivism in Moral Theory," 522.

20 Thanks to David Estlund for this terminology as well.

21 Arneson, "Defending the Purely Instrumental Account of Democratic Legitimacy"; Brennan, *Against Democracy*.

22 Rawls, *A Theory of Justice*, 318.

23 Rawls, *A Theory of Justice*, 314–15.

procedure is "better than random."[24] If the procedure is generally reliable, then the failures of the procedure are "honest mistakes," and honest mistakes do not undermine legitimacy and authority.[25] In the remainder of the paper I argue that these instrumental requirements are insufficient.

2. DOXASTIC PROCEDURES

First, let us consider an analogy between doxastic and political proceduralism. If they are both forms of normative proceduralism, then presumably their failure conditions have similarities. I will argue that, since doxastic procedures fail to transmit normative properties in cases of barely reliable procedures and in circumstances of predictable failure, we should take this to be true of political procedures as well.

2.1. Barely Reliable Procedures

Estlund's Epistemic Proceduralism has only one instrumental appropriateness condition: the procedure must get the right result more than 50 percent of the time or perform better than chance. In contrast, doxastic proceduralists (process reliabilists) usually take the bar to be much higher for knowledge. And though justification might be conferred by a barely reliable procedure, the belief that is the result of such a process will similarly be barely justified.

Consider a scenario in which a barista is trying to determine whether the customers in line want a cappuccino or a latte. Unbeknownst to him, every single customer in the line would like a latte. He decides to employ the following procedure to reach his belief. He will select a customer and flip a coin. If it comes up heads he will believe the customer to want a cappuccino, and if tails he will believe the customer to want a latte. Also unbeknownst to him, the coin is weighted such that 50.01 percent of the time it will land on tails. In this scenario, the barista will likely form the correct belief more than half the time. The process reliabilist does not require that one understand that or why their belief is justified or constitutes knowledge for it to be justified or knowledge. So if the requirement were merely better than random, the barista would be justified, if at all, to a small degree in believing in accordance with the coin flip.[26]

The structural similarities between doxastic and political proceduralism al-

24 Estlund, *Democratic Authority*, 116.

25 Estlund, "On Following Orders in an Unjust War," 221.

26 Reliabilism is a form of externalism. See Huemer, "Phenomenal Conservatism and the Internalist Intuition" for one helpful discussion of the difference between internalists and externalists on this matter.

low us to infer something about the latter from this. Doxastic procedures confer either epistemic justification or knowledge. Legitimacy will be like one of these in the sense that it either comes in degrees or is a threshold concept. If legitimacy is like epistemic justification, then barely reliable political procedures will transmit barely any legitimacy. This would mean that even weak countervailing reasons could override the legitimacy of the political decision. Certainly not everyone understands legitimacy to come in degrees, though this strikes me as a natural thought.[27]

For those who take legitimacy to be a threshold concept, they ought to accept an analogy between legitimacy and knowledge. If like knowledge, then barely reliable procedures will confer no legitimacy; some significantly higher level of reliability will be needed for that. Either option requires that the political proceduralist not settle for simply better than random reliability. Given the stakes of political decisions—they can cost lives rather than coffee preferences—we need something more than merely better than random. I return to this point below.

Rejecting this argument would require one to explain why structurally similar forms of justification have different reliability requirements. One might think that, since we need political procedures for our social coordination, the reliability requirements for legitimacy are less demanding. But this would be to abandon the distinctively proceduralist form of justification formalized in TT in favor of a consequentialist justification.

2.2. *Predictable Failures*

Procedures can be deficient not only in terms of general reliability, but also in

27 And some proceduralists do endorse this way of thinking about authority, at least. Estlund writes, "Epistemic proceduralism generates *more* legitimacy and authority with less demanding epistemic claims" (*Democratic Authority*, 106; emphasis added). On the other hand, some have objected to me that the notion of "degrees of authority" or legitimacy just does not make sense. It seems to me, however, that other popular approaches to authority at least help us make sense of this notion, even if philosophers rarely speak of degrees of authority. For instance, since reasons come in degrees of strength, any view of authority in which it is a power to give reasons will be in principle compatible with degrees of authority. This is because one might have the power to give only weak reasons, whereas another has the power to give very strong reasons. Enoch defends a reasons-giving account of authority, though does not explicitly endorse the view that authority comes in degrees like Estlund does ("Authority and Reason-Giving"). Of course, for those (like Simmons, discussed in note 1) who take legitimacy to be voluntarist, the externalist reasons that are produced by reliable procedures will be immaterial to legitimacy. A view like this may have a harder time accommodating, or may even be incompatible with, a notion of "degrees of authority." Proponents of a voluntarist conception of legitimacy will have parted ways with the instrumental proceduralists much earlier in the dialectic, so I think this incompatibility is not a serious problem. Thanks to a referee for comments on this point.

terms of how reliable they are in certain circumstances. Not all failures are equal. Some justified beliefs will be false, and some unjust political outcomes will be legitimate. That, after all, is one of the goals of political proceduralism. But some failures are *predictable*, either because the design is ill suited to a particular application or because the procedure's input is inappropriate. When these occur, failure is to be expected. Predictable failures, I will argue, undermine the transmission of properties to the output of a procedure.

This should already be familiar enough. The bribed judge example cited above is an instance of a predictable procedural failure. And even if a procedure is well designed, it needs to be "fed" the appropriate material. Meteorological models require accurate data as an input; no matter how well designed the model-construction procedure is, it will not work if it is not fed accurate data. In part because certain sources of data (e.g., buoys in the ocean) often fail to collect accurate data, and because computer models have known weaknesses, the National Hurricane Center employs forecasters instead of issuing guidance based on computer modeling alone. The political proceduralist is already in position to accept this revision. According to Rawls, we "may think of the political process as a machine which makes social decisions when the views of representatives and their constituents are fed into it."[28]

I want to highlight a more subtle kind of predictable failure. Take as our example one possible procedure for acquiring justified beliefs and knowledge about geography. Suppose you have a desire to have these sorts of beliefs. In particular, you are curious about the relative sizes of Germany and Belgium. To satisfy your desire, you consult a world map of the common Mercator projection variety. As you look at the map, it is clear that Germany is much larger than Belgium, for Germany takes up significantly more space on the map than does Belgium.

Because a major cartography company made your map, and because this map and ones very much like it have successfully guided navigation for some time, you are justified in believing that Germany is indeed much larger than Belgium. Your belief is the result of a reliable process, and therefore the procedure transmits justification to its output. And because in this instance the procedure is highly reliable, and your belief is true, it also counts as knowledge.

Suppose, however, you were curious about the relative sizes of Greenland and Africa. You consult the same map, and you form the belief that Greenland is comparable in size to Africa. If you were familiar with the way in which the Mercator projection distorts landmasses far from the equator, you would know that the process used was far from reliable *in this circumstance*, and you would refrain

28 Rawls, *A Theory of Justice*, 171–72.

from forming the belief. But even if you did not know this, and you did form this belief, you would have no or very little justification.

