ANNETTE ZIMMERMANN

Criminal Disenfranchisement and the Concept of Political Wrongdoing

Abstract: Disagreement persists about when, if at all, disenfranchisement is a *fitting* response to criminal wrongdoing of type X. Positive retributivists endorse a permissive view of fittingness: on this view, disenfranchising a remarkably wide range of morally serious criminal wrongdoers is justified. But defining fittingness in the context of criminal disenfranchisement in such broad terms is implausible, since many crimes sanctioned via disenfranchisement have little to do with democratic participation in the first place: the link between the nature of a criminal act X (the 'desert basis') and a fitting sanction Y is insufficiently direct in such cases. I define a new, much narrower account of the kind of criminal wrongdoing which is a more plausible desert basis for disenfranchisement: 'political wrongdoing', such as *electioneering, corruption, or conspiracy with foreign powers*. I conclude that widespread blanket and post-incarceration disenfranchisement policies are *overinclusive*, because they disenfranchise persons guilty of serious, but non-political, criminal wrongdoing. While such overinclusiveness is objectionable in any context, it is particularly objectionable in circumstances in which it has additional large-scale collateral consequences, for instance by perpetuating existing structures of racial injustice. At the same time, current policies are *underinclusive*, thus hindering the aim of holding political wrongdoers accountable.

Philosophy & Public Affairs (Early View, 2019). This is the accepted version. Please cite the published version (https://doi.org/10.1111/papa.12153).

Over one in five African American citizens in Florida, Kentucky, Tennessee, and Virginia are currently disenfranchised due to a present or past felony conviction.[2] 6.1 million US citizens were disenfranchised as of November 2016, although 1.4 million citizens became re-eligible to vote again on January 8th, 2019, when Florida's constitutional amendment 4 *(Voting Rights Restoration for Felons Initiative)* came into effect. Irrespective of these recent changes, however, the lasting political impact of criminal disenfranchisement remains significant, particularly in the US, where many states disenfranchise not only those currently serving prison time, but also parolees and ex-convicts. Studies argue that even if the US had enfranchised only those felons who already

I am grateful to Chuck Beitz, Ben Berger, Anuja Bose, Dan Butt, Shuk Ying Chan, Tom Christiano, Lars Christie, Stefan Eich, Cécile Laborde, Melissa Lane, David Miller, Jan Werner Müller, Alan Patten, Talia Schaffer, Leonie Schulte, audiences at Princeton University's Political Theory seminar and the University of Oxford's Nuffield Political Theory Workshop as well as students in the 2019 Princeton Prison Teaching Initiative class "Punishment: Theory and Practice", and two anonymous editors of this journal, for their illuminating and generous comments. For insightful extensive written correspondence I am especially indebted to Cécile Fabre, Élise Rouméas, Annie Stilz, and Steve White.

[2] Christopher Uggen, Ryan Larson, and Sarah Shannon, "6 Million Lost Voters: State-Level Estimates of Felony Disenfranchisement, 2016," *The Sentencing Project* (2016): http://www.sentencingproject.org/publications/6-million-lost-voters-state-level-estimates-felony-disenfranchisement-2016/

served their time in prison, Democrats might have controlled the Senate throughout the 1990s, and that the outcomes of several presidential elections—most notably in 2000—might have been different.[3] More importantly, when blanket disenfranchisement policies are implemented in a wider empirical context of mass incarceration with salient disparate consequences for oppressed minorities, they risk further entrenching, and indeed exacerbating, conditions of structural racial injustice.

But even if the number of people at risk of long-term democratic exclusion is lower in a given polity than it is in the US, criminal disenfranchisement raises difficult questions for democratic theory. The UK, where the prison population is much smaller,[4] and thus, the number of disenfranchised prisoners is much lower, imposed a blanket disenfranchisement policy for all prisoners until the end of 2017. This number will be even lower going forward, since the UK government has recently, after a twelve year standoff, decided to change its policy to comply with a 2005 European Court of Human Rights judgment in the landmark case *Hirst v. UK*, which stated that the UK's blanket ban violated Article 3 of Protocol No. 1 to the ECHR.[5] Currently, the blanket ban continues to apply to all prisoners in principle, with a few exceptions: for those released on temporary licence or on home detention, those on remand, and those committed to prison for contempt of court or for default on paying fines.[6] Yet it would be incorrect to infer that the practice is less objectionable given the comparatively smaller number of disenfranchised criminals, *especially* since the racially disparate, and thus structurally unjust, impact of disenfranchisement also applies to the UK.[7]

The view that *temporary* disenfranchisement, contemporaneous with an offender's term of incarceration, is much less morally troubling than post-incarceration or permanent disenfranchisement is widespread. While it is plausible to think that, in general, temporary

---

[3] Pamela Karlan, "Convictions and Doubts: Retribution, Representation, and the Debate over Felon Disenfranchisement," *Stanford Law Review* 56 (2004): 1157; Hugh LaFollette, "Collateral Consequences of Punishment: Civil Penalties Accompanying Formal Punishment," *Journal of Applied Philosophy* 22 (2005): 241.

[4] In March 2018, there were 83,263 prisoners in England and Wales. This number includes short-term prisoners, those on remand, those subsequently acquitted, and foreign nationals. "Offender Management Statistics Bulletin," London: UK Ministry of Justice (2017), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702297/omsq-q4-2017.pdf

[5] *Hirst v. UK* (No. 2), ECtHR (2005), http://hudoc.echr.coe.int/eng?i=001-70442

[6] "Communication from the United Kingdom Concerning the Case of Hirst (No. 2) v. the United Kingdom: Action plan DH-DD(2017)1229," *Council of Europe: Secretariat of the Committee of Ministers* (2017): https://rm.coe.int/1680763233

[7] For a UK-specific report on racial injustice and incarceration, see UK Ministry of Justice Analytical Services, "Exploratory analysis of 10-17 year olds in the youth secure estate by black and other minority ethnic groups," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/641481/Exploratory-analysis-of-10-17-year-olds-in-the-youth-secure-estate-by-bame-groups.pdf

punishments of type P are less severe than permanent punishments of the same type P, this general claim by itself does not have a clear necessary upshot for the issue at stake in this paper: whether criminal disenfranchisement, including in its temporary form, is morally permissible *all things considered*. The severity of a punishment is one obviously relevant consideration for determining whether imposing that punishment is morally permissible, since some severe punishments may be unduly burdensome for the criminal offender; but it not the *only* consideration: a plausible theory of punishment must also identify the proper *positive* normative justification for imposing punishment in the first place, such as the nature of the criminal act at stake and the associated benefits or aptness of a given punitive sanction. While such positive justifications underpinning views defending temporary disenfranchisement on deterrentist, expressivist, and positive retributivist grounds are more plausible *prima facie* than their respective counterparts demanding post-incarceration and permanent criminal disenfranchisement, we must still assess each positive justification's merits in turn. I argue that we have strong reasons to reject each of these views in favour of a different, more restricted view.

I define the set of rights that can be temporarily forfeited during criminal disenfranchisement as the full set of democratic participation rights—including, for instance, the right to vote, the right to run for office in a given election and to continue to hold public office once elected, or the right to contribute financially to political campaigns, but excluding the more general, higher-order right to free expression.[8] This is a broader view of criminal disenfranchisement than the majority of contributors to the philosophical debate have adopted, many of whom focus exclusively on the right to vote, thus failing to consider other forms of (dis-)enfranchisement. Even if a criminal retains her right to vote, revoking other participation rights is a possible punishment, and in that case, it would be counterintuitive to say that she continues to be *fully* enfranchised. Her set of participation opportunities in democratic decision-making has been punitively limited. Therefore, criminal disenfranchisement is best construed as a broader notion, accommodating the intuition that disenfranchisement can be a matter of degree.

This paper approaches criminal disenfranchisement from the view that there is a *prima facie* presumption in favour of universal suffrage, which implies an especially high justificatory threshold for any deviation from that presumption. As the ECtHR ruling in *Hirst v. UK* put it, "[a]ny departure from the principle of universal suffrage risks undermining the democratic validity of the legislature thus elected and the laws which it promulgates" (§62). The value of political inclusion is particularly high because the

[8] Thanks to Chuck Beitz for pushing me to clarify this point.

realization of that value via universal suffrage is *important* for securing other important values and rights in a democratic state, including individual liberty rights.[9] This is an instrumentalist justification of the value of political inclusion: given political realities of unequal power, protecting the *equal* liberty of democratic citizens will depend in part on every citizen's ability to publicly exercise partial but meaningful democratic control over how individual rights and liberties are protected within a democratic state: democratic participation rights tend to be protectors of other rights. Thus, while punishments *not* involving disenfranchisement—such as incarceration—may well be considered morally troubling on the grounds that they limit individual liberty rights, they are less morally troubling than disenfranchisement in that limits on universal democratic participation rights prevent those who are temporarily or permanently excluded from democratically contesting the ways in which democratic states may permissibly limit their liberty rights, and the liberty rights of others.[10] The presumption in favour of universal suffrage thus suggests that disenfranchisement as a deviation from an important norm merits particularly careful normative justification, especially in light of our historical experience with widespread unjust types of disenfranchisement, such as the exclusion of women, persons of color, and members of the working class. Rather than thinking of full positive enfranchisement as conditional on certain standards being met, we should therefore think of disenfranchisement as an *exception* permissible only on the condition that particularly stringent justificatory standards are met.

