Sanctuary Cities and Non-Refoulement

Michael Blake & Blake Hereth

Ethical Theory and Moral Practice
An International Forum

ISSN 1386-2820

Ethic Theory Moral Prac
DOI 10.1007/s10677-020-10082-3

Springer

Ethical Theory and Moral Practice
https://doi.org/10.1007/s10677-020-10082-3

Sanctuary Cities and Non-Refoulement

Michael Blake[1] · Blake Hereth[2]

Accepted: 8 April 2020/Published online: 22 May 2020

Abstract
More than two hundred cities in the United States have now declared themselves to be sanctuary cities. This declaration involves a commitment to non-compliance with federal law; the sanctuary city will refuse to use its own juridical power – including, more crucially, its own police powers – to assist the federal government in the deportation of undocumented residents. We will argue that the sanctuary city might be morally defensible, even if deportation is not always wrong, and even if the federal government is legally permitted to demand that states participate in the process of deportation. We defend this conclusion with reference to a simple, but powerful, norm of international law: that of non-refoulement. As we will discuss, this norm articulates the idea that no state may rightly use its coercive power to move a person into a position in which their basic rights are at risk. We take this norm as a morally defensible principle, and ask what could follow from its acceptance. We argue that this norm has implications for a variety of actors – especially when we notice that those who are facing unjust risks of death are, in general, entitled to use defensive violence against their aggressors. This simple fact, we argue, can help offer a novel defense of the sanctuary city.

Keywords Sanctuary City · *Non-Refoulement* · Refugees · Doxxing · Self-defense · Non-violent resistance

✉ Michael Blake
miblake@uw.edu

Blake Hereth
sbhereth@uark.edu

[1] University of Washington, Savery Hall, Box 353350, Seattle, WA 98195, USA

[2] Visiting Assistant Professor of Philosophy, University of Arkansas, 713 N Rupple Rd, Fayetteville, AR 72704, USA

1 Introduction

More than two hundred cities in the United States have now declared themselves to be sanctuary cities.[1] What this entails is not always entirely clear.[2] The core of the concept, though, involves a commitment to non-compliance with federal law; the sanctuary city will refuse to use its own juridical power – including, more crucially, its own police powers – to assist the federal government in the deportation of undocumented residents.[3] This means, in practice, that the sanctuary city will refuse to hold undocumented residents for transfer to the Immigration and Customs Enforcement (ICE) agency of the federal government – even when ordered to do so by that agency itself.

The legitimacy of this sort of refusal is, to put it mildly, rather controversial. The public discourse about this phenomenon has been dominated by two themes. The first is effective policing. Those opposed to sanctuary cities have sought to paint those cities as fostering crime, by removing the ability to effectively detain undocumented criminals; Donald Trump, on the campaign trail, promised to "end the sanctuary cities that have resulted in so many needless deaths."[4] Those in favor of sanctuary cities have emphasized how they enable effective policing, by encouraging the undocumented to report crime without fear of deportation. Thus, the police chiefs of Houston and Dallas led a coalition of police officers in speaking out against efforts to punish sanctuary cities: demanding that police participate in ICE activities would "further strain the relationship between law enforcement and [their] diverse communities."[5] The second theme in the public debate has been federalism. Those in favor of sanctuary cities have sought to emphasize that subsidiary political jurisdictions are entitled to make their own decisions about the policies guiding the acts of their employees; cities should not be reduced to tools with which federal authorities can execute law.[6] Those opposed to sanctuary cities argue in response that cities are not entitled to interfere with the effective administration of federal

[1] The Center for Immigration Studies maintains a list of jurisdictions – including cities, counties, and states – that identify as sanctuary jurisdictions. The list is available at https://cis.org/Map-Sanctuary-Cities-Counties-and-States

[2] We focus on the sanctuary city as a phenomenon which has been deployed primarily in the United States, but do not intend to imply that any of our conclusions depend upon any factors unique to the United States. We would note, further, that the phenomenon has been used in resistance against state-level deportation practices in a variety of contexts – including, most recently, Italy, where mayors of cities such as Palermo have used this concept to explain their refusal to assist the Italian government's current regime of deportation. See Jason Horowitz, "Italy's Crackdown on Migrants Meets a Grass-Roots Resistance," *The New York Times,* February 1, 2019.