This illustration shows that a procedure can be reliable in some respects but fail predictably in others. We can ignore the specific features of the Mercator projection that make it useful and reliable in some respects and not in others here, and note simply that it does not allow for reliable comparisons of the size of two areas, one of which is close to the equator and one that is not. We should not, however, throw our Mercator projections away. Rather, we should recognize that, as a tool, it is suited for some purposes and not for others, and confers justification or knowledge on some beliefs but not others.[29]

Doxastic procedures can be highly reliable in most circumstances but fail to confer justification or knowledge in circumstances of *predictable* failure. This, it seems, is a feature of normative proceduralism in general. Highly reliable political procedures too can fail to transmit normative properties in circumstances of predictable failure. Indeed, some argue that electoral mechanisms predictably fail to produce representative government.[30] On these grounds, one might reject an election's ability to legitimate its results.

3. CRIMINAL PROCEDURES

Two lessons emerged from the last section: the appropriateness conditions of TT for instrumental normative procedures include high reliability and an anti-predictable failure requirement. We will see that this yields plausible results in the context of criminal justice procedures. An analysis of these procedures demonstrates that there is one more element we must attend to: the distribution of failure.

3.1. Guilt by Coin Flip

Instrumental proceduralism attempts to explain why the outputs of criminal justice procedures are legitimate. As Estlund notes, corrections officers should not help prisoners escape even if they suspect that they were wrongly convicted, and they do not act wrongly in detaining them.

29 These cases are not rare. Consider the procedure one might use to acquire reliable beliefs about which way north is. A compass is part of a reliable procedure. But it fails predictably in certain circumstances—in close proximity to magnets, for instance. For more high-powered doxastic procedures, one could look to the various kinds of models that go into hurricane forecasts. Some are better suited to predicting wind strength, others for storm tracking. Forecasters avoid procedural failure by using them as part of distinct doxastic procedures.

30 Guerrero, "Against Elections."

But consider a barely reliable criminal trial procedure: a flip of a coin determines guilt. If the coin comes up heads, the defendant is found guilty. If the coin comes up tails, the verdict is innocent. The coin is weighted and the composition of the pool of defendants is such that 50.01 percent of the time it gets the right result. This procedure succeeds at a rate better than chance. I suspect that not only would no one take the verdicts to be legitimate or authoritative, people would find this procedure deeply unjust. People were rightly outraged at footage showing police officers use a coin flip to determine whether to arrest a speeding motorist.[31] Like doxastic procedures, then, trial procedures must be highly reliable.

Actual trial procedures are not coin flips. Yet they are quite a bit like the doxastic procedure involving maps discussed above in that they can fail in predictable ways. I turn to that now.

3.2. *Hungry Judges and Biased Juries*

Social scientists have investigated whether, and what sort of, irrelevant factors influence sentencing in criminal cases. This body of scholarship provides us with empirical evidence of predictable failures of trial procedures. These procedures are aimed at achieving justice by presenting evidence and arguments to judges and juries. These individuals are supposed to be as close to the "ideal observer" as possible. That is, they should not have a bias in favor of guilt, or in favor of one party or another, and they should be competent at evaluating evidence and arguments. Indeed, these individuals are professionally trained to approximate the ideal, and that they come close to achieving the ideal is a large part of the foundation of the institution's legitimacy.

There are many ways in which one can deviate from the ideal. Legal realists sometimes disparagingly say that "justice is what the judge ate for breakfast."[32] Clearly what a judge ate for breakfast is irrelevant when it comes to what the outcome of a procedure *should* be, and if hunger leads to bad decisions, then this is nonideal. Although there may be no way to determine the exact correct sentencing, if it turns out that judges give harsher sentences when they are further from their most recent meal, then presumably this is a result of a deviation from the ideal state. This notion underlies the procedure of "comparative sentence review" or "proportionality review" that some courts undergo to determine whether a sentence is appropriate. Many states legally require the state supreme court to perform these reviews in cases where defendants receive the death pen-

31 Amiri, Sacks, and Sanders, "Georgia Officers on Leave after Coin-Toss App Used before Decision to Make Arrest."

32 Danziger, Levav, and Avnaim-Pesso, "Extraneous Factors in Judicial Decisions."

alty. Some empirical research supports the realist's concern: "the likelihood of a favorable ruling is greater at the very beginning of the workday or after a food break than later in the sequence of cases."[33] Though there may be good reason to be skeptical about this empirical claim, it demonstrates a clear possible example of predictable failure in criminal justice procedures.

We might also expect judges and juries to be influenced by various psychological biases when reaching their decisions. A study of over seventy-seven thousand sentences found that while white offenders had an average sentence of thirty-two months, Black offenders had an average sentence of sixty-four months. Further, the study concluded that ethnicity accounted for over half of the variance.[34] Another study had different results, finding that seriousness of offense accounted for the majority of variations in sentence length. What they did find, however, was that Black defendants with more "Afrocentric" facial features received longer sentences than Black defendants with less "Afrocentric" facial features.[35] Other studies produce broadly similar results for sentencing and bail setting.[36] When a murder victim is white, in death penalty jurisdictions, the defendant is more likely to receive the death penalty.[37] This is even more likely when the defendant is Black.[38] Empirical research provides evidence that racial biases play a role in this: people tend to seek retribution more strongly when the victim is white.[39]

It is possible that there are explanations of these data that do not rely upon race. Social psychologist Neil Vidmar presents as possible alternatives the defendant's demeanor, manner of speaking, reports prepared by probation officers, or bail conditions set by magistrates.[40] These possibilities just push the problem of

33 Danziger, Levav, and Avnaim-Pesso, "Extraneous Factors in Judicial Decisions," 6890.

34 Mustard, "Racial, Ethnic, and Gender Disparities in Sentencing."

35 Blair, Judd, and Chapleau, "The Influence of Afrocentric Facial Features in Criminal Sentencing"; Vidmar, "The Psychology of Trial Judging."

36 Rachlinski et al., "Does Unconscious Racial Bias Affect Trial Judges?"; Ayres and Waldfogel, "A Market Test for Race Discrimination in Bail Setting"; Monaghan, Van Holm, and Surprenant, "Get Jailed, Jump Bail?"

37 Baldus, Pulaski, and Woodworth, "Comparative Review of Death Sentences."

38 Baldus et al., "Evidence of Racial Discrimination in the Use of the Death Penalty"; Eberhardt et al., "Looking Deathworthy." Though Pierce and Radelet find that the race of the defendant is not a predictor of receiving the death penalty in Baton Rouge, Louisiana, the race of the victim is (Pierce and Radelet, "Death Sentencing in East Baton Rouge Parish, 1990–2008"; Radelet and Pierce, "Race and Death Sentencing in North Carolina, 1980–2007"). Here their results are in agreement with Baldus et al.

39 Levinson, Smith, and Young, "Devaluing Death."

40 Vidmar, "The Psychology of Trial Judging," 59.

bias back a step. Further, none of these alternatives serve as justification of the sentencing disparities. It remains the case that Black defendants tend to receive longer sentences than white defendants for comparable offenses. Unless we can point to reasons other than the defendant's race to explain this disparity, this is another predictable procedural failure (predictable in light of our knowledge of the serious problems of racism).