If one agrees that there is a *prima facie* presumption in favour of universal suffrage, one must also agree that the burden of proof rests on the shoulders of those wishing to deny the vote to a particular person to show why, if ever, criminal disenfranchisement is justifiable in a specific case. I begin by outlining deterrence-based and expressivist justifications of criminal disenfranchisement, and showing why such justifications fail. Another historically popular justification of criminal disenfranchisement is the positive retributivist view that disenfranchisement is a *fitting* response for a wide range of serious crimes, such as murder or assault. I argue that positive retributivism ultimately fails to provide persuasive justifications for defining the desert basis for disenfranchisement as broadly as it currently is. While positive retributivists are right that fittingness *matters* for

[9] This does not imply that in a state in which anyone is disenfranchised, *no-one's* liberty is secured: rather, departures from the presumption in favour of political inclusion warrant especially careful justification so as not to endanger the equal liberty of democratic citizens.

[10] Of course, given that other forms of punishment like incarceration pose significant burdens on individual liberty, they may not be defensible in their current form either. Much like there is a presumption in favour of universal suffrage, there is a presumption in favour of ending mass incarceration. So while there is a *special* presumption against disenfranchisement, it is not obvious that there is *no* presumption against other forms of punishment like incarceration.

justifiable punishment, it does not follow that disenfranchisement is a punishment which plausibly fits *all* serious crimes. Furthermore, a purely retributivist account relying solely on a desert-based fittingness argument is an insufficient *positive* justification for why we ought to disenfranchise particular criminals: given the special moral weight of the democratic presumption in favour of universal suffrage, and the morally costly consequences of disenfranchisement, a supplementary positive argument—compatible with the original presumption—is required. In other words, desert is necessary, but not sufficient for justifying disenfranchisement. Building on structurally similar negative retributivist contributions to the philosophy of punishment, I develop a negative retributivist account of disenfranchisement, combining a novel conceptual account of the appropriate desert basis for disenfranchisement with a positive instrumentalist justifiability condition. I argue the relevant kinds of wrongs for disenfranchisement are *political wrongs*, a very narrow subset of the set of public wrongs (i.e. crimes). Since no contributor to this debate has done so thus far, I give a systematic account of political wrongdoing. I then show why only political wrongdoers temporarily forfeit their right not to be punished via disenfranchisement, and why disenfranchising political wrongdoers fulfils both the negative retributivist desert criterion and the instrumentalist justifiability condition, according to which disenfranchisement is justified if and because it deters and temporarily incapacitates political wrongdoers, which in turn ultimately helps to protect and enforce the presumption in favour of universal suffrage of all democratic subjects as equals.

I. The Non-Retributivist Case for Criminal Disenfranchisement, and Why It Fails

Constitutional and criminal law scholars have cast significant doubt on the notion that criminal disenfranchisement is justifiable given existing laws and constitutional provisions. In the US, criminal disenfranchisement has been subject to scrutiny[11] because of its possible violation of the 14th amendment's Equal Protection Clause,[12] the 8th amendment's ban of cruel and unusual punishment, and various laws prohibiting disenfranchisement on racial grounds, such as the 15th amendment and the US Voting

[11] The relevant body of scholarship has far earlier origins than one might expect: for instance, several contributions to the very first issue of the APSR (1906) considered the role of racially motivated discriminatory intent for criminal disenfranchisement policies. See e.g. John C. Rose, "Negro Suffrage: The Constitutional Point of View," *American Political Science Review* 1 (1906): 17-43.

[12] Legal scholarship in this area has mainly reacted to the US Supreme Court's ruling in the landmark case *Richardson v. Ramirez* 418 U.S. 24 (1974), which stated that felon disenfranchisement does not violate the 14th amendment if voting restrictions are necessary due to a "compelling state interest". For an influential critique, see Karlan, "Convictions and Doubts," 1147-70.

Rights Act (1965). While clashes between empirical practices of disenfranchisement and contemporary legal provisions do carry moral significance, they do not decisively settle the *normative* question of which theory of punishment can plausibly justify criminal disenfranchisement, irrespective of current empirical practices. Here, I discuss possible normative justifications of criminal disenfranchisement based on dominant theories of punishment *in general*: deterrence accounts, expressivist accounts, and—in a separate section—retributivist accounts.

1. Deterrence Accounts and the Problem of Democratic Incentives

Deterrence theories of punishment claim that punishment is justified only if it reduces the likelihood of potential offenders actually offending.[13] The point of punishment, then, is forward-looking crime reduction. To justify criminal disenfranchisement, deterrence theorists must show that it will in fact serve these aims. In principle, deterrence views are compatible with punishing the innocent, and with disproportionately punishing the guilty.[14] But deterrence theorists can respond to this problem by arguing that well-functioning democratic societies have protections against unjust punishments.[15] While this response may be persuasive with respect to constructing a deterrence justification of punishment in general, it fails in the specific context of justifying criminal disenfranchisement. Why is this so?

> **No Incentives Objection**: According to the deterrence theorist, protections from unjust punishment rely on interventions by elected public officials. If democratic states disenfranchise some incarcerated members (including some innocent ones), those members cannot appeal their own disenfranchisement, or lobby for protections against unjust punishment. Elected officials have no incentive to consider the interests of the unjustly punished in having protections against unjust punishment, since the latter will not be able to penalise them at the ballot box anyway. This applies even to cases of temporary disenfranchisement, if the term of incarceration with which disenfranchisement is contemporaneous coincides with federal, state,

[13] Examples of recent accounts of deterrence theory include Anthony Ellis, "A Deterrence Theory of Punishment," *Philosophical Quarterly* 53 (2003): 337-51; Zachary Hoskins, "Deterrent Punishment and Respect for Persons," *Ohio State Journal of Criminal Law* 8 (2011): 369-84; Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011).

[14] David Boonin, *The Problem of Punishment* (Cambridge: Cambridge University Press, 2008), 41-52; Igor Primoratz, *Justifying Legal Punishment* (Atlantic Highlands, NJ: Humanities Press, 1989), 33-72; William D. Ross, *The Right and The Good* (Oxford: Clarendon Press, 1930), 56-64.

[15] For similar arguments, see John Rawls, "Two Concepts of Rules," *The Philosophical Review* 64 (1955): 3-32; John Braithwaite and Philip Pettit, *Not Just Deserts* (Oxford: Oxford University Press, 1990). See also Tadros, *The Ends of Harm*, 42, stating explicitly that one of the central strengths of his constrained deterrence account is its ostensible ability to avoid punishing the innocent. But see Patrick Tomlin, "Innocence Lost: A Problem for Punishment as Duty," *Law and Philosophy* 36 (2017): 225-54 for an argument (which succeeds, in my view) about why the most plausible version of Tadros's account is in fact compatible with punishing the innocent.

> and local elections as well as with other opportunities to exercise democratic participation rights.

Deterrence theorists may respond that other members of the democratic state—the enfranchised—can pressure elected officials to create and uphold protections against unjust punishments. Doing so would be in their self-interest, given that they themselves may be unjustly disenfranchised in the future. However, this response tells us nothing about why unjust punishments are morally permissible. Instead, it simply acknowledges that unjust punishments are a moral problem, and then states that it is possible (though not certain) to rectify the problem during the democratic process. A theory of punishment that excludes the possibility of punishing the innocent from the start would be more appealing. But such a theory cannot be a pure deterrence theory, for it would have to introduce additional principles justifying why unjust punishment is intrinsically wrong in all cases. This kind of judgment is not possible if a theory of punishment makes the justification for criminal disenfranchisement contingent on its context-dependent effects, and not on the context-independent wrongness of punishing innocents.

2. Expressivism and the Problem of Self-Contradiction

The possession of voting rights and other democratic participation rights carries symbolic significance. As Bennett and Viehoff put it,

> [t]he democratic presumption in favour of enfranchisement reflects [a] non-outcome-related [...] value [...]. The right to vote [...] identifies its bearer as someone who counts. [...B]y enfranchising someone we *publicly affirm* that her interests are *worthy* of the protection that the right to vote provides.[16]

Even if disenfranchisement's effect on morally weighty interests is trivial, it may constitute a non-trivial communicative wrong against the rights-holder. But those who endorse expressivist justifications of criminal disenfranchisement[17] may insist that the expressivist wrong truly at stake goes in the opposite direction: enfranchising criminals communicates disrespect to the law-abiding demos members. Therefore, disenfranchising criminals is justified because it expresses a democratic polity's commitment to its own laws.[18]

---

[16] Christopher Bennett and Daniel Viehoff, "Written evidence for the UK's Joint Committee on the draft Voting Eligibility (Prisoners) Bill," https://www.parliament.uk/documents/joint-committees/Draft-Voting-Eligibility-Prisoners-Bill/prisonervoting-evidvol%20(4).pdf, 97 [my emphases].

[17] For expressivist views, see Joel Feinberg, "The Expressive Function of Punishment," *The Monist* 49 (1965): 397-423. For other versions of expressivism (also known as 'communicative theories of punishment') see Elizabeth S. Anderson and Richard H. Pildes, "Expressive Theories of Law: A General Restatement," *University of Pennsylvania Law Review* 148 (2000): 1503-75; R. A. Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001).