[3] A good summary of the phenomenon can be found in Darla Cameron, "How sanctuary cities work, and how Trump's executive order might affect them," *Washington Post*, January 18, 2017. The Trump administration's summary of its position is available in a brief issued on March 20, 2018, entitled "Sanctuary Cities Undermine Law Enforcement and Endanger Our Communities," available at https://www.whitehouse.gov/briefings-statements/sanctuary-cities-undermine-law-enforcement-endanger-communities/

[4] Quoted in Christopher Ingraham, "Trump says sanctuary cities are hotbeds of crime. Data say the opposite." *Washington Post,* January 27, 2017. This article discusses the research of Tom K. Wong, who has studied the effects of sanctuary policies on crime. His research has consistently found lower crime rates in sanctuary jurisdictions than in non-sanctuary jurisdictions. See Tom K. Wong, "The Effects of Sanctuary Policies on Crime and the Economy," *American Progress,* January 26, 2017.

[5] David Pughes and Art Acevedo, "Texas police chiefs: Do not burden police officers with federal immigration enforcement," *Dallas Morning News,* April 28, 2017.

[6] Thus, Jacques Derrida hopes that the sanctuary city as an idea might act as a catalyst for the revision of the Westphalian state system. Although he refers primarily to the European sense of the sanctuary city, his point applies with equal strength to the North American sense. Jacques Derrida, *On Cosmopolitanism and Forgiveness* (London: Routledge, 2001).

law; Jeff Sessions, then Attorney General of the United States, sought to cut off federal funding to any jurisdiction that refuses to cooperate with federal instructions regarding deportation. Whether this threat is Constitutional is not yet clear.[7] Federal law mandates that subsidiary jurisdictions may not "in any way restrict" federal agencies from acquiring information about citizenship or immigration status.[8] Recent constitutional litigation, however, has emphasized that states may not be "pressed into service" at the command of the federal government.[9] The current legal fight over sanctuary cities will likely be dominated by the interplay between these two visions of federalism and subsidiarity.

In this chapter, however, we want to ask a more distinctively *moral* question about sanctuary cities, and we do this by simply stipulating that Sessions' efforts will succeed. We will stipulate, that is, that the federal government is, in general, Constitutionally entitled to insist upon the cooperation of city and state resources in the administration of migration policy. We will further stipulate that it is not, in general, morally wrongful that the federal government engage in practices of deportation. We are thus assuming, in the present chapter, that there are at least some cases in which the forcible removal of a current resident from her country of residence is not inherently immoral.[10] What, then, might be said about the morality of the sanctuary city, after these stipulations? Would it not follow that the sanctuary city is morally wrongful, if the federal government is entitled to the legal and moral deference we imagine?

We will argue, in contrast, that the sanctuary city might be morally defensible, even if deportation is not always wrong, and even if the federal government is legally permitted to demand that states participate in the process of deportation. We defend this conclusion with reference to a simple, but powerful, norm of international law: that of *non-refoulement.* As we will discuss, this norm articulates the idea that no state may rightly use its coercive power to move a person into a position in which their basic rights are at risk. We take this norm as a morally defensible principle, and ask what could follow from its acceptance. We argue that this norm has implications for a variety of actors – especially when we notice that those who are facing unjust risks of death are, in general, entitled to use defensive violence against their aggressors. This simple fact, we argue, can help offer a novel defense of the sanctuary city.

We will make this argument in several stages. We will begin with the principle of non-refoulement (henceforth PNR), which we take to articulate a powerful moral norm about harm. We will then argue that this principle entails that those who are subject to violations of that norm have some limited right to use harm in defense of their interests. We then argue that this defensive harm might be rightly invoked not only by the victims of those whose rights are violated, but by those who are mere bystanders to the harm in question. This is, of course, likely to be especially important in the case of refugees; since they are marked out precisely by their lack of effective power, we have a reason to ask about the moral status of acts done by the more powerful on their behalf. We then argue, finally, that where there is a right to engage in harm on behalf of the victim, there is a right to engage in other forms of otherwise impermissible act on behalf of that victim – including, at the very least, the refusal to abide by a valid law issued by an otherwise morally legitimate state. We apply these ideas to the concept of the sanctuary city, and use them to defend the legitimacy of the sanctuary city under

[7] The threat is currently the subject of litigation. For a discussion of the constitutional issues, see Ilya Somin, "Federalism, the Constitution, and sanctuary cities," *Washington Post,* November 26, 2016.
[8] 8 U.S.C. §1373(a).
[9] *Prinz v. United States,* 521 U.S. 898 (1997) at 905.
[10] It should be noted that not all of us are actually committed to the truth of these stipulations; we merely ask what would follow from that truth, were it to obtain.

current political realities. If current federal practices of deportation involve a serious risk of violation of basic rights, then the city has a limited grant to do what is normally wrong by means of opposition to those practices. We conclude by examining some objections to these arguments – as well as some implications that might emerge from arguments such as our own.