Predictable failures can arise in a variety of ways. In adversarial trial systems, the hope is that by having both parties battle it out the truth will prevail. But even if all the other components of the procedure are functioning ideally, if one legal team is less skilled than the other, it becomes possible for the truth not to prevail. Perhaps the skill differential needs to be substantial before the procedure will fail. This remains, however, a clear possible input problem for adversarial legal procedures.[41]

The courts themselves recognize something like this. Defendants have a right to competent legal defense, and individuals can argue before judges that their conviction was the result of incompetent legal representation. If the defendant can show that the legal advice or defense they received was indeed incompetent, and that this caused the conviction, the conviction can be overturned. Because our actual legal procedures recognize that what I call predictable failures undermine the legitimacy of their decisions, and for this reason have ways of rectifying the failure, proceduralist justifications of political legitimacy must recognize this as a constraint on justification as well. This is especially true for proceduralists who appeal to legal procedures as part of their case for a fallibilist theory of legitimacy and authority.

Some might object at this point that the examples described above are not instances of the same procedure. Perhaps there are really two separate judicial procedures, one for white defendants and one for Black defendants. And since only the former procedure meets the reliability bar we do not need the notion of predictable failure to explain why even the courts have recognized that their decisions are illegitimate and non-authoritative in certain cases. The problem with the suggestion is not only that it is ad hoc, but it also does not capture the way that proceduralists typically think about procedures. It threatens to make every political or legal decision the singular output of a one-time procedure. Each relevant difference in the reliability of a procedure would generate a new procedure, and this gives each procedure a perfect reliability or a perfect anti-reliability. But then we would be back to a correctness theory of legitimacy.

41 In fact, empirical research shows that this is a problem. See Frederique, Joseph, and Hild, "What Is the State of Empirical Research on Indigent Defense Nationwide?"; Abrams and Yoon, "The Luck of the Draw."

3.3. *The Distribution of Failure*

There is a long history of a presumption in favor of innocence. To paraphrase a famous remark from William Blackstone (and many others), it is better that ten guilty escape than one innocent suffers.[42] This commitment to lenience does not, by itself, tell us how often a trial system can yield a false verdict of guilty before losing legitimacy, but the numbers suggest that it must be *considerably* more reliable than better than random. Because this presumption of innocence is widely endorsed, most should think that the "better than random" standard is not sufficient for a legal procedure to transmit normative properties to their outputs. We cannot look simply at the failure *rate*. We must also consider the failure *distribution*.

The distribution of failure is especially important for political procedures. Legal and political decisions can be seriously harmful, and in exercising political power they threaten to undermine political equality. For this reason, they must be justifiable to whom they affect. But the predictable failures discussed above highlight an important requirement for the success of proceduralist justifications of legal and political decisions. Some procedures can be highly reliable, meeting the first requirement. They might fail in predictable ways only on rare occasions. But, *for whom* they fail is significant, and there are relevant groups beyond the innocent. If a minority group in a political community experiences the vast majority of the procedural failures, then the procedure cannot be justified to them. They are within their rights to ask, "Why should I obey the output of this procedure? It clearly does not work *for us*." They can rightly deny that these failures are honest mistakes, even if the failure does not stem from intentional malfeasance.

Let us suppose that the criminal legal system is 90 percent reliable, thereby meeting the general reliability criterion. If the 10 percent failure rate falls entirely or mostly on certain portions of the population, we have a situation in which the system delivers justice for most people, but no justice for some. This highlights the need for a more sophisticated assessment of the reliability of a procedure. We cannot expect or demand that people obey the outcome of a procedure on the grounds that it tends to be reliable if they shoulder most of the burden of the unreliability. Some proceduralists accept this point. Rawls claims that the duty to obey the law is difficult to establish for minority groups who regularly bear a disproportionate amount of the burdens of procedural failure.[43]

So to recap, even if the legal procedure is highly reliable, the transmission

42 See Laudan, *Truth, Error, and Criminal Law*, 63.

43 Rawls, *A Theory of Justice*, 312.

of relevant normative properties to political decisions is blocked in cases of predictable failure. This is true for doxastic procedures, and intuitively for legal procedures like a criminal trial as well. Even if the procedure does not fail predictably, and even if highly reliable, certain distributions of failure can render it inappropriate and block the transmission of legitimacy.

4. DEMOCRATIC PROCEDURES

Criminal legal procedures are a kind of political procedure. So we should extend these considerations to other political procedures as well. That is the task of this section.

We have seen that Estlund offers an analogical argument between criminal trials and democratic procedures as a component of his overall case for the legitimacy and authority of democratic decisions. Rawls similarly takes the imperfect procedures to be authoritative even when they go wrong, but not always. This is made most clear in his discussion of civil disobedience. Civil disobedience is permitted, says Rawls, when injustices are clear and substantial, where "injustice" is understood as a state of affairs that violates the principles of equal liberty and fair equality of opportunity.[44] Christiano takes democratic decisions usually to be authoritative so long as they do not violate their fundamental commitment to equality of citizens (e.g., by disenfranchising some of the population). But were democratic institutions to nearly always generate seriously unjust outcomes, they would lack authority, for there "is no good reason for thinking that matters of distributive justice, individual rights and the common good are less normatively important than democratic principles."[45]

4.1. The Democracy/Jury Analogy

Since jailers should enforce sentences even when they are unjust because the legitimacy and authority of the decision is not canceled whenever it is wrong, the decision of a democratic institution is still authoritative even when it is wrong (unjust). The argument gets its force from the fact that no one endorses the heterodox implication that jailers should release prisoners if the jury or judge made an honest mistake.

If my argument is correct that the output of a trial procedure (culminating in the decision of a judge or jury) is not legitimate or authoritative in cases where the three appropriateness conditions are not met, then the analogical argument

44 Rawls, *A Theory of Justice*, 326.

45 Christiano, "The Authority of Democracy," 269.

justifies a more restrictive view of legitimacy or authority. The heterodox conclusion is compatible with—indeed, follows from—proceduralist considerations, rather than a rejection that the conclusion is actually heterodox.[46] When democratic procedures are not highly reliable, when they fail in predictable ways, or when they distribute burdens in an objectionable way, they too fail to transmit legitimacy.

Perhaps the proceduralist will want to reject the argument on the grounds that the democratic procedures are unique in a way that insulates them from this problem. But it would be surprising if democratic procedures were unlike trial procedures in this regard. Furthermore, the proceduralist must then take on the task of identifying how democratic procedures are like trial procedures such that the analogical argument establishes the authority of democratic decisions, but different in a way that undermines my modification of the view.

4.2. Political Procedural Failures

Determining whether various political procedures meet the reliability standard is a difficult task that I shall not take up here. Instead, it is worth focusing on the problems of predictable failure and the inappropriate distribution of failure. I want to raise the possibility that democratic procedures can fail in ways that undermine the transmission of the relevant normative property according to the instrumental proceduralist approach.

During close US presidential elections the electoral college is the subject of much discussion. This is a design problem: Is the electoral college a legitimacy-conducive feature or not? Here is one reason for thinking it is not. The electoral college has the effect of making votes in swing states count for more than votes in other states. This means that the votes of citizens are not equal. They are unequal in terms of their effect on the outcome, but in another important way as well. The disproportionate impact can motivate politicians to elevate their interests over the interests of their "safe vote" constituents, producing unjust electoral outcomes. This design problem threatens the appropriateness of the procedure.[47]

Consider now an input problem. If voters lack robust policy preferences and an ability to assess politicians holistically, they are susceptible to exploitation,

46 Cf. Brennan, *When All Else Fails*, 143.

47 This might seem like a pure procedural element, but it need not be. If we think that equalizing political power is important for high-performing political institutions producing high-quality governance, then design elements that create political inequality will degrade performance.

and the system subject to capture.[48] Voters, the empirical evidence suggests, form their policy preferences based on their self-identity and loyalty to groups.[49] If one has a partisan loyalty, they are likely to come to prefer the policies defended by that group. Voters are unlikely to form partisan loyalties based on prior policy preferences. Furthermore, voters are more likely to support the incumbent party when there is an economic uptick during the quarter leading up to the election. There is evidence that politicians exploit this by generating short-term economic bumps to coincide with elections. Political scientists have called this the "economic-electoral cycle."[50] This is only one instance of the myopia of voters. If such exploitation leads to the procedure getting things wrong, then this could be a case of predictable failure.