[18] For a critical account, see López-Guerra, *Democracy and Disenfranchisement*, 116-7.

However—first—criminal disenfranchisement is not *necessary* to achieve expressivist aims: we can communicate our disapproval of criminal behaviour in other ways which do not involve the infringement of key political rights, for instance, during sentencing. Other kinds of punishment, like incarceration, can express the same message. Expressivists would have to rely on additional criteria to show why *criminal disenfranchisement in particular*—as opposed to other kinds of punishment—is necessary for communicating disapproval. Second, criminal disenfranchisement is not *sufficient* for achieving expressivist aims in all cases. After all, the expressivist message may get 'lost in translation'—for instance, in societies facing pervasive disagreement about the justness of existing laws, and the duty to expressively affirm them as a collective. If an offender does not (want to) understand the expressivist message, this poses a problem for expressivism, unless expressivists can show that affirming democratic values *specifically by disenfranchising criminals* has *intrinsic* value.[19]

Expressivists may respond that they need neither claim that disenfranchisement is necessary nor that it is sufficient, but that disenfranchisement is simply an *appropriate* way of expressing the expectation that full participation rights are conditional on respecting other citizens' rights. Temporary criminal disenfranchisement, much like temporary incarceration, would then simply be part of a punitive package with one expressivist purpose: communicating the insistence that this standard be met.[20] The *Standards View* is a *declarative* account of expressivism, rather than an *instructive* one geared towards the moral education of criminals.[21] But we cannot consider the illocutionary act of declaring standards for enfranchisement in isolation from its perlocutionary effects in a given social context. Grant that democratic states have a right to engage in the illocutionary act of declaring standards, and that one such standard is non-criminality. This stronger version of expressivism is not vulnerable to the objections above, and thus seems intuitively plausible under (near-)ideal circumstances. However, in a non-ideal world, disenfranchisement as a response to the failure to meet such standards involves expressing condemnation, and thus stigmatising, criminal offenders in a way that often disproportionately affects persons of

[19] One might think that it ultimately does not matter whether the offender understands the expressivist message, since the real addressee is the polity as a whole. However, analogous problems arise: if significant disagreement persists in society about the justness of particular laws and punishments, the purpose and content of the expressivist message may be lost on a large part of the democratic constituency.

[20] Andrew Altman proposes a similar view in "Democratic Self-Determination and the Disenfranchisement of Felons," *Journal of Applied Philosophy* 22 (2005): 265. Note that Altman's view does not suggest that criminal disenfranchisement as an insistence on certain standards being met is morally *required,* nor does it suggest that criminal disenfranchisement is necessarily a matter of *punishment*. Instead, Altman's view is that disenfranchisement is merely a morally permissible decision for democratic constituencies to make given their right of self-determination.

[21] This terminology relies on Andrei Poama and Tom Theuns, "Making Offenders Vote: Democratic Expressivism and Compulsory Criminal Voting," *American Political Science Review* 113 (2019): 796–809.

colour. In the specific non-ideal context of the contemporary US, this 'badge of dishonor' has a distinctive historical connotation tied to the Jim Crow era. Given this, disenfranchisement with racially disproportionate effects constitutes a culturally and socially specific expressive harm.[22] For expressivists, this should count as an overriding normative reason, *on expressivist grounds*, not to disenfranchise criminals, because it would be morally repugnant to insist on exercising a declarative, standard-setting right if doing so leads to the imposition of severe, unequally distributed expressive harms.

Expressivists may make another claim about disenfranchisement: rather than simply expressing a demos's judgment, it may serve the aim of morally educating or motivating criminals to abide by a polity's laws. Sigler claims that criminal disenfranchisement may "heighten offenders' sense of civic responsibility by establishing the expectation of restored political participation",[23] and Hampton states that "the suspension of voting rights [may] change [...] the wrongdoers' way of thinking about himself and society".[24] Even if an offender does not want to receive the expressivist message of disapproval, expressivists hope that temporarily losing participation rights will promote an offender's civic virtue. Yet aside from the fact that this claim is implausibly speculative about the actual empirical consequences of disenfranchisement, it reveals a deeper normative tension. On the one hand, expressivists are committed to the view that the collective *exercise* of democratic participation rights educates people, and instils civic virtue in them because participation constitutes an expressive re-affirmation of democratic values.[25] On the other hand, by endorsing disenfranchisement, expressivists seem to assume that it is exactly the deprivation of participation rights that instils civic virtue in the demos's members. Which is it? Given these internal tensions, we should reject the expressivist justification of criminal disenfranchisement. In fact, the expressivist case *for* enfranchising criminals is

[22] It is beyond the scope of this paper to develop a complete conceptual account of expressive harms of this kind. For further discussion see Richard Pildes and Richard Niemi, "Expressive Harms, 'Bizarre Districts,' and Voting Rights: Evaluating Election-District Appearances after Shaw v. Reno," *Michigan Law Review* 92 (1993): 483–587.

[23] Mary Sigler, "Defensible Disenfranchisement," *Iowa Law Review* 99 (2014): 1728.

[24] Jean Hampton, "Punishment, Feminism, and Political Identity: A Case Study in the Expressive Meaning of the Law," *The Canadian Journal of Law & Jurisprudence* 11 (1998): 43.

[25] Jean Hampton, "The Moral Education Theory of Punishment," *Philosophy & Public Affairs* 13 (1984): 40-44. This expressivist commitment coheres with the presumption in favour of universal suffrage, and thus with the view that political inclusion has an especially high value in comparison to other important values in a democratic state, such as individual liberty rights. Given this, expressivists can coherently defend incarceration while rejecting disenfranchisement: the internal tension in the expressivist justification of disenfranchisement does not arise in the same way for expressivist justifications of incarceration without disenfranchisement, since expressivists can plausibly argue that temporary incarceration fulfils the expressivist aim of instilling civic virtue in offenders while *at the same time* allowing offenders to expressively reaffirm their own commitment to democratic values by exercising participation rights. By *not* disenfranchising incarcerated criminals, democratic states can publicly express their commitment to the value of citizens' equal liberty.

stronger than for disenfranchising them, if doing so would allow criminals to join the rest of the demos members in expressing their commitment to democratic values.

II. The Retributivist Challenge

1. Positive Retributivism and the Fittingness Problem

Positive (or 'pure') retributivists are committed to two basic assumptions: the guilty *deserve* to be punished in a way that *fits* their crime in a proportionate way, and no moral consideration relevant to punishment *outweighs* the offender's criminal desert.[26]

Positive retributivists face the difficult problem of identifying what it is about an act of criminal wrongdoing that makes disenfranchisement *proportionate*, to explain why the criminal *deserves* disenfranchisement instead of, or in addition to, other available punitive sanctions like incarceration. The ECtHR ruling in *Hirst v. UK* discusses this problem explicitly: "the principle of proportionality requires a *discernible and sufficient link* between the sanction and the conduct and circumstances of the individual concerned".[27]

One way of justifying why criminal disenfranchisement is a fitting response to criminal wrongdoing is to argue that criminals have, by breaking the law, shown that they are *morally*—though not necessarily rationally—*unfit* to participate politically. In the US context, the most influential statement of the moral fitness view which is still often invoked today is the 1884 ruling of the Alabama Supreme Court in *Washington v. Alabama*, according to which the aim of disenfranchisement is to

> preserve the *purity of the ballot box*, which is the only sure foundation of republican liberty [...] one rendered *infamous* by conviction of felony [...] is *unfit* for the *privilege* of suffrage, or to hold office, upon terms of equality with freemen who are clothed by the State with the toga of political citizenship.[28]

Contemporary versions of this argument rely on the intuition that committing a serious criminal offence is sufficient for showing that a criminal is morally unfit to participate politically in society. If a person wilfully or recklessly breaks democratically made laws, the intuition goes, it is fitting to (temporarily or permanently) revoke her right to participate in

---

[26] Michael S. Moore, "The Moral Worth of Retribution," in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology,* ed. Ferdinand D. Schoeman, (Cambridge: Cambridge University Press, 1987), 179-219; Morris, "Persons and Punishment", 475-501; Andrew von Hirsch, *Doing Justice: The Choice of Punishments* (New York: Hill & Wang, 1976). Not all positive retributivists are desert theorists, but I focus on this dominant type of positive retributivism here.

[27] *Hirst v. UK,* §71.

[28] *Washington v. State*, 75 Ala. 582, 585 (1884) [emphases added].

the democratic making of laws.[29] Even if her *primary* intention was not to express disregard for the value of democratically made laws, she has in effect *acted* out of disregard for that value. On this view, a criminal deserves temporary suspension from participating as a co-deliberator because of her failure to willingly subject herself to the rules for regulating society. Importantly, what motivates this version of the fittingness intuition is *not* the severity of *harm* inflicted by perpetrators of criminal wrongdoing, such as the harm inflicted by murder or assault, but the wrong of breaking a law that has been democratically agreed upon: criminal disenfranchisement is *fitting*, on this view, not primarily because the crime in question is particularly morally serious and because disenfranchisement is a weighty punitive sanction on par with the crime's seriousness, but because disenfranchisement temporarily deprives the criminal of opportunities to co-author the system of laws contrary to which she has acted, and thus of opportunities to exercise power, to some extent, over what *other* members of society may (not) do.