Some final notes might help situate this argument, with reference to what it claims – and what it does not claim. We do not, in particular, identify anything of special moral importance about the city, taken as one sort of political entity. The arguments we make on behalf of the sanctuary city might as easily be applied to any number of federal sub-units, from cities to counties or states. Instead, what we seek to do is to argue that the moral principle undergirding the PNR offers a powerful resource for political agents seeking to explain their non-compliance with otherwise legally binding federal orders. What we seek to demonstrate, in short, is that the PNR offers a moral principle that has some normative force in refusing to act upon that federal state's orders – not simply because the PNR has considerable moral power, but because the moral force of that principle is implicitly recognized by the federal state itself, insofar as it acts within the framework of international law.

2 The Principle of *Non-Refoulement*

The PNR is one of the cornerstones of modern humanitarian law. It is enshrined in Article 33 of the 1951 Convention relating to the Status of Refugees:

> No Contracting state shall expel or return ("refouler") a refugee to the frontiers of territories where his life or freedom would be threatened on account of his race, religion, nationality, membership of a particular social group or political opinion.

The Convention emerged from the horrors of the Second World War and focused primarily upon the evil most central in those horrors: namely, persecution on the basis of group identity. The idea that states ought not return vulnerable people to the places from which they have sought refuge, however, predates the Convention.[11] The idea, moreover, has in more recent documents been applied to people whose circumstances do not fall within the Convention's definition of the refugee. The 1966 International Covenant on Civil and Political Rights, for instance, defines a broader set of human rights that can be used to ground a claim of immunity from *refoulement*; these rights include freedom from slavery, forced labor, torture, degrading treatment, and the arbitrary imposition of death. These rights – which make no reference to group membership – are taken as core international human rights, whose violation can ground a claim under the PNR.[12]

There are any number of conceptual and legal difficulties with the PNR. There are enormous disputes, both in theory and in practice, about what exactly the norm prevents; the Australian practice of intercepting ships of asylum seekers at sea, for instance, may or may not breach the norm, as might the earlier United States practice of intercepting Haitian ships

[11] See the Convention Relating to the International Status of Refugees of 1933, which contains an earlier statement of *non-refoulement.*

[12] The concept of "complementary protection" is often invoked in these discussions; see Jane McAdam and Fiona Chong, *Refugees: Why Seeking Asylum is Legal and Australia's Policies Are Not* (Sydney: UNSW Press, 2014) 33.

prior and returning those ships to Haiti.[13] There are also disputes about the nature of the norm itself: While it clearly binds those parties that have agreed to be bound by it, there are some international lawyers who contend that the norm has acquired the status of a universal norm binding on all – a *jus cogens* norm, to use the lawyer's term.[14] In what follows, we will ignore these complexities, and focus on the PNR as a statement of a particular *moral* view. The norm is, on our view, a clear statement of what ought to be entailed by the recognition of the fact that humans have certain rights, and that some states exist in which these rights are inadequately respected. We will thus take the PNR to speak not only a legal, but a moral truth: that we fail to respect these rights, when we use coercive force to return people to circumstances in which their rights are subject to systematic violation.[15]

We use the norm in this way, because we are interested not simply in how rights constrain the acts of states over those residents within their territory. However poorly it is respected in practice, the relationship between the norm against torture (say) and the rights of the state is at least conceptually clear; the state has no right to torture those within its coercive grasp. We are interested, instead, in how the rights and obligations of third parties are affected by the recognition of human rights. The PNR is a good place for us to start, because it identifies one way in which the human right to freedom from torture constrains even those states who do not themselves use torture; namely, it precludes the use of force to return a foreign citizen to circumstances in which she will be tortured. Our contention is that these ideas can ground a defense of the practice of the sanctuary city. To make this case, though, we must first discuss the ways in which norms such as the PNR can ground individual rights to respond to wrongdoing elsewhere. It is to this task that we now turn.