There are other ways in which the input to or design of democratic political procedures is problematic that do not rely upon claiming that the average citizen has inappropriate or wrongheaded policy preferences. Gerrymandering of districts to influence electoral outcomes is a well-known and blunt (though effective) design manipulation. The arbitrary or prejudicial restriction of suffrage is another example. Contemporary voter ID laws are candidates for a pernicious input problem. Felony disenfranchisement is another. Exacerbating this problem is the practice of counting, for the purpose of apportioning political representation, inmates in the districts where their prisons are located. Not only are racially biased prison populations deprived of their right to vote, but they also increase the political power of largely white, rural districts. When gerrymandering or disenfranchisement leads to unjust results we have another case of predictable failure.

Let us move to the distribution of failures. Suppose that the typical citizen has sensible policy preferences, and that gerrymandering and suffrage restriction do not constitute procedural failures sufficient to call into question the authority of the procedure's outputs. Still, it turns out that the majority sometimes has little influence on policy decisions. Here is how Martin Gilens and Benjamin I. Page characterize their findings:

> When the preferences of economic elites and the stands of organized interest groups are controlled for, the preferences of the average American appear to have only a minuscule, near-zero, statistically non-significant impact upon public policy.[51]

48 Guerrero, "Against Elections."

49 Achen and Bartels, *Democracy for Realists*.

50 Tufte, *Political Control of the Economy*; Achen and Bartels, *Democracy for Realists*.

51 Gilens and Page, "Testing Theories of American Politics," 575.

Instead, economic elites and organized groups representing business interests have far more influence on public policy. Regulatory capture and the existence of legislation written by lobbyists make this an unsurprising result. And when we know that democratic procedures are susceptible to the possibly pernicious influence of interest groups, and that this often leads to non-economic elites bearing much of the burden of the unjust political outcomes, public justification becomes significantly more difficult. There are no instrumental proceduralist grounds for insisting that groups of citizens obey a political decision when the procedure usually fails that group. The mechanism here, we might think, involves predictable failure, but the distributional concern raises an additional legitimacy problem.

As a disclaimer, I should emphasize that the success of these arguments does not rest entirely on the strength of these examples. One might reject the Gilens and Page account of the power of organized groups seeking concentrated benefits, or the hungry judges phenomenon, but I offer them as plausible illustrations of the kinds of failures instrumental proceduralists must take seriously. Just like legal procedures, our democratic procedures can fail to satisfy the appropriateness conditions needed to transmit relevant normative properties to their outputs. We have *proceduralist reasons* for restricting the scope of incorrect outputs that we regard as legitimate, authoritative, or just.

5. THE INTERACTION AND AMPLIFICATION OF PROCEDURAL FAILURES

One might object that, in at least some of these failure cases, the fault does not lie with the procedure. When Black defendants get harsher outcomes, it is not necessarily the procedure that causes this, but rather the background social facts the procedure is embedded in.[52] Surely there will be cases where procedural failures are not purely internal to the procedure. The conclusions I have defended are, however, compatible with this. Consider, for example, the Mercator projection example discussed above. There, the problem is not that the procedure is internally flawed. Rather, it is used in inappropriate circumstances. That is what explains the predictable nature of its occasional failures. Pairing the arguments so far with a structural injustice framework can illuminate the full scope of procedural failures along these lines and contribute to a response to this objection.

Structural injustices are distinct, according to Iris Marion Young, from "two other forms of harm or wrong, namely, that which comes about through individual interaction, and that which is attributable to the specific actions and

52 Thanks to a referee for raising this objection.

policies of states or other powerful institutions."[53] They are *emergent* injustices not directly attributable to particular culpable actions. Structural injustices can create inappropriate backgrounds that render our political procedures prone to predictable failures or inappropriate failure distributions. Overpolicing of certain areas that funnels a racially disproportionate set of offenders into the trial system, combined with racism or bias in the population, sets a background in which a procedure that includes prosecutors involved in the selection of jurors predictably results in the procedure failing (by, e.g., making it such that Black defendants predictably get unjustly harsh sentences).

Structural injustices can also contribute to the inappropriate distribution of procedural failures. The failures of misdemeanor systems, for example, include being detained for months for an offense that might deserve only a fine or a short jail term. These failures largely burden impoverished people precisely because they cannot afford bail. This exacerbates problematic distributions along other dimensions (e.g., race, since it is correlated with poverty). The predictable failure (here too a disproportionate punishment) is caused not only by features internal to the misdemeanor process, but also background structural injustices that play a causal role in the inappropriate failure distribution.[54]

Structural injustice, therefore, helps draw our attention to the causes, scope, and significance of these kinds of failures. It also draws our attention to the kind of changes that we must make to rectify various procedural failures (simply increasing funding for public defenders in an effort to improve the procedure itself is likely to be insufficient). Further, structural injustices can themselves be explained by procedural failures.[55] It is for this reason crucial to recognize the relationships between the imperfect procedures that constitute the institutions we live in. Procedural failures in one domain can lead to failures in others. Failures of judicial procedures fall disproportionately on certain groups of the population. These interact to generate failures for our democratic procedures, in turn risking the creation of vicious feedback loops.

Felony disenfranchisement serves as an obvious example. In many states in the US, felons are unable to vote. In other states felons are re-enfranchised upon

53 Young, *Responsibility for Justice*, 45.

54 For a thorough look at the many problems in the misdemeanor system, see Natapoff, *Punishment Without Crime*.

55 Housing policies that prevent the construction of new housing, leading to price increases and gentrification, are plausibly a manifestation of procedural failure (they exclude those who would benefit from new housing from the political procedure to determine whether a new development will be permitted). Unaffordable housing was, of course, a component of Young's major example (*Responsibility for Justice*, 43). For discussion, see Sankaran, "Structural Injustice as an Analytical Tool."

release, and in others upon completion of parole (and paying off sometimes prohibitively expensive fines). Some states require felons to apply for the right to vote, and their application can be denied. When our criminal legal procedures fail such that certain groups are more likely to make their way into the criminal justice system and receive felony convictions, this leads to the failure of political procedures.

Millions of US citizens are disenfranchised because they have been convicted of a felony. In some states the number is large enough to have a causal influence on political elections at state and national levels. This issue rose to prominence in the Bush versus Gore election of 2000, in which the election likely would have gone to Gore had some of the disenfranchised population been permitted to vote.[56] Here we have an example of procedural failure in judicial and political procedures changing the outcome of actual elections. If the procedure is responsible for transmitting legitimacy to the output, the procedural failure might make this outcome illegitimate.