But it is not obvious why we should accept this claim. First, not all criminal wrongdoing is morally objectionable to an extent that sufficiently demonstrates a moral unfitness to participate in *all* democratic decisions during the duration of their punishment, rather than a specific instance (if at all) of moral wrongdoing: consider, for instance, petty theft out of necessity. But even in cases of morally serious criminal wrongdoing, such as murder, it is unclear how the fact that a criminal has violated one democratically made law substantiates the view that the criminal has in effect acted in disregard of *all* democratically made laws. A murderer, for instance, might genuinely believe in the rule of law's legitimacy as a whole, and adhere to all laws unrelated to murder. By this, I do not mean to suggest that criminal wrongdoing like murder is *not morally serious*. But acting out of disregard for one democratically made law and acting out of disregard for all democratic laws is qualitatively different: we cannot necessarily infer the latter from the former. Intuitions about fittingness (and thus, desert) must be informed in part by our moral judgments about *how close* the fittingness link between desert basis and desert object must be. This judgment will depend on which specific underlying values are at stake in light of the imposition of desert-based punishment. In the case of disenfranchisement, this would be the value of political inclusion in accordance with the presumption in favour of universal suffrage, which—as I have argued—is particularly high for instrumental reasons: its full realisation is important for the protection of other central democratic values, such as the equal liberty of citizens realised through the joint exercise of democratic control. Thus, the justificatory

[29] This is commonly taken to imply that positive retributivism points towards a *broadly defined desert basis* for criminal disenfranchisement, which extends to *all* perpetrators of serious criminal wrongdoing.

burden for disenfranchisement is both *very high* and *very specific*, which is why it is not surprising that only a smaller group of criminals will meet this justificatory threshold, in comparison to the group of criminals who deserve punishments like fines and incarceration. Morally serious crimes, which are not *primarily* wrong *because* they undermine others's ability to meaningfully exercise democratic control within robust democratic institutions, provide an *insufficiently close* desert basis for disenfranchisement, given the particularly high value at stake.

Second, to contrast, consider a person who never commits any crimes, but nevertheless holds a disregard for the system of democratically made laws. She might not violate existing laws, but from her perspective, the legitimacy of democratically made laws may not constitute a *reason* for her to act in a certain way. Thus the moral fitness argument does not succeed in plausibly capturing the cases it is trying to capture, and it is both under- and overinclusive because the cases it does capture are the wrong cases.

Some positive retributivists combine the moral fitness argument with a claim about fairness: as Herbert Morris argues, "[i]f a person fails to exercise self-restraint even though he might have [...] he renounces a burden which others have voluntarily assumed and thus gains an advantage which others, who have restrained themselves, do not possess",[30] in which case it would be unfair not to punish the criminal. The *combined* argument from moral fitness and fairness assumes that all criminals derive benefits from lawbreaking. The relevant 'advantage' would be the fact that the criminal has allowed herself to commit a crime, usually (though not always) for self-interested reasons, which other demos members have—on this basis—restrained themselves from committing. If demos members have collectively agreed to restrain themselves, the argument goes, it would be unfair of criminals to renege on this democratic agreement.

However, the notion that criminals derive unfair advantages from criminal wrongdoing, and that this justifies the imposition of punishment—such as criminal disenfranchisement—faces serious problems: not all crimes lead to unfair net benefits for the criminal, and while many crimes are committed with the specific *intent* of deriving benefits, not all of them—including morally serious ones—are, for instance those involving ignorance or negligence. While the sentence in those latter cases may be more lenient, the offence may still be serious enough to result in a conviction that in practice triggers disenfranchisement.

[30] Morris, "Persons and Punishment," 477.

But more importantly, I doubt that the fairness argument helps justify criminal disenfranchisement in the way that positive retributivists might think.[31] Criminal disenfranchisement comes with the significant *democratic cost* of making it impossible for prisoners to have a say in how democratic decisions affect their interests—not only general democratic decisions, but also specific decisions affecting them more directly, such as the use of public funds towards improving prison conditions.[32] Furthermore, disenfranchisement *itself* seems unfair especially in cases where persons reasonably disagree about the criminal law's content and scope. If what constitutes a crime, and why, is subject to democratic debate, then democratic societies must be particularly careful about depriving lawbreakers of opportunites to voice disagreement.[33] Contrary to Morris's view, blanket *dis*enfranchisement—even if temporary—thus seems intuitively unfair, not enfranchisement. By contrast, incarceration (without simultaneous disenfranchisement) is not unfair in the same way, since it would still allow incarcerated criminals to engage in democratic deliberations about the status of (non-) criminal acts subject to reasonable disagreement. Importantly, the distinct perspective of currently and previously incarcerated individuals, which likely differs from the perspectives of non-incarcerated citizens, may *enrich* such deliberations. Ultimately, positive retributivism fails to articulate a plausibly constrained desert basis for disenfranchisement.

2. The Negative Retributivist Alternative

Negative retributivists like myself reject the standard ('positive', or 'pure') retributivist view that desert provides a *sufficient positive reason to punish* the guilty.[34] Instead, negative retributivism imposes *constraints* on a particular justification of punishment (in this case, criminal disenfranchisement): we should disenfranchise only those criminals who deserve it, and only in proportion with their respective desert.[35] Negative retributivism thus articulates the *lack* of a desert basis: when have persons *not*

[31] A commitment to fairness as a side constraint for punishment does, however, help determine the correct scope of disenfranchisement once it has been justified—see Part IV.
[32] On this point see Kleinig and Murtagh, "Disenfranchising Felons," 229.
[33] Obviously, some crimes are *uncontroversially* morally wrongful, such as murder. But the point still holds that when the morally wrongfulness of some crimes *is* controversial, we have strong fairness-based reasons to be cautious about disenfranchisement, given that disenfranchisement would prevent the participation of a subset of people in democratic deliberations concerning the wrongfulness of controversially morally wrongful crimes.
[34] John D. Mabbott, "Punishment," *Mind* 48 (1939): 150-67; Anthony M. Quinton, "On Punishment," *Analysis* 14 (1954): 133-42; Richard L. Lippke, "Some Surprising Implications of Negative Retributivism," *Journal of Applied Philosophy* 31 (2014): 49-62.
[35] For a classic definition of negative retributivism see J. L. Mackie, "Retributivism: A test case for ethical objectivity," reprinted in *Philosophy of Law*, eds. Joel Feinberg and Hyman Gross, fourth edition (Belmont CA: Wadsworth, 1991), 677-84. See also Duff, *Punishment, Communication, and Community*, 11-12.

temporarily forfeited their right not to be punished via criminal disenfranchisement? Of course, this does not by itself amount to a full positive justification of why, and when, criminal disenfranchisement is justified.[36] Therefore, negative retributivism must be supplemented by additional, non-retributivist principles, which can accomplish this normative task. This, however, is a strength of the account, not a weakness. Unlike positive retributivism, negative retributivism can incorporate plausible forward-looking *instrumentalist* concerns—such as deterrence, but also incapacitation, or more specifically, the concern for the integrity of the democratic procedures and institutions enabling the protection and enforcement of the rights of democratic citizens in the first place.[37] At the same time, negative retributivism remains immune to my earlier objections to pure deterrence accounts, as negative retributivism is incompatible with punishing the innocent. Similarly, negative retributivism is not internally contradictory like pure expressivist justifications of disenfranchisement, though it can accommodate supplementary expressivist considerations when it comes to providing a positive justification of disenfranchisement. Furthermore, negative retributivism is compatible with *side constraints* introduced specifically to ensure coherence with the higher-order moral concern of protecting the integrity of democratic procedures and institutions, which will be specified later.

But what exactly would a plausible mixed justification for criminal disenfranchisement built on negative retributivist foundations look like? In my view, it must proceed from two claims:

1. *Desert as a necessary but not sufficient condition (Negative Retributivism)*: Deserving punishment is a necessary condition for its overall justification. Those (and only those) guilty of wrongdoing of type X deserve to be punished in proportion to their desert.
2. *Compatibility Claim*: Criminal disenfranchisement is compatible with the presumption in favour of universal suffrage and the higher-order moral goal of the equal protection of the democratic rights of citizens, provided that (i) it adheres to sentencing constraints[38] which do not undermine that goal, and that (ii) it disenfranchises only those members of X who deserve *disenfranchisement* specifically, not simply punishment in general.

[36] For a similar point, see Richard L. Lippke, "Criminal Offenders and Right Forfeiture," *Journal of Social Philosophy* 32 (2001): 78.

[37] This does not mean that I endorse a complete instrumentalist account of punishment, such as standard deterrence theory, which I have rejected above. But I do endorse plausible *aspects* of that theory but not the view that it is sometimes morally justifiable to punish the innocent, and to disproportionately punish the guilty.

[38] These constraints will be specified below.

In what follows, I develop this argument by distinguishing between different types of criminal wrongdoing, and arguing that a special, narrowly defined type of crime is a more plausible desert basis for disenfranchisement: *political wrongdoing*. I show why negative retributivism justifies the criminal disenfranchisement of political wrongdoers—and political wrongdoers only.