2.1 Defensive Violence

We are interested, as we have noted, in the ways in which the PNR constrains the acts of third parties. We want to explore this norm, however, by looking first at the ways in which the threat of wrongdoing alters the rights of the one against whom wrong is threatened. Take, to begin, the norm against torture, as discussed above. A proposed violation of that norm tends to bring with it the possibility of legitimate defensive violence. Consider, for example, the following case:

> *Freelance Torturer*
> Victim is at home reading when Torturer, who was hired by Hater to torture Victim since Hater despises Victim, breaks in and begins to torture Victim. After several hours of torture, Victim has an opportunity to grab a knife and inflict mild and non-fatal defensive wounds on Torturer, allowing Victim to escape.

We take it that that Victim's use of defensive violence in this case will be widely regarded as morally permissible, and this judgment is reflected in the literature.[16] Why should this be so?

[13] On the latter, see *Sale v. Haitian Centers Council, Inc.* 509 U.S. 155 (1993), in which the Supreme Court upheld the practice's legitimacy, on grounds that have been criticized by international lawyers.
[14] Jean Allais, "The *jus cogens* nature of non-refoulement," 13(4) *International Journal of Refugee Law* (2002).
[15] For this reason, we do not consider moral interpretations of the PNR where refugees flee *natural* disasters such as earthquakes, droughts, and hurricanes. This is because the PNR supports protection from threats that are made with discriminatory intent, and natural disasters lack the requisite *mens rea*.
[16] The case is similar to Thomson (1991: 283), Doggett (2011: 220), and McMahan (2005: 390). Frowe (2014: 175) considers a case in which a mafia boss raises funds from fellow mobsters to hire an assassin to impermissibly kill an individual, and defends the suggestion that the mere financiers of the assassination are liable to defensive harm. Frowe's judgment in this case reflects her strong conviction that the assassin is liable to defensive harm.

The explanation we favor, and one that has considerable attraction in the contemporary literature, is that moral permissibility more readily attaches to *less unjust* actions than *more unjust* actions.[17] We then make the relatively weak assumption that because Victim is less responsible than Torturer for the fact that some harm will befall either Victim or Torturer, and because Victim is not liable to any harm from Torturer, justice favors prioritizing avoiding harming Victim over avoiding harming Torturer. Since this is an agent-neutral fact of justice, it is one that applies to Victim's actions, and thus Victim is permitted to act so as to prioritize justice. Thus, Victim is permitted to bring it about that the harm befalls Torturer and not Victim.

So far, so good. Suppose, however, that Torturer was not a freelance employee but an employee of the state doing state-sanctioned (but still morally unjustified) work. Would Victim still be permitted to engage in defensive violence against Torturer? It seems so, unless we are to endorse the implausible view that defensive violence against state agents doing state-sanctioned work is *never* permissible. Put another way: If defensive violence against state-sanctioned violence is ever permissible, it's permissible in Victim's case against Torturer. Moreover, this is true even if Torturer is a *reluctant* perpetrator of the state's official purposes. For even if we suppose that it's unjust for the state to coerce Torturer to assault Victim and thus also unfair for Torturer to suffer defensive injuries as a result, it's nevertheless *more* unfair for Victim to suffer these injuries. After all, it is *Torturer* who initiates the unjust attack against Victim, not Victim, and thus it is Torturer who bears more responsibility for the fact that an unjust harm will befall one of them. Those with more responsibility for an unjust harm's occurrence should (other things being equal) suffer more of that unjust harm than those with less responsibility.[18] This is why in a conflict where harm will befall either an unjust aggressor, a non-liable victim who's fighting back, and a bystander, justice prefers the harm to befall the unjust aggressor primarily, the non-liable victim secondarily, and the bystander thirdly.[19]

Thus far, we have considered cases where the harm inflicted against victims is *direct* in the sense that it is the perpetrators themselves who do the harming. This conclusion has some usefulness for refugees: It explains why certain malign actors in their home states might be liable to defensive harms from those refugees and perhaps also from other states, as is the case in humanitarian wars. What it does not explain, and what we hope to explain presently, is what permissions attach to cases where the harm is inflicted by the host state (as opposed to the home state) is *indirect* in the sense that the host state commits no violence itself against the refugees, but instead exposes the refugees to violence. To begin, consider the following case:

> *Malfunctioning Airplane*
> Agent's job is to place Migrant on a plane that leaves the country. The only plane available, however, is a malfunctioning autonomous airplane. It is likely, but not absolutely certain, that the plane will crash, resulting in Migrant's death. As Agent attempts to force Migrant onto the plane, Migrant has an opportunity to knock Agent over the head, mildly injuring Agent, and escape.