This is not unique to the 2000 election. In 2016, Florida had 1.5 million citizens disenfranchised due to felony convictions.[57] During the presidential election that year, Donald Trump received 4,617,886 votes and Hilary Clinton received 4,504,975.[58] So in an election decided by 112,911 votes there were 1.5 million who were not permitted to vote. Further, one study estimates that 35 percent of felony disenfranchised citizens would vote in presidential elections—525,000 in Florida's 2016 election.[59] If we combine these observations with information about how some of the disenfranchised population would likely vote, then we can reasonably conclude that in some elections procedural failures could change the outcome of the election. And again, if procedural failures block the transmission of legitimacy, then we can conclude that these outcomes lack legitimacy.

In addition to felony disenfranchisement, mere contact with the criminal justice system has been shown to decrease political participation.[60] If unjust laws that are the result of procedural failure and unjust enforcement of those laws fall disproportionately on certain groups, then those groups are also less likely to

56 Uggen and Manza, "Democratic Contraction?"

57 Uggen, Larson, and Shannon, "6 Million Lost Voters."

58 Division of Elections, Florida Department of State, https://results.elections.myflorida.com/index.asp?electiondate=11/8/2016.

59 Uggen and Manza, "Democratic Contraction?" 786.

60 Being "arrested reduced the likelihood of voting by 7%; being convicted reduced the odds of turning out by 10%; being sentenced to jail or prison reduced it further by 17%, and serving more than 1 year reduced the likelihood of voting by nearly one third" (Weaver and Lerman, "Political Consequences of the Carceral State," 828).

be politically active. This has the further effect of politically marginalizing these groups and making it less likely that their interests will be taken into account when political groups determine whether to overturn or change the enforcement of existing laws. This is how various procedural failures generate vicious circles of failure.

What is important about this sort of example is that is does not rely on the view (which instrumental proceduralists would reject) that only optimific procedures transmit legitimacy or authority. The problem is not that the outcome of some presidential elections is not optimal or ideal, but rather that it is different from what it would be without predictable, nonuniformly distributed procedural failures.

There are plenty of procedures that fail silently, rather than loudly. In other words, the failures are not so egregious that we immediately notice them, or that proceduralists are happy to deny are legitimate. What I have suggested is that we need to refine our evaluation of imperfect procedures to look for instances of less obvious failure. One implication of the wide scope of procedural failure is that more work on proceduralism needs to be done in the realm of nonideal theory. When we idealize away these problems we are left with no real answer to questions about the legitimacy or authority of actual political institutions.

Unevenly distributed procedural failures are dangerous because they interact with other procedures, tending to generate and amplify additional failures. This can be self-reinforcing and amplifying in many contexts. For this reason in particular, instrumental proceduralists cannot rely on a reliability rate alone. We must attend also to the possibility of predictable and unevenly distributed failure.

The instrumental proceduralist, by offering a fallibilist account of legitimacy, encourages us to draw a distinction between what we can call *culpable mistakes* from *honest mistakes*. A bribed judge produces a culpable mistake that blocks legitimacy, but a mistake from a properly functioning procedure is an honest mistake involving no culpable wrongdoing and therefore preserves legitimacy. But the structural injustice literature demonstrates that looking only at culpably unjust actions is overly narrow and deprives us of an important evaluative tool. Similarly, the revised set of appropriateness conditions argued for here demonstrates the importance of another kind of mistake. These are non-culpable but seriously unjust procedural errors (arising from insufficiently reliable, predictably failing, or inappropriately distributing procedures). They are a kind of disqualifying mistake that is not dishonest in the sense of being a result of individual culpability, but neither are they exactly honest in the way that a fallibilist sensibility would lead us to begrudgingly accept as legitimate. When we

reflect on the nature of instrumental proceduralist justification we see that the view provides us with the resources to diagnose these problems. Proceduralists should embrace those resources and take on board more stringent success conditions for instrumental proceduralist justification.[61]

University of New Orleans
jakemonaghan@me.com

REFERENCES

Abrams, David S., and Albert H. Yoon. "The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability." *University of Chicago Law Review* 74, no. 4 (2007): 1145–78.

Achen, Christopher H., and Larry M. Bartels. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton Studies in Political Behavior. Princeton: Princeton University Press, 2016.

Amiri, Farnoush, Ethan Sacks, and Kerry Sanders. "Georgia Officers on Leave after Coin-Toss App Used before Decision to Make Arrest." NBC News, July 13, 2018. https://www.nbcnews.com/news/us-news/georgia-officers-leave-over-after-coin-toss-used-decision-make-n891306.

Arneson, Richard J. "Defending the Purely Instrumental Account of Democratic Legitimacy." *Journal of Political Philosophy* 11, no. 1 (March 2003): 122–32.

Ayres, Ian, and Joel Waldfogel. "A Market Test for Race Discrimination in Bail Setting." *Stanford Law Review* 46, no. 5 (May 1994): 987–1047.

Baldus, David C., Julie Brain, Neil A. Weiner, and George Woodworth. "Evidence of Racial Discrimination in the Use of the Death Penalty: A Story from Southwest Arkansas (1990–2005) with Special Reference to the Case of Death Row Inmate Frank Williams, Jr." *Tennessee Law Review* 76, no. 3 (2008): 555–614.

Baldus, David C., Charles Pulaski, and George Woodworth. "Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience." *Journal of Criminal Law and Criminology* 74, no. 3 (Autumn 1983): 661.

Blair, Irene V., Charles M. Judd, and Kristine M. Chapleau. "The Influence of Afrocentric Facial Features in Criminal Sentencing." *Psychological Science* 15, no. 10 (October 2004): 674–79.

Brennan, Jason. *Against Democracy*. Princeton: Princeton University Press, 2016.

61 I am grateful to Ryan Muldoon, Jerry Gaus, David Estlund, David Boonin, Kirun Sankaran, Danny Shahar, and J. P. Messina for feedback on earlier drafts of this article.

———. *When All Else Fails*. Princeton: Princeton University Press, 2019.
Buchanan, Allen. "Political Legitimacy and Democracy." *Ethics* 112, no. 4 (July 2002): 689–719.
Christiano, Thomas. "The Authority of Democracy." *Journal of Political Philosophy* 12, no. 3 (September 2004): 266–90.
Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. "Extraneous Factors in Judicial Decisions." *Proceedings of the National Academy of Sciences* 108, no. 17 (April 26, 2011): 6889–92.
Dretske, Fred. "Conclusive Reasons." *Australasian Journal of Philosophy* 49, no. 1 (May 1971): 1–22.
———. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press, 1981.
Eberhardt, Jennifer L., Paul G. Davies, Valerie J. Purdie-Vaughns, and Sheri Lynn Johnson. "Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes." *Psychological Science* 17, no. 5 (May 2006): 383–86.
Enoch, David. "Authority and Reason-Giving." *Philosophy and Phenomenological Research* 89, no. 2 (September 2014): 296–332.
Estlund, David. "On Following Orders in an Unjust War." *Journal of Political Philosophy* 15, no. 2 (June 2007): 213–34.
———. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press, 2009.
Feldman, Richard. *Epistemology*. Hoboken, NJ: Prentice Hall, 2003.
———. "Reliability and Justification." *Monist* 68, no. 2 (April 1985): 159–74.
Frederique, Nadine, Patricia Joseph, and R. Christopher Child. "What Is the State of Empirical Research on Indigent Defense Nationwide? A Brief Overview and Suggestions for Future Research." *Albany Law Review* 78, no. 3 (2015): 1317–40.
Gettier, Edmund. "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (June 1963): 121–23.
Gilens, Martin, and Benjamin I. Page. "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens." *Perspectives on Politics* 12, no. 3 (September 2014): 564–81.
Goldman, Alvin I. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press, 1988.
———. *Reliabilism and Contemporary Epistemology: Essays*. New York: Oxford University Press, 2012.
Guerrero, Alexander. "Against Elections: The Lottocratic Alternative." *Philosophy and Public Affairs* 42, no. 2 (March 2014): 135–78.