III. Rethinking the Desert Basis: Political Wrongdoing

1. What is Political Wrongdoing?

a. Crimes as Public Wrongs: Not All Public Wrongs Are Created Equal

Contemporary practices of criminal disenfranchisement in the UK, the US, and many other states operate on the assumption that the relevant kind of wrongdoing for which we may justifiably disenfranchise offenders is the set of serious crimes (or 'felonies') that *also* justify incarceration. This includes, most obviously, standard *mala in se* crimes like murder, but also *mala prohibita* like the illegal use of (some) drugs. Why disenfranchise such a wide set of criminals? A closer inspection of what crimes *are* is insightful here. As Duff and others have argued, what distinguishes crimes of the kind mentioned above from other kinds of wrongdoing (e.g. torts) is that they are public wrongs: "[w]rongs done to individual members of the community are then wrongs against the whole community [...] insofar as the individual goods which are attacked are goods in terms of which the community identifies and understands itself".[39] This makes crime *everybody's* business.[40] Public wrongs thus necessarily affect the interests of every member of a democratic polity. However, it would be too quick to infer from this conceptual point that the kinds of wrongdoing *warranting criminal disenfranchisement* simply are crimes, i.e. the full set of public wrongs. Why?

While I do not wish to reject the widely shared view that crimes are public wrongs, I do not think that all public wrongs are *primarily* wrong because they wrong the public. To see the force of this intuition, contrast the following two cases:

---

[39] Sandra E. Marshall and R. A. Duff, "Criminalization and Sharing Wrongs," *The Canadian Journal of Law and Jurisprudence* 11 (1998): 20.

[40] For a direct statement of this point, see R. A. Duff, "Responsibility, Citizenship, and Criminal Law," in *Philosophical Foundations of Criminal Law*, eds. R. A. Duff and Stuart P. Green (Oxford: Oxford University Press, 2011), 139: "Public wrongs are our wrongs as citizens—wrongs in which we take a proper *interest*, to which we should collectively respond, for which we claim the right [...] to call the perpetrator to answer to us. Some such wrongs are public in their material impact: if we ask who is wronged the answer is 'the public'; this is true of, for instance, [...] wrongs that threaten public institutions; of frauds committed against the public purse. Others take individuals as their direct victims, but count as 'our' wrongs because they violate our public values, and because we share them with the victim: our concern for the victim as our fellow citizen *makes them our business*." [emphases added]

> ***Jane the Murderer***: Jane murders Gary. Her only reason for murdering Gary is her personal feeling of hatred towards him. Jane is not trying to make a public statement by murdering Gary—in fact, she will try everything possible to avoid becoming a publicly known suspect in the subsequent investigation.

By committing the murder, Jane has not only wronged Gary as an individual, but also the public. But this is not the primary reason why it was wrong for Jane to murder Gary. Murder is in and of itself morally wrong (hence, *malum in se*). Arguably, this is the more obvious reason why we call murder a wrongful act. The additional wrong towards the public is secondary to this initial wrong. The public's interest in punishing murderers is an interest *on behalf of murder victims*: murders are "[public] wrongs that the victim ought to pursue—that it would be wrong to shrug off or ignore".[41] Contrast *Jane the Murderer* to

> ***Jane's Get Out of Jail Free Card***: Jane is a suspect in the investigation of Gary's murder. The prosecutor in Jane's case, Mary, knows about Jane's crime. Jane knows that Mary knows. Jane bribes Mary to lie in court to protect Jane, so that Jane will not be convicted of murder. Mary accepts the bribe. Jane does not get convicted for murdering Gary.

Both public wrongs here differ significantly from the public wrong in *Jane the Murderer*. First, Jane's criminal wrongdoing (bribery of a public official for private gain) is wrong primarily because it obstructs justice, thereby undermining vital features of well-functioning democratic institutions: it is *primarily* wrong because it wrongs the public, and thus differs in this respect from Jane's initial wrongdoing (murder). Second, Mary has committed a cluster of public wrongs (accepting bribes, obstruction of justice, being derelict in her special obligations as a public official), and again, these wrongs are *primarily* directed at the public. Obviously, Mary has also wronged particular individuals: Gary's relatives perhaps, who have an interest in justice being served. But having wronged Gary's relatives individually is arguably not the main reason why Mary's actions were wrongful, as her actions as a public official have larger-scale implications. Mary, by accepting the bribe, would have still wronged the public even if Jane had been convicted despite Mary's lies, such that justice would have been served.

Contrasting *Jane the Murderer* and *Jane's Get Out of Jail Free Card* gives us strong reasons to differentiate between different kinds of public wrongs. If only some public wrongs are *primarily* wrong because they wrong the public, we must ask what exactly defines

[41] Sandra E. Marshall and R. A. Duff, "Public and Private Wrongs," in *Essays in Criminal Law in Honour of Sir Gerald Gordon*, eds. James Chalmers, Fiona Leverick, and Lindsay Farmer, (Edinburgh: Edinburgh University Press, 2010), 83.

that subset of cases—and whether negative retributivism supports disenfranchising those who commit any or all types of public wrongs.

b. Defining Political Wrongs

I define political wrongdoing as the subset of public wrongs which are *primarily* wrong because they wrong the public by (i) wrongfully imposing a non-trivial risk of harm on the integrity of democratic institutions or procedures, *and* (ii) by committing such wrongs with a wrongful intention which makes a modal difference for the potential harm caused (not necessarily by the political wrongdoer herself) should the risk eventuate. Here, wrongful risks of harm (i) and wrongful intentions (ii) are necessary elements of a sufficient set (NESS),[42] where (i) and (ii) are *jointly* sufficient for a public wrong to count as political wrongdoing.

Risks of Harm

I say 'wrongfully imposing a non-trivial risk' because I endorse the view that *some* risk impositions can be wrongful even if the harm never eventuates, provided that the risk of harm's probability and severity is non-trivial.[43] Including '*risks* of harm', rather than 'actual harm' *only*, is particularly important because public wrongs also include inchoate offences, such as attempts to rig elections, to obstruct justice, or to silence or otherwise harm political opponents. Consider

> ***Jane's* Permanent *Get Out of Jail Free Card***: Jane is plotting future murders. Jane hires a helper for the specific purpose of bribing any public official—not just Mary—involved in any future criminal prosecutions of Jane's crimes. Jane has not committed her second murder yet, but she has created a system that increases her chances of successfully obstructing justice. The presence of this system itself constitutes a non-trivial risk of harm to the integrity of democratic procedures and institutions in Jane's state.

To clarify, this example also applies to *failed* attempts. When Jane commits her second murder, Jane's helper's bribe may well fail to induce public officials to act in specific ways: for instance, due to bad luck, incompetence, or unwillingness to carry out Jane's orders in this instance. But Jane's helper will not necessarily fail again when Jane commits her third murder. Jane has created a *system* which continues to threaten and undermine an

[42] For an influential account see Richard W. Wright, "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts," *Iowa Law Review* 73 (1988): 1019.

[43] For defences of this view, see (amongst others) John Oberdiek, *Imposing Risk: A Normative Framework* (Oxford: Oxford University Press, 2017), 84-91; Judith Jarvis Thomson, *Rights, Restitution, and Risk: Essays in Moral Theory* (Cambridge, MA: Harvard University Press, 1986), 181.

institution integral for well-functioning democratic states, and in turn, for states' ability to ensure equal consideration for every citizen's morally weighty interests by enforcing every citizen's equal right to democratic participation.

The point about risks of harm is obviously not restricted to the *specific* political wrong of obstructing justice. Analogous points apply to electioneering (a helper has been hired so that if Jane ever runs for office, the helper rigs the votes in her favour and intimidates her opponents), as well as other types of political wrongdoing.

Importantly, 'risks of harm' already includes 'actual harm'—my earlier point was simply that we should not narrow down the definition to include actual harm only, because many important instances of political wrongdoing involve inchoate offences. For illustrations of *actual* harm resulting from political wrongdoing, consider real cases of successful electoral fraud, such as vote buying, tampering with voting machines, intimidating or attacking people at polling stations, or systematically spreading misinformation about electoral procedures.

Impersonating someone else while voting might be another relevant actual harm. This type of voter fraud is often discussed in the public debate, and in a particularly sensationalist way, in the context of the 2016 US presidential election. Consider, for instance, Donald Trump's statement—reported by CNN on April 5th, 2018, that "[i]n many places the same person in California votes many times [...] Millions and millions of people".[44] But as many academic studies have shown, voter fraud of this kind is extremely rare in the recent US context. For example, as a 2007 report by the Brennan Center for Justice points out, "[i]t is more likely that an individual will be struck by lightning than that he will impersonate another voter at the polls".[45] Therefore, rare individual instances of voter fraud will typically not constitute large-scale political wrongdoing of the kind that I have in mind here.

Other significant examples of political wrongdoing include significant abuses of power by public officials, treason, and the funding of or participation in terrorist activities with the explicit aim of undermining democratic institutions; or the criminal misuse of economic participation rights,[46] such as contributing financially to political campaigns above

---

[44] See Eli Watkins, "Trump Repeats Debunked Voter Fraud Claim," *CNN* (April 5, 2018) https://edition.cnn.com/2018/04/05/politics/trump-voter-fraud-california/index.html

[45] Justin Levitt, "The Truth About Voter Fraud," *Brennan Center for Justice at New York University School of Law*, https://www.brennancenter.org/sites/default/files/legacy/The%20Truth%20About%20Voter%20Fraud.pdf, 4. For a review of empirical studies regarding voter fraud in the US, which all support the view that voter fraud of this type is rare, see the 2014 Government Accountability Office Report to Congressional Requesters, "Elections: Issues Related to State Voter Identification Laws", http://www.gao.gov/assets/670/665966.pdf.