[17] We say "more readily attaches" so as to avoid the stronger commitment that avoiding less unjust actions is *necessarily* permissible. We suspect this is true, but we assume a weaker commitment to avoid unnecessary controversy.

[18] It doesn't follow from this that the individual responsible for the creation of *justified* threats should (other things being equal) suffer more of the unjust harm than those with less responsibility.

[19] See, for example, Frowe (2016: 160–161) and Frowe (2014: 44–45).

As a backstory, let us suppose that Migrant had previously been legally admitted into Agent's country. Let us also assume that Agent does not intend Migrant's death, but instead intends merely to remove Migrant from the country knowing it's likely that doing so will result in Migrant's death. Assuming Migrant is not liable to the harm they would suffer if their plane were to crash, would Migrant act permissibly if they knocked Agent unconscious? It seems so, but there is more than mere moral intuition to offer here. The harm to be inflicted on Migrant is unjust, and even if we maintain that any harm suffered by Agent is also unjust, it appears *less* unjust than the harm to be suffered by Migrant.[20] For again, Migrant has done nothing and Agent has done something, and thus justice favors the protection of Migrant over Agent.

When certain migrants, such as refugees, are pushed out of the country hosting them, they are not sent in malfunctioning planes that descend like a bullet into a blaze of fire, but they *are* sent back to their home countries to be fired upon by the bullets of unjust aggressors. Examples in which a refugee is returned to their abusive home country or to a country in a state of civil war present the clearest cases. The case of Malala Yousafzai is one of them. After resisting Taliban efforts to keep girls out of school, Malala (and several of her friends) were shot by the Taliban in an effort to silence their resistance. Malala was evacuated from Pakistan to the United Kingdom where she received treatment for her critical injuries and where she and her family currently reside. The threat to Malala from the Taliban in Pakistan still exists, but it is not as severe as it is for many refugees. Let's suppose that Malala, if she returned to Pakistan, would be under severe threat from the Taliban—that she would, for example, be executed on the tarmac after her plane landed. If that were true, then forcing Malala to return to Pakistan would be morally analogous to forcing Migrant onto the malfunctioning plane. If Malala were able to effectively save her own life by escaping British immigration custody through the use of defensive violence, she would be permitted to do so, just as Migrant would be.

Of course, deploying *effective* violence against state officials is not always feasible. Were most refugees to violently resist being deported to their home countries, they would fail to secure their own safety. Malala might evade the British immigration agent at her gate only to be accosted by security at another checkpoint. It might be thought, then, that violent self-defense on the part of non-liable refugees faced with deportation is more often impermissible than permissible since it accomplishes nothing. This is sometimes called the Success Condition for permissible self-defense, according to which "otherwise immoral acts can be justified under the right to self-defense only if they are likely to achieve defense from the perceived threat."[21] Daniel Statman has offered a powerful challenge to this claim, however, in an example where someone is under threat of death:

> John Wayne is hunted by a bunch of bad guys who want to kill him. His shelter is surrounded by them, and he can expect no help from anyone. Only a miracle could make

[20] It might be objected that this defense is problematic because it implicitly assumes that Migrant's death would be unjust if it were caused by, as it were, a natural event. Plane crashes, like tree falls, are natural events, and the harms one suffers as the result of mere natural events can't be unjust. This view seems at least implicit in, for example, John Rawls (1971). In response to this objection, it will not do to appeal to Agent's wrongful action as grounding the possibility of injustice, since we seek to *show* that Agent's action is unjust, and thus we can't assume the action to be unjust without vicious circularity. However, our argument is not *that* Agent's action is unjust; it's that *if* Agent's action is unjust, then Migrant is permitted to employ certain defensive measures against Agent. What's more, since refugees are typically under the threat of harm by other agents, the possibility of injustice is very much a live one.