Huemer, Michael. "Phenomenal Conservatism and the Internalist Intuition." *American Philosophical Quarterly* 43, no. 2 (2006): 147–58.

Larmore, Charles. "The Moral Basis of Political Liberalism." *Journal of Philosophy* 96, no. 12 (December 1999): 599–625.

Laudan, Larry. *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*. Cambridge: Cambridge University Press, 2006.

Levinson, Justin D., Robert J. Smith, and Danielle M. Young. "Devaluing Death: An Empirical Study of Implicit Racial Bias on Jury-Eligible Citizens in Six Death Penalty States." *New York University Law Review* 89 (2014): 515–81.

Monaghan, Jake, Eric Joseph van Holm, and Chris W. Surprenant. "Get Jailed, Jump Bail? The Impacts of Cash Bail on Failure to Appear and Re-Arrest in Orleans Parish." *American Journal of Criminal Justice* 47, no. 1 (February 2022): 56–74.

Mustard, David B. "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts." *Journal of Law and Economics* 44, no. 1 (April 2001): 285–314.

Natapoff, Alexandra. P*unishment Without Crime: How Our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal*. New York: Basic Books, 2018.

Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.

Peter, Fabienne. *Democratic Legitimacy*. New York: Routledge, 2009.

———. "Pure Epistemic Proceduralism." *Episteme* 5, no. 1 (February 2008): 33–55.

Pierce, Glenn L, and Michael L Radelet. "Death Sentencing in East Baton Rouge Parish, 1990–2008." *Louisiana Law Review* 71, no. 2 (2011): 647–72.

Rachlinski, Jeffrey J, Sheri Johnson, Andrew J. Wistrich, and Chris Guthrie. "Does Unconscious Racial Bias Affect Trial Judges?" *Notre Dame Law Review* 84, no. 3 (2009): 1195–1246.

Radelet, Michael L., and Glenn L. Pierce. "Race and Death Sentencing in North Carolina, 1980–2007." *North Carolina Law Review* 89 (2011): 2119–60.

Rawls, John. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77, no. 9 (September 1980): 515–72.

———. *A Theory of Justice*. Rev. ed. Cambridge, MA: Belknap Press of Harvard University Press, 1999.

Sankaran, Kirun. "'Structual Injustice' as an Analytical Tool." *Philosophy Compass* 16, no. 10 (October 2021): 1–12.

Simmons, A. John. "Justification and Legitimacy." *Ethics* 109, no. 4 (July 1999): 739–71.

Tufte, Edward R. *Political Control of the Economy*. Princeton: Princeton University Press, 1978.

Uggen, Christopher, Ryan Larson, and Sarah Shannon. "6 Million Lost Voters: State-Level Estimates of Felony Disenfranchisement, 2016." Sentencing Project, October 6, 2016. https://www.sentencingproject.org/publications/6-million-lost-voters-state-level-estimates-felony-disenfranchisement-2016.

Uggen, Christopher, and Jeff Manza. "Democratic Contraction? Political Consequences of Felon Disenfranchisement in the United States." *American Sociological Review* 67, no. 6 (December 2002): 777–803.

Vallier, Kevin. "Against Public Reason Liberalism's Accessibility Requirement." *Journal of Moral Philosophy* 8, no. 3 (January 2011): 366–89.

Vidmar, Neil. "The Psychology of Trial Judging." *Current Directions in Psychological Science* 20, no. 1 (February 2011): 58–62.

Weaver, Vesla M., and Amy E. Lerman. "Political Consequences of the Carceral State." *American Political Science Review* 104, no. 4 (November 2010): 817–33.

Young, Iris Marion. *Responsibility for Justice*. New York: Oxford University Press, 2011.

Journal of Ethics and Social Philosophy
VOL. 22, NO. 1 · MAY 2022
https://doi.org/10.26556/jesp.v22i1.1647

CONSTRAINED FAIRNESS IN DISTRIBUTION

Daniel M. Hausman

GERARD VONG addresses intriguing problems in which it may be impossible to give an equal chance of receiving a good to a set of equal claimants.[1] After developing Vong's views in sections 1 and 2, in section 3, I point out an implausible feature of algorithms that attempt to integrate concerns about comparative fairness and what Vong calls "absolute fairness." I then argue in section 4 against attempting to integrate concerns about comparative and absolute fairness.

1. INTRODUCTION

Following John Broome, Vong takes an individual *Q* to have a "claim" to a good *G* on some agent *A* if and only if *A* has a *pro tanto* duty to provide *Q* with *G*.[2] When individuals have equal claims to some good, it seems comparatively fair to give them equal shares of the good or, if the good is indivisible, equal chances of getting the good. In the cases Vong has identified, it is impossible to provide *G* to some individuals without also providing it to everyone in some group to which they belong. The good goes to all and only group members. These division problems resemble those discussed by John Taurek, where a drug can save the life of one person or five persons.[3] In these cases, unlike the conflicting claims to some indivisible good that can be possessed by only one person, the distribution of the good determines how many as well as which people get the good, and, *contra* Broome, Vong maintains that equal claimants need not be given equal chances. Indeed, in the case of overlapping groups (where individuals can be benefitted through their membership in more than one group), equal chances may be impossible. For example, suppose that the chance that any of the six individuals *A*, *B*, *C*, *D*, *E*, and *F* gets a good *G* depends on the chances that

1 Vong, "Weighing Up Weighted Lotteries."

2 Broome, "Fairness"; Vong, "Weighing Up Weighted Lotteries."

3 Taurek, "Should the Numbers Count?"

G will go to one of the following four couples: *A* & *B*, *A* & *C*, *D* & *E*, or *D* & *F*.[4] There is no way to give the six individuals equal, nonzero chances of enjoying *G*. All possible lotteries, other than one that gives no one any chance, assign unequal chances to individuals. In this case, Vong suggests that it is fair to give equal chances to each couple, even though that means that individuals *A* and *D* are twice as likely to receive *G* as are the others. What principles imply that this unequal lottery is fairer than others?[5]

Vong maintains that fairness is an amalgam of two species.[6] One is comparative, which counts distributions as fair if chances or shares of the good are in proportion to the strength of claims.[7] The other measures the fairness of a distribution by how many claims it satisfies and by how fully it satisfies them, regardless of comparisons to how fully the claims of other individuals are satisfied. A distribution that awarded everyone half of what they claim, when their claims could have been completely satisfied, is comparatively fair and absolutely unfair.

Vong seeks some criterion that reflects the moral importance of both comparative and absolute fairness.[8] I argue in section 4 that it is better to offer separate assessments of the absolute and comparative fairness of distributions, whose weights vary with context. Until then, I will follow Vong and consider which distributions are fairest "overall." I shall impose the constraint that lotteries be *efficient*: the probabilities they assign to overlapping groups add up to one and the shares of divisible goods that are assigned to overlapping groups exhaust the good. This constraint can be defended both on the grounds of absolute fairness and on welfarist grounds.