[46] For further discussion see Annette Zimmermann, "Economic Participation Rights and the All-Affected Principle," *Global Justice: Theory Practice Rhetoric* 10 (2017): 1-21.

legal limits, or transferring campaign funds from one campaign to another when the two ought to be legally separate.

Wrongful Intentions[47]

My point about 'wrongful *intentions'* may seem puzzling at first sight. After all, why criminalise wrongful intentions—via criminal disenfranchisement or otherwise? What I have in mind here are intentions coupled with a *plan*, or, in Bratman's terms, a "reflective endorsement".[48] In many paradigmatic cases of political wrongdoing, such a plan might involve thinking about which criminal acts (my own, or others') must occur in order for me to *benefit*. In the context of political wrongdoing, I might plan how I will benefit from electioneering should I, or someone else, gain the power to engage in electioneering. Here, neither I nor anyone else has engaged in electioneering yet. But, critically, I have already endorsed the idea of benefitting from electioneering. Maybe I have planned how to increase my chances, or someone else's, of gaining electioneering opportunities. The point is that I am not merely daydreaming about rigging elections. I have intentionally developed and reflectively endorsed a plan about how I can benefit from electioneering. Such a plan in itself is part of the *actus reus*. As Goodin and Pasternak put it, "the actions that practical intending involves (forming the intention or reflectively endorsing it) [...] are not merely a precursor towards some subsequent 'genuine' doing but rather already part of that doing".[49]

But when, and why, is the intention to benefit already part of the wrongful action, thus making the intention to benefit *wrongful itself*? Call a wrongful intention to benefit ξ. I would argue that intentions to benefit are wrongful, and thus constitute a necessary element of a sufficient set, if and only if the presence of such intentions makes a *difference* for whether a risk of harm, or actual harm, occurs. Consider

> ***Jane for Mayor***: Jane, Mary, and Kate are all running for mayor. Jane is plotting to murder her opponent Mary. Kate is plotting to intimidate Mary so that she drops out of the race. Kate becomes aware of Jane's plans. Kate intends to benefit from Jane's action, and thus fails to warn Mary of her imminent murder. In fact, Kate checks routinely whether Mary has been murdered yet, which counts as a reflective endorsement of her intention to benefit.

[47] Due to this restriction, the concept of political wrongdoing *excludes* acts of disobedience to the law with a *non-wrongful* intention, such as exercising one's moral right to engage in civil disobedience with the aim of upholding justice.
[48] Michael E. Bratman, *Structures of Agency: Essays* (Oxford: Oxford University Press, 2007), 22-5 and 33-46.
[49] Robert E. Goodin and Avia Pasternak, "Intending to Benefit from Wrongdoing," *Politics, Philosophy & Economics* 15 (2016): 285.

Kate has never formed the wrongful intention φ to murder Mary herself. However, the fact that she has formed the wrongful intention ξ to benefit from Mary's murder makes a difference to the situation. Ordinarily, a third agent who gains knowledge of a murder plot would attempt to prevent the murder, for example by sharing their knowledge with law enforcement, or by warning the prospective victim. And in fact, Kate might have done so before she formed intention ξ. But Kate is not an ordinary third agent. The fact that Kate formed ξ makes a difference for whether actual harm to Mary will occur, in comparison to a scenario in which Kate does not have ξ. As Goodin and Pasternak explain, when ξ is a NESS condition of φ, "ξ makes a *modal difference*. It makes it possible for wrongdoing φ to occur [...]. That makes the [holder of ξ] potentially causally responsible for wrongdoing φ occurring and morally responsible for that".[50]

Apart from illustrating this point about wrongful intentions (ξ) making a modal difference, *Jane for Mayor* also shows that the *nature*—not only the wrongfulness—of intentions matters for whether an act is an ordinary public wrong, or indeed a political wrong. Earlier, I used *Jane the Murderer* to illustrate exactly this distinction, arguing that standard *mala in se* crimes like murder are not primarily political. However, in this case, murder becomes part of political wrongdoing, such that the *primary* aim of the actus reus in *Jane for Mayor* is to circumvent fair democratic procedures.

Note that on my account, the class of ξ is not *restricted* to wrongful intentions to *benefit*, as long as the intention itself is wrongful *and* makes a modal difference for an associated wrongful act φ. An example might be an anarchist terrorist sabotaging democratic procedures and institutions with the sole aim of creating chaos. Here, the wrongful intention is vandalistic malice, rather than benefitting personally from their wrongful act, and it is plausible to assume that this wrongful intention is a difference-maker for a wrongful act φ (sabotaging the integrity of democratic institutions) much like a wrongful intention to benefit ξ would be a difference-maker. Thus, while wrongful difference-making intentions are a necessary element of a sufficient set of wrong-making features on this account, wrongful difference-making intentions *to benefit* are worse-makers, not wrong-makers: they are aggravating but not necessary features of political wrongdoing.[51]

[50] Goodin and Pasternak, "Intending to Benefit from Wrongdoing," 286.

[51] In what follows, I focus again on cases involving wrongful intentions *to benefit*, since such cases are a particularly good characterization of *typical* acts of political wrongdoing, yet this does not mean that my conceptual account is restricted to that subset of intentions.

One definitional aspect remains: I say 'harm caused (not necessarily by the political wrongdoer herself)' because political wrongdoing often involves joint enterprises, or complicity in a larger scheme, where an individual political wrongdoer does not necessarily plan to execute the act leading to wrongful harm herself. Not all agents involved in political wrongdoing need to be directly causally responsible for its resulting harms, but they all intend to benefit from such acts. Consider

> ***Jane's Hitman***: Jane is running for mayor. She hires a hitman to murder her political opponent Mary. Jane's hitman kills Mary.

Both Jane and the hitman had wrongful intentions to benefit here, although only the hitman actually murdered Mary: the hitman anticipated a material reward for his own actions, and Jane anticipated a certain win. Note that on my account, it is strictly speaking not necessary that the political wrongdoer *intends to do wrong* (φ), either by committing the wrong herself or by working with other wrongdoers. It is merely necessary that she intends to benefit from any present or future political wrongdoing (φ), even if she did not intend to bring about the original wrong φ. My claim is that the wrongful intention to benefit (ξ) is conceptually distinct from the wrongful intention to commit a wrong (φ), and that, much like φ is wrong in and of itself, irrespective of the actual consequences, ξ is also wrongful in and of itself.[52]

My claim about ξ might be less intuitive than my claim about φ. To see how this thought plausibly applies to political wrongdoing in particular, consider

> ***Big League Conspiracy***: Officials of state R conspire with the staff of a presidential campaign team T to influence an upcoming election in the US by disseminating false information online. Meanwhile, officials of a third country S contemplate the possibility of conspiring with the same team T for the same reasons, since the political interests of R and S are aligned in this case: both want the candidate of team T to win. However, R is unaware of S's activities. S, in turn, *is* aware of R's activities. S therefore knows that, because of R's plans, S will likely achieve its aim of T winning the election. S is willing to take the small chance that R may be unsuccessful in its endeavour. Thus, S abandons the intention to take any action to support T.

In *Conspiracy*, both R and S are political wrongdoers. However, R intends to do wrong (φ) *and* possesses the wrongful intention to benefit (ξ) from wrongdoing (φ); whereas S possesses *only* the latter intention (ξ). The latter suffices for fulfilling our 'wrongful intentions' criterion, because S *would have* supplemented an intention to benefit (ξ) with an intention to do wrong (φ) had R suddenly dropped out of the picture. Political

[52] For a related point, see Robert Fullinwider, "Preferential Hiring and Compensation," *Social Theory & Practice* 3 (1975): 318: it is "wrong to seek out benefits from the wrongdoing of others, even if there is nothing wrong with unavoidably coming into possession of them".

wrongdoing was unavoidable in this case in the sense that either R or S would have carried it out—it was not necessary for them to act *together*. What is interesting about *Conspiracy*, however, is that one agent's knowledge about the presence of wrongful intentions (φ) altered their own intentions to commit wrongdoing (φ), but not their wrongful intentions to benefit from wrongdoing (ξ).

IV. Negative Retributivism In Context: Constraints and Limitations

1. Developing the Negative Retributivist Argument

Why does negative retributivism justify the criminal disenfranchisement of political wrongdoers, and political wrongdoers only? Recall the first two parts of the negative retributivist argument—the restriction of the desert basis to crimes of type X and the compatibility claim—which I outlined earlier. Based on my account of the concept of political wrongdoing, we can now add:

> 3. *Relevant Types of Wrongs*: X denotes the domain of *criminal* wrongdoing. Crimes are public wrongs. But not all public wrongs are morally wrong *primarily* because they wrong the public, and specifically, the integrity of democratic institutions or procedures. Only political wrongdoing, a *subset* of public wrongs, is primarily wrong for this reason.