[21] Statman (2008: 659).

> his defensive acts effective in protecting his life. On [the Success Condition], using his last bullets to kill some of those people could not be licensed on grounds of self-defense…. If this is what some moral or legal theory demands of us, it seems like a reductio of the theory.[22]

Statman's example should make us reconsider whether the Success Condition really is a necessary condition for permissible self-defense. During the course of explicitly considering Statman's proposal, Helen Frowe makes the following remarks:

> Wayne might not prevent other people from killing him by killing two of his attackers, but he prevents *those men* from killing him. There is no reason to think that the targets are not liable to be killed and that such killings could not be all-things-considered permissible on most accounts of self-defense. Such accounts do not typically demand that one's defense be necessary for making oneself better off over all, but only necessary for averting a threat. I may, for example, kill a (culpable) person who will otherwise kill me even if I am certain to die of an illness the next day.[23]

Frowe's analysis strikes us as an accurate moral analysis. Each of the outlaws attempting to kill John Wayne represents an individual threat. Wayne is under threat of only one harm: death. But five people are threatening that same harm, each of them trying to bring it about through their own individual efforts. Thus, it's one harm but *five* wrongdoings at play, and Wayne is therefore permitted at least one defensive attempt per wrongdoing. In Wayne's case, that means killing two of the five men. In Malala's case, that means harming one of the agents even if she can't evade the rest. The judgment generalizes to refugees beyond Malala: Refugees are permitted to defend themselves violently against those who attempt to return them to exceptionally dangerous conditions – even if it is not clear that the violence will succeed in preventing that return.

We should not stop here, however, but should extend our moral judgment to *third-party defense*. While it is not always true that third parties are permitted to defend individuals under threat of unjust harm, it *is* true that third parties enjoy a *presumptive* moral permission to do so. Here again, we can again appeal to the justice-based account of self-defense utilized earlier: It is an agent-neutral fact that justice prioritizes protecting Malala over her unjust aggressors, and this explains why Malala is permitted to self-defend. But since it is an agent-*neutral* fact, it also explains why third parties are (again, presumptively) permitted to defend Malala. For it is a fairly uncontroversial assumption about morality that if justice prefers one action to another, the former action is one moral agents are presumptively permitted to pursue. Again, this judgment generalizes to refugees beyond Malala: Third parties are permitted to defend refugees violently against those who attempt to return the refugees to exceptionally dangerous conditions.

3 Non-Violent Defense

Thus far, our arguments have addressed *violent* self-defense and other-defense. Typically, as in Migrant's case, this involves physically attacking someone. This might be of theoretical

[22] Statman (2008: 664).
[23] Frowe (2014: 111).

interest, but it is likely to be of little practical importance. Violent resistance to even the most unjust regime of migration control is unlikely to effectively constrain that regime – and refugees are the least likely of all to have the ability to deploy violence effectively. Refugees, after all, are marked out precisely as those people least able to deploy power on their behalf; they are refugees precisely because they do not have any ability to control that sort of coercive force that is the hallmark of a functioning state. What is more interesting, though, is how a permission to use violent resistance might entail a permission to engage in other, less violent forms of resistance – which will ground our defense of the sanctuary city. The permission to do what is usually wrongful might extend not only to violence, that is, but to deception and non-participation.

We will therefore begin by examining some forms of non-violent wrongdoing. The first two examples are those of *threatening* and *doxxing*. We shall discuss them first since, among the forms of defensive resistance we discuss, threats of physical violence and doxxing are morally closest to violence, and because our moral analysis of what makes them permissible and how we know they are permissible sheds further light on the permissibility of other forms of defensive resistance and, ultimately, sanctuary cities. To begin, consider an example involving threats of physical violence:

3.1 Threatening Phone Call

After attempting to escape Agent, Migrant is re-captured and watched closely by Agent until the next flight is available. During this time an ally of Migrant's, Caller, phones Agent and sincerely threatens to injure Agent unless Agent releases Migrant.