2. EXCLUSIVE COMPOSITION-SENSITIVE LOTTERIES

Vong considers several ways to distribute chances among groups in order to treat claimants fairly, and he favors what he calls "exclusive composition-sensitive lotteries" (hereafter EXCS lotteries).[9] The characterization is complicated, and the reader may want to skip to the example in the following paragraph. In EXCS lotteries, each of the *n* equal claimants is assigned an initial baseline weight of 1/n.

4 Vong, "Weighing Up Weighted Lotteries," 324.

5 One answer: it maximizes the minimum chance that any individual will win.

6 Vong, "Weighing Up Weighted Lotteries," 326–27.

7 Broome, "Fairness." Like Broome, I regard fairness as comparative, but in this essay I follow Vong's terminology, expressing later some skepticism about whether absolute and comparative fairness have the same normative source.

8 Vong, "Weighing Up Weighted Lotteries," 332.

9 Vong, "Weighing Up Weighted Lotteries," 335.

Each individual *j*'s baseline weight is distributed among the groups in which *j* is a member. The fraction of *j*'s weight assigned to a group depends on how many members in the group are "distributively relevant" to *j*, divided by the total number of members distributively relevant to *j* in all the groups.[10] A member *k* of a group containing *j* is distributively relevant to *j* in that group if it matters to *j* how *k*'s baseline probability is distributed among groups. If *k* is in some groups that do not include *j*, then it matters to *j* how *k*'s baseline probability is distributed and *k* is distributively relevant to *j*. If every group containing *k* also contains *j*, then *k* is not distributively relevant to *j*. If an individual, *j*, is in only one group, then *j*'s entirely baseline probability is assigned to that group.

For example, consider:

> *Problem**: There are four equal claimants, Ann, Bill, Chuck, and Diane (*A*, *B*, *C*, and *D*). It is possible to distribute chances of getting some good to them only by distributing chances of getting the good to the groups *A* & *B*, *A* & *B* & *C*, *C* & *D*, and *B* & *C*. The baseline probability for each individual is ¼. *A* is not distributively relevant to *B*, because every group containing *A* also contains *B*. *B* is distributively relevant to *A*.

Table 1 lists the distributive relevancies and calculates the chances in the lottery.

Table 1

Group	Distributive Relevancies	Calculation	Chance
A & *B*	*B* to *A* (1 of 4)	¼ × ¼	1/16
A & *B* & *C*	*A* to *C* (1 of 1); *B* to *A* and *C* (2 of 4); *C* to *A* and *B* (2 of 4)	¼ (1 + ½ + ½)	½
C & *D*	*C* to *D* (1 of 4); *D*'s full baseline (1)	¼ (1 + ¼)	5/16
B & *C*	*B* to *C* (1 of 4); *C* to *B* (1 of 4)	¼ (½)	⅛

We are not quite done. Because it is unfair (and inefficient) to assign any nonzero probability to a subset of another set, the chances assigned to *A* & *B* and to *B* & C should be distributed among the sets containing these subsets, in this case, *A* & *B* & *C*.[11] In Problem*, the EXCS lottery assigns an 11/16 chance to *A* & *B* & *C*

10 Where I speak of *k* being "distributively relevance" to *j*, Vong speaks of *j* as "exclusive" to *k*. I find that this change makes Vong's proposal easier to follow.

11 Vong, "Weighing Up Weighted Lotteries," 342.

and a 5/16 chance to $C \& D$. This implies: $\Pr(A) = \Pr(B) = 11/16$, $\Pr(D) = 5/16$, and $\Pr(C) = 1$.

3. Problems with EXCS and Other Lotteries

Because the groups $B \& C$ and $A \& B$ are subsets of $A \& B \& C$, the EXCS lottery quite rightly gives them no chance of getting the good. Yet, as table 2 shows, the chance that the claims of individuals in different groups are satisfied depends on whether claims could be satisfied via the two subset groups, even though it would never be fair to give them any chance of getting G.

Table 2

Group	Distributive Relevancies	Calculation	Chance
ABC	*C* to *A* and *B* (2 of 3); *A*'s and *B*'s full baselines (2)	$(1/4)(1 + 1 + 2/3)$	2/3
CD	*C* to *D* (1 of 3); *D*'s full baseline (1)	$(1/4)(1 + 1/3)$	1/3

The lotteries derived in tables 1 and 2 assign chances to the same equal claimants, and both assign nonzero chances only to groups $A \& B \& C$ and $C \& D$. Yet which distribution to the four individuals is fair depends on whether one employs Vong's two-step procedure to decide how to distribute chances among the four groups, or whether one starts by ruling $A \& B$ and $B \& C$ out of the lottery on the grounds that they must wind up with a zero probability. In that case, *A*'s and *B*'s chances would be lower (2/3 rather than 11/16), and *D*'s chances higher (1/3 rather than 5/16). This result is implausible. Regardless of the status that groups have in other contexts, their only role here is to specify which distributions among individuals are possible. Whether an assignment of chances treats the four equal claimants fairly should not depend on whether they belong to groups to which no chance is given. This is not a bargaining problem, wherein the possibility of individuals getting the good by themselves or via coalitions gives them a threat advantage.[12]

There are alternatives to EXCS lotteries to consider. Suppose one weights each alternative by the proportion of the individual claimants it contains and then multiplies each weight by the reciprocal of the sum of the weights so that the weights add up to one. This method implies that $\Pr(A) = \Pr(B) = 7/9$, $\Pr(D) = 2/9$,

12 Moreover, since every fair distribution gives *C* the good, the distribution of *C*'s baseline probability should be irrelevant.

and $\Pr(C) = 1$.[13] If, however, one begins by eliminating the groups with zero probabilities, then the chances for the two groups $A \& B \& C$ and $C \& D$ should be ⅗ and ⅖, and the probabilities among the four individuals are: $\Pr(A) = \Pr(B) = ⅗$, $\Pr(C) = 1$, and $\Pr(D) = ⅖$. Proportional lotteries, like EXCS lotteries, imply that the fairest weighted lottery among equal claimants depends on the treatment of groups to which the lottery assigns zero probability.

Vong discusses and criticizes a third method of assigning chances to lotteries, which he calls "equal composition-sensitive lotteries."[14] In these "EQCS lotteries," the chance of each group is the sum of fractions consisting of the baseline probabilities for each individual divided by the number of groups in which the individual is a member. The values EQCS lotteries assign to $A \& B \& C$ and $C \& D$ also vary depending on how one deals with the zero-probability groups.[15]

There is an easy way to avoid the untoward dependence on membership in groups to which fair lotteries assign no chance: simply delete all groups that are subsets of other groups before calculating the chances. But that solution does not explain why these methods of assigning chances when there are overlapping groups are responsive to whether there are groups to which fair lotteries assign zero probabilities of benefitting. Nor does it help us decide among EXCS, EQCS, and proportional lotteries.[16]

4. ADJUDICATING AMONG LOTTERIES

Vong offers an example that he believes supports employing EXCS lotteries and undermines the employment of EQCS lotteries.[17] I draw different conclusions. Consider the groups, G_1, G_2, and G_3. G_1 contains claimants 1 through 500. G_2 contains claimants 501 to 1,000. G_3 contains claimants 2 to 999. The EQCS lottery

13 This adopts Frances Kamm's proportionality proposal (*Morality, Mortality*, 124) and renormalizes so that the weights assigned to groups add up to 1. In this example, the weights assigned to $A \& B$, $A \& B \& C$, $C \& D$, and BC would be ¼, ¾, ¼, ¼. The sum is 9/4. Multiplying by 4/9, the groups' chances would be ⅔, ⅓, ⅔, and ⅔. Donating $B \& C$'s and $A \& B$'s probabilities to $A \& B \& C$, the result is $\Pr(A \& B \& C) = ⅞$ and $\Pr(C \& D) = ⅔$.