Now, any negative retributivist justification of criminal disenfranchisement must introduce a (lack of) desert claim and a claim about temporary rights forfeiture:

> 4. *Negative Retributivist Lack of Desert Claim*: Wrongdoers guilty of public wrongs (crimes), but not of political wrongs, do not deserve to be punished by being temporarily disenfranchised (though they may deserve other kinds of punishment).
> 5. *Negative Retributivist Temporary Forfeiture Claim*: Political wrongdoers temporarily forfeit their right not to be punished via criminal disenfranchisement.[53]

By drawing on higher-order democratic principles, negative retributivists can give substance to these claims. While negative retributivists—or any retributivists, for that matter—need not commit themselves to the claim that criminals necessarily temporarily forfeit exactly the right that they have violated, they must provide an argument to the effect that it would not be *fitting* to restrict some or all of the democratic participation rights of a criminal who is

[53] Although an ordinary language usage of the term 'forfeiture' might intuitively suggest a permanent loss of rights, the term is used more broadly in the relevant philosophical literature so as to include temporary losses of rights. See Christopher Heath Wellman, "The Rights Forfeiture Theory of Punishment," *Ethics* 122 (2012): 386.

not a political wrongdoer. The intuition underpinning such a 'lack of desert' claim is the following. If A's crime had nothing do with democratic participation, did not inhibit the democratic participation of others, and was not pursued with the primary intention of eroding central democratic procedures and institutions, and possibly—though not necessarily—benefitting personally from doing so, it would not be fitting to punish A by depriving her of democratic participation rights, the exercise of which is entirely unrelated to her criminal conduct. Otherwise, the connection between the substantive nature of the sanction and the nature of the criminal conduct at stake would simply be too tenuous. Those insisting that A deserves criminal disenfranchisement must explain why disenfranchisement *specifically* is appropriate in this case. Why *not* (or not only) incarcerate A, fine A, or force A to do community service? Why *not* (or not only) publicly shame her? Deprive her of the right to buy and own guns, or the right to take up employment in the medical or educational sector? As López-Guerra argues, "[i]t would be difficult to argue that a person deserves to be deprived of the right to influence the making of all laws just for breaking one, or a few of them".[54] To this, negative retributivists like myself will add: "... unless the criminal has broken laws concerning the integrity of democratic institutions, which ensure that *everyone else* can influence the making of all laws during legitimate democratic procedures, and that everyone's right to do so is protected equally." While criminals who commit public wrongs that are *not* primarily wrong because they wrong the public in this way *lack the kind of desert* necessary for showing that they temporarily forfeit their right not to be punished via criminal disenfranchisement, *political* wrongdoers do not lack this kind of desert. On my account, the desert basis (political wrongdoing) is thus closely linked to the desert object (disenfranchisement) in an intuitively clear way.[55] Negative retributivists must restrict justifiable criminal disenfranchisement to political wrongdoers, and political wrongdoers only.

---

[54] López-Guerra, *Democracy and Disenfranchisement*, 116.

[55] Here, I am relying on influential terminology in Joel Feinberg, *Doing and Deserving* (Princeton, NJ: Princeton University Press, 1970), 55-94, where he defines "[d]esert bases [as follows:] if a person is deserving of some sort of treatment, he must, necessarily, be so *in virtue of* some possessed characteristic or prior activity" (58). Some will of course simply not share my intuition that the connection between non-political criminal wrongdoing and the sanction of criminal disenfranchisement is too tenuous, and that the connection between political wrongdoing and disenfranchisement is sufficiently non-tenuous. Negative retributivism might therefore seem overly restrictive. In that case, the debate about the appropriate desert basis for criminal disenfranchisement leads to an impasse, since evaluations of desert bases must always to some extent rely on intuitionistic reasoning. Ultimately, my discussion here may nevertheless be useful for those denying the intuitive force of my negative retributivist account in the sense that my account helps clarify one amongst several *possible* stances on criminal disenfranchisement, and presents a systematic conceptual analysis of a *possible* desert basis (political wrongdoing), which no other contributor to this debate has to my knowledge done so far. I am indebted to Annie Stilz and Tom Christiano for pushing me to clarify this.

Compare negative retributivism to the positive retributivist stance, which relies on a positive desert claim to justify the disenfranchisement of criminals irrespective of the precise nature (including the intentions and possible harms) of their wrongdoing. Positive retributivists could incorporate my concept of political wrongdoing into their account, and then introduce additional normative principles to restrict their account to political wrongdoers only. This might yield similar results—it would justify the criminal disenfranchisement of the same group of people—but it would do so for different—and in my view, *wrong*—reasons: as I have argued above, positive retributivism is not the most plausible account available for justifying criminal disenfranchisement, for reasons unrelated to restricting the account's scope. Thus, I continue to endorse *negative* retributivism *specifically* here. Analogous arguments apply to the expressivist account: due to the internal tensions in the expressivist justification *of the specific punitive sanction of criminal disenfranchisement*, which would persist even if the account were restricted to political wrongdoers, expressivism remains less plausible than my negative retributivist justification for criminal disenfranchisement.

Having reached the preliminary conclusion that some criminals—political wrongdoers—temporarily forfeit their right not to be punished via criminal disenfranchisement, we still have to show why the criminal disenfranchisement of political wrongdoers is *justified*.[56] Such a justification must be compatible with the morally weighty democratic presumption in favour of universal suffrage: any disenfranchisement policy's purpose must be to ultimately enable the enforcement of that presumption, which is why the justification must be restricted to incapacitating and deterring political wrongdoers in pursuit of the goal of protecting the integrity of the democratic institutions enabling equal democratic participation in the first place.

6. *Positive Instrumentalist Justification for Disenfranchisement*:[57]

    a. Protecting the integrity of democratic institutions and procedures is a valuable goal because doing so is important for securing the equal rights of democratic citizens.
    b. Criminal disenfranchisement by definition requires the restriction of the right to participate in democratic decisions.

[56] Note that this is not a claim about *all-things-considered* justifiability, given that additional constraints will be specified below.

[57] Many other contributors to the debate on negative retributivism agree that combining retributivist elements with instrumentalist ones is a plausible strategy. For an overview of accounts that take this stance, see Mitchell N. Berman, "Two Kinds of Retributivism," in *Philosophical Foundations of Criminal Law*, eds. R. A. Duff and Stuart P. Green (Oxford: Oxford University Press, 2013), 433-59.

c. Such restrictions are justified iff, and because, they serve the aim of protecting the integrity of democratic institutions or procedures.

Of course, not all political wrongdoers will be effectively deterred, and temporarily incapacitated, by criminal disenfranchisement: an election-rigging public official will continue to rig elections even if she loses her individual right to vote. This is why (c) is phrased conditionally. Suppose now that our election-rigging official can only rig elections because of her privileged access to voting machines and administrative procedures related to the electoral process: simply *being a public official* is a constitutive element of her election-rigging efforts. Here, disenfranchising her by revoking democratic participation rights *other than the right to vote*, such as her right to remain in office, and perhaps her right to run for office again, effectively incapacitates her while the punitive sanction of disenfranchisement is in place, and it may deter her and other future political wrongdoers from rigging elections.

A sceptical reader may ask: if citizens' liberty rights are conditional on their rights-respecting conduct, why are citizens' democratic participation rights not similarly conditional? Must the fittingness link be similarly close when we justify incarceration, given that individual liberty's value is also rather high, much like the value of political inclusion? The view that there must be a tight fittingness link may seem implausibly restrictive, given that it may imply that only those criminals who have infringed on another person's liberty deserve a punitive restriction of their own liberty. But most would agree that we need an account of punishment that explains why incarcerating a wide range of serious criminals is permissible. Note that just because the imposition of punishments like disenfranchisement and incarceration is *similarly* conditional, this does not imply that the fittingness link must be *equally tight* in both cases, such that we would have to incarcerate and disenfranchise all serious criminals as part of a single punitive package. The lines can and should be drawn differently for each type of punishment, depending on the nature of the criminal conduct at stake in each case, and conditional on independent normative criteria going beyond a similarity claim. One such criterion is the moral requirement not to impose unnecessary, excessive punishments. In the case of incarceration, that criterion might be the effective incapacitation of criminals, in which case incarceration would not be unnecessary and excessive. By contrast, disenfranchisement would be unnecessary for serious criminals not guilty of political wrongdoing, who have not endangered the integrity of democratic institutions in a sufficiently encompassing way. Local rights violations, even morally serious ones like murder, do not have this significant effect of endangering the entire democratic

system as such. Negative retributivism, then, is not an eye-for-an-eye view: it does not demand that we deprive each criminal of exactly the right they have violated. But it does require that we do not 'pile on' excessive punishments on people as a 'package'. Each part of a punitive package must be justified individually via a desert-independent positive instrumentalist justification, and if criminal disenfranchisement is unnecessary for serving plausible instrumentalist aims, it is not justified.