14 Vong, "Weighing Up Weighted Lotteries," 334.

15 In this example, $\Pr(A \& B) = 5/24$, $\Pr(A \& B \& C) = 7/24$, $\Pr(C \& D) = ⅓$, and $\Pr(B \& C) = ⅙$. The fair lottery if one starts with four groups gives A and B a ⅔ chance ($^{(5+7+4)}/_{24}$) and D a ⅓ chance. If one starts with two groups, A and B each have a ⅝ chance while D has a ⅜ chance. C, of course, is sure to win.

16 Vong ("Weighing Up Weighted Lotteries," 338) also discusses an iterated version of Timmerman's individualist lottery, which I shall not discuss; see Timmermann, "The Individualist Lottery."

17 Vong, "Weighing Up Weighted Lotteries," 339–40.

assigns a chance of a little more than a quarter to each of G_1 and G_2, and a little less than one half to G_3.

Vong finds this result intolerable:

> A theory of fairness that utilizes the equal composition-sensitive lottery procedure gives the startlingly implausible result that it is fair to give a greater than 50 percent chance to save [members of] either one of G_1 or G_2, making it more likely that 500 claimants rather than 998 claimants will be saved. This is an affront to absolute fairness because benefiting the much larger group of 998 claimants is less likely than benefiting one of the much smaller groups containing 500 claimants.[18]

Vong's EXCS lottery, in contrast, gives about a 96 percent chance to G_3. The EXCS lottery probably satisfies many more claims than the EQCS lottery. It is far fairer absolutely. However, Vong's EXCS lottery gives individuals 1 and 1,000 a vastly lower 2 percent chance of getting *G*. On Broome's view of comparative fairness as requiring equal chances for equal claimants, G_1 and G_2 should have equal chances of ½. On Kamm's proportional view with the renormalization discussed above, G_1 and G_2 should have a little more than a 25 percent chance and G_3 a little under a 50 percent chance. So individuals 1 and 1,000 will have about a 25 percent chance of getting the good, while everyone else will have about a ¾ chance. This seems fairer comparatively, but, as Vong argues, less fair absolutely. Vong's proposal, with its focus on distributive relevance—that is, whether *j*'s benefitting affects *k*'s benefitting—makes the magnitude of expected claim satisfaction the dominant factor here: the larger the chance of G_3, the greater the "absolute" fairness.

There are two moral considerations here—in Vong's terminology, absolute and comparative fairness. Whereas Vong sees these as two faces of the same coin, I see one as a matter of how one shows respect to individuals, while the other is focused on satisfying duties to individuals. What is absolutely fairest is to give the good to G_3, which fully satisfies 998 claims. What is, on Broome's view, fairest comparatively is to give everyone the same ½ chance by giving that chance to G_1 and G_2. Vong accepts the comparative unfairness of the EXCS lottery, because he seeks a rule for assigning chances that integrates absolute and comparative fairness.

I think that Vong's search for a context-invariant compromise between absolute and comparative fairness is a mistake. It is more perspicuous to separate the questions concerning absolute and comparative fairness and to allow the trade-

18 Vong, "Weighing Up Weighted Lotteries," 340.

off to respond to details of the specific circumstances, which may include other ethically relevant aspects. These sometimes call for compromises and sometimes respond to one consideration, passing over the other. In the case concerning G_1, G_2, and G_3, what is comparatively fairest is so different from what is absolutely fairest that compromises are not plausible: one should give the good to G_3 despite its comparative unfairness if the good that individuals have claims to is a lifesaving medicine. This is far better on the grounds of well-being as well as absolute fairness. On the other hand, if the good were seats at a presidential inauguration, it may be more important that people be treated equally than that so many more with claims to attend are able to do so.

There are other cases where the demands of absolute and comparative fairness should affect the distribution. The quandaries concerning the allocation of COVID-19 vaccines might be examples. What I am questioning is whether integrations of comparative and absolute fairness concerns, such as EXCS, EQCS, and proportional lotteries, are helpful in guiding ethical decisions. Having determined what is comparatively fairest and what is absolutely fairest, one needs to decide how to distribute the chances, taking into account other relevant moral considerations. There is no reason to insist on a uniform adjudication of just two of the considerations.

Depending on the method and whether one ignores the two groups in Problem* to which no chance is given, we have seen arguments for several different assignments of chances. In the specific problem, all of the different ways of apportioning chances among the four individuals seem plausible in the abstract. I see no good argument for defending one of these as the overall fairest without attending to the characteristics of the good and of the claims to the good.

Eschewing the determination of which distribution is fairest overall leaves one with the tasks of judging which distributions are fairest comparatively and which are fairest absolutely. I suggest that the comparatively fairest distribution assigns shares and chances to equal claimants that are as equal as possible or that maximizes the minimum chance of receiving the good. In the case of Problem*, giving *A* and *B* a 6/11 chance and *D* a 5/11 chance minimizes the variance. However, giving *A*, *B*, and *D* each a one-half chance of getting *G* maximizes the minimum chance and perfectly equalizes the chances for everyone except *C*, who is in any case guaranteed to get the good and whose chance is hence arguably irrelevant to which distribution is fairest. Absolute fairness is not simple either, unless it is just a matter of how many claims are satisfied, as is the case here, where giving the good to the group *A* & *B* & *C* guarantees that three of the four individuals will have their claims satisfied. Which distribution is overall fairest, let alone best, all things considered, depends on the context. It may be what is comparatively fair-

est, what is absolutely fairest, some compromise, or an unfair distribution that is ethically attractive on other grounds.

5. CONCLUSION

Overlapping groups pose theoretical problems concerning how to distribute goods or chances fairly. Compromises such as Vong's EXCS lotteries have implausible implications, which can be avoided by addressing separately the comparative and absolute fairness of distributions of chances or goods. Rather than seeking some general algorithm to assign the proper significance to these separate moral considerations, allocators should look to the details of the context to prioritize these separate considerations of fairness and other relevant ethical considerations such as well-being.[19]

Rutgers University
dhausman@cplb.rutgers.edu

REFERENCES

Broome, John. "Fairness." *Proceedings of the Aristotelian Society* 91 (1990): 87–101

Kamm, Frances. *Mortality, Morality*, vol. 1. New York: Oxford University Press, 1993.

Taurek, John. "Should the Numbers Count?" *Philosophy and Public Affairs* 6, no. 4 (Summer 1977): 293–316.

Timmermann, Jens. "The Individualist Lottery: How People Count, but Not Their Numbers." *Analysis* 64, no. 2 (April 2004): 106–12.

Vong, Gerard. "Weighing Up Weighted Lotteries: Scarcity, Overlap Cases, and Fair Inequalities of Chance." *Ethics* 130, no. 3 (April 2020): 320–48.

19 I am indebted to Gerard Vong for helpful conversations concerning his essay and this comment.