2. Sentencing Constraints: Proportionality and Fairness

Proportionality

For negative retributivists, proportionality is a necessary sentencing constraint[58] because it protects negative retributivism's forward-looking instrumentalist component from the pitfalls of pure forward-looking theories like deterrence accounts, such as allowing the disproportionate punishment of the guilty as well as the punishment of the innocent.[59] In contrast to a non-comparative notion of cardinal proportionality, *ordinal proportionality* involves multidimensional comparative assessments of available punishments. Following von Hirsch's account of ordinal proportionality,[60] I define these criteria as follows:

> *Parity*: "when offenders have been convicted of crimes of similar seriousness they deserve penalties of comparable severity".
> *Rank Ordering*: "[the] relative severity [of punishment] reflects the seriousness-ranking of the crimes involved".
> *Spacing*: "Suppose crimes X, Y, and Z are of ascending order of seriousness; but that Y is considerably more serious than X but only slightly less so than Z. Then, to reflect the conduct's gravity, there should be a larger space between the penalties for X and Y than for Y and Z".

One way of making the disenfranchisement ordinally proportionate is to ensure parity for political wrongdoers guilty of *similar* kinds of wrongdoing (e.g. obstruction of justice with respect to the same institutional body). But whenever this is not possible, we can approximate at least rank ordering and spacing by depriving political wrongdoers of different kinds of participation rights. For instance, when sentencing two political wrongdoers guilty of qualitatively different crimes (such as conspiring with a hostile foreign power vis-à-vis bribing a local county official), restricting the right to vote for both political wrongdoers may be inconsistent with *rank ordering*. It may thus be more proportionate to deprive the former political official not only of her right to vote in local elections, but also

[58] Importantly, this implies that criminal disenfranchisement ought to be (contrary to current practice) an *explicit* part of sentencing, rather than a mere collateral consequence automatically applied to all offenders convicted of a particular crime.
[59] On a similar point see Lippke, "Some Surprising Implications of Negative Retributivism," 53.
[60] Andrew von Hirsch, *Censure and Sanctions* (Oxford: Clarendon Press, 1993), 18.

her right to run for office in a given election cycle, thus serving the incapacitative aim of preventing further conspiracy or obstructions of justice by elected officials. It is beyond the scope of this paper to defend a complete account of how to achieve an ordinal ranking of different types of political wrongdoing, and a corresponding ordinal ranking of temporarily forfeitable participation rights that applies to all possible sentencing contexts. My point is simply that, in principle, we can feed comparative considerations concerning the moral seriousness of different types of political wrongdoing into a context-specific, ordinally proportionate judgment about which specific democratic participation rights a given political wrongdoer temporarily forfeits.

Fairness

A concern for fairness may constrain the scope of all-things-considered justified criminal disenfranchisement in two different ways: first, disenfranchisement must not be an unfair kind of punishment *itself*, and second, disenfranchisement must not exacerbate the adverse impact of structurally unfair background conditions such as racial injustice.

The scope of a punishment—the set of temporarily forfeited rights—can vary depending on the empirical context in which a punishment is imposed, and on whether the actual sentence adequately reflects that context. For example, contrast disenfranchising a prisoner between elections, so that the loss of voting rights remains inconsequential in the sense that there would have been no opportunity to exercise it,[61] with disenfranchising a prisoner whose sentence overlaps with elections.[62] Asymmetries may occur even if everything else is held constant—the severity of the crime, the length of the prison sentence, the kind of punishment imposed. So even ostensibly *de jure* fair punishments may be unfair *de facto* due to arbitrarily different effects,[63] which courts must prevent and mitigate: for example, by disenfranchising only those criminals whose prison sentence exceeds the duration of one election period, so that all criminals disenfranchised in that jurisdiction will experience the same punitive effect—not being able to vote—at least once, irrespective of *when* they are sentenced, the latter of which is a morally arbitrary empirical fact.

Consider also the separate, and more complex, issue of unfair background conditions. Members of racial minorities are disproportionately impacted by stop-and-frisk

[61] But disenfranchisement may still be objectionable all-things-considered even if it does not *actually* lead to a tangible loss of participation rights due to a lack of opportunities to exercise them, since it may nevertheless entail a morally objectionable loss of equal social standing.
[62] See Nicholas Munn, "The Limits of Criminal Disenfranchisement," *Criminal Justice Ethics* 30 (2011): 223-39.
[63] Susan Easton, "Electing the Electorate: The Problem of Prisoner Disenfranchisement," *The Modern Law Review* 69 (2006): 445-8; and *Hirst v. UK,* §76.

policies,[64] are more likely to be arrested,[65] and are "nearly *twice* as likely to be charged with an offense carrying a mandatory minimum sentence",[66] even when controlling for a criminal's individual characteristics and criminal history. They are thus more likely to receive a sentence entailing criminal disenfranchisement as a collateral consequence. This disparate impact is in tension with existing laws,[67] such as section 2 of the US Voting Rights Act (1965), which prohibit restrictions of the electorate if such restrictions disproportionately decrease a minority's equal opportunities to influence political outcomes via voting.[68] This applies, as the 1982 amendment to the Voting Rights Act clarifies, irrespective of whether the *intent* of such restrictions is discriminatory—what matters is the discriminatory *effect*. This suggests that regardless of why and how unfair background conditions arise, once they exist courts must consider them when determining the appropriate scope of punishment during sentencing:[69] for instance, by considering a criminal's personal background, and by determining whether the proposed punishment matches punishments imposed on criminals from different backgrounds in a fair and potentially more lenient way, in light of racial bias in policing and sentencing.

Of course, doing so would neither compensate for the historical and ongoing presence of structural injustice, nor address its sources directly. However, it would help avoid *exacerbating* structural injustice further via disenfranchisement policies with an unfair political impact. Ultimately, the relevant question we must pose when we decide whether to implement the negative retributivist account of criminal disenfranchisement all things considered is whether we are in an empirical context in which *any* disenfranchisement policy, even a severely restricted one, would end up exacerbating racial injustice, and other structurally unjust inequalities. If that was the case, this would constitute an overriding

---

[64] See this recent study published by the criminal justice reform organisation *The Sentencing Project*: Nazgol Ghandnoosh, "Black Lives Matter: Eliminating Racial Inequity in the Criminal Justice System" (2013), https://sentencingproject.org/doc/publications/rd_Black_Lives_Matter.pdf, 3.

[65] The same study finds that in Ferguson, Missouri, African-Americans were twice as likely to be arrested after being stopped in traffic. This disparity is particularly alarming given that the study also found that police stopped more black persons than white persons (12% vs. 7%) in the first place, but found less contraband with black persons (22%) than white persons (34%). The report drew on data by the Office of the Missouri Attorney General, including Racial Profiling data (2013) by Ferguson PD, http://ago.mo.gov/docs/default-source/public-safety/2013agencyreports.pdf?sfvrsn=2

[66] Sonja B. Starr and M. Marit Rehavi, "Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of *Booker*," *The Yale Law Journal* 123 (2013): 28-9.

[67] Karlan, "Convictions and Doubts", 1147-70.

[68] The relevant parts of section 2 are: "(a) No voting qualification [...] shall be imposed [...] which results in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color [...]. (b) A violation of subsection (a) is established if [...]the political processes leading to nomination or election [...] are *not equally open to participation* by members of a class of citizens [...] in that its members have *less opportunity than other members of the electorate to participate* in the political process and to elect representatives of their choice." *Voting Rights Act of 1965 (1982 amendment)*, 52 U.S.C. 10101 (2018) [emphases added].

[69] For a related argument see *Thornburg v. Gingles*, 478 U.S. 30, 47 (1986).

reason not to implement the policy, thus trumping the negative retributivist account and any other justification of criminal disenfranchisement.

Acknowledging possible overriding reasons not to disenfranchise *anyone* is compatible with my account: the negative retributivist argument does not provide a context-insensitive justification for why it is *all things considered justified* for democratic states to disenfranchise political wrongdoers. The argument merely specifies which narrow set of criminals temporarily forfeit their right *not* to be punished by disenfranchisement, but I have not attempted to develop a full account of whether that argument *outweighs* competing moral reasons not to disenfranchise, such as reasons pertaining to the goal of establishing overall fair conditions in a democratic state.

V. Conclusion

Widespread current disenfranchisement policies disenfranchise the *wrong* set of people. They are *overinclusive*, because they disenfranchise persons guilty of serious, but non-political, criminal wrongdoing. While overinclusiveness would be objectionable in any context, it is particularly objectionable in circumstances in which failing to articulate a plausible narrow desert basis for criminal disenfranchisement has additional large-scale collateral consequences, for instance by perpetuating existing structures of racial inequality and exclusion. At the same time, existing disenfranchisement policies are also *underinclusive*, because they fail to disenfranchise some persons guilty of serious political wrongdoing, thus failing to articulate appropriate punitive sanctions which would help hold political wrongdoers accountable.

I have argued for this view by critiquing deterrentist, expressivist, and positive retributivist justifications of criminal disenfranchisement—including justifications of *temporary* criminal disenfranchisement—and by developing an alternative, severely restricted negative retributivist account. Importantly, the negative retributivist justification is not an all-things-considered justification, which is why it is subject to context-specific constraints and responsive to overriding reasons. Furthermore, the negative retributivist account is not committed to the view that criminal disenfranchisement is *the best way* of responding to the political wrongdoing: other measures may be more effective or more desirable, morally and politically speaking. Nevertheless, the more constructive upshot of the argument is that we have strong, yet underacknowledged reasons to question the underlying intuitions of those accounts of criminal disenfranchisement which currently dominate the philosophical debate—and those reasons point us towards approaching the question of whether we ought to disenfranchise *anyone* with much more nuance and scepticism.