In the example, Caller does something presumptively wrong by threatening to injure Agent. To say this is just to say that, other things being equal, it is wrong to threaten harm to Agent. As we argued previously, however, it would *not* in fact be wrong for Migrant to harm Agent as justice favors prioritizing avoiding harm to Migrant over avoiding harm to Agent. Moreover, since this is an agent-neutral fact about justice, it follows that Caller is likewise (presumptively) permitted to assist Migrant in harming Agent. What's more, if Caller is permitted to assist Migrant in actually harming Agent, then Caller is also permitted to *intend* to do this. This offers strong support for the conclusion that Caller is not only permitted to harm or intend to harm Agent, but also to threaten to harm Agent. Further support can be offered. Consider two scenarios in which a law enforcement officer is attempting to save a young child from someone who threatens to kill the child. In the first scenario, the officer (at no risk to the child) shoots the criminal without warning. In the second, the officer (again at no risk to the child) issues a warning and shoots the criminal only after the warning is ignored. If we assume that the officer acted permissibly in the former case, it seems as though they acted permissibly in the latter case as well. The mere issuance of a warning—a threat to impose harm unless the criminal ceases to threaten the child—does not itself make the officer's action impermissible. Indeed, in cases where the criminal does not threaten imminent harm to the child, it might seem even *obligatory* for the officer to issue a warning. All of this is strong evidence that Caller threatening Agent is permissible, at least presumptively.

The case of doxxing is more complicated. By 'doxxing', we mean publishing safety-sensitive or harassment-enabling identifying information to the general public with coercive or punitive intent. In 2015, the hacker group Anonymous doxxed several hundred members of the Ku Klux Klan by posting their identities in publicly accessible online forums in order to shame those individuals for their support of racist causes. More recently, in 2016, former

Politico editor Michael Hirsh used a social media platform to publish the home address of white nationalist Richard Spencer. These examples do not involve the direct infliction of physical harm, and nor are they explicit threats of harm. Their function is to *enable* harm or harassment by publicly outing the personal profile or location of individuals likely to incite widespread and intense social anger or rejection.

Doxxing is typically regarded as presumptively impermissible for two reasons. First, it is a presumptive violation of the doxxed individual's right to privacy. Second, and more fundamentally, it is a presumptive violation of the doxxed individual's right to safety. In order to show that that doxxing individuals is permissible, it is crucial to show that these primary presumptions against doxxing are not decisive.

Consider, first, the right to privacy. It is commonly accepted that an individual's right to privacy is not absolute. Amber Alerts, for example, typically publish as much information about suspected kidnappers as is then known, yet we do not regard this as a violation of the kidnapper's right to privacy. Plausibly, if one individual unjustly threatens another and publishing certain private information of the threatening individual would aid the threatened individual, it does not violate the threatening individual's right to privacy to release that private information.[24] Thus, if Agent is unjustly threatening Migrant and doxxing Agent would defensively aid Migrant, then Agent lacks a right to privacy with respect to certain key information.[25] While this does not entail an all-things-considered moral permission to dox Agent, it does entail that doing so is not a violation of *Agent's* rights, since Agent is liable to having this information released. This is true even if Agent is acting in full legality, for our claim here is not a legal one but a moral one. Were it legal for law enforcement in the United States to kill Afghani immigrants on-sight, their doing so would nonetheless constitute an unjust assault against Afghani immigrants and would render the perpetrators liable to defensive harms. Justice, to be short, is not partial to actions that are legal if they are immoral; instead, justice prefers moral courses of action to legal courses of action. This is because, at the very least, justice is not principally a fixture of legality but of morality. On the view of defensive permissions we have defended and relied upon, what justice favors is what permissibility more readily attaches itself to, and thus the moral presumption in favor of doxxing defended here is not undermined by the mere fact that the doxxed individual acts legally.

The presumption against an individual's right to safety (or, more negatively put: their right against unjust harm) can be similarly treated. Individuals who are liable to defensive harm lack a right to safety because they have *forfeited* that right. The right to safety, like the right to privacy, is not absolute: It protects individuals *only* from unjust harm. The cases we are presently considering are ones in which Agent threatens unjust harm to Migrant without justification. Because Agent's threat to Migrant is unjust and because Agent is morally responsible for unjustly threatening Migrant (or at least is more responsible than Migrant is), a threat against Agent qua agent of injustice is not itself unjust, and thus not the sort of harm

[24] This is not to say that releasing *any* private information would be permissible, however. A kidnapper has no right to keep their location or license plate number private if they are transporting a kidnapped individual in their car (or are reasonably suspected of doing so), but they might maintain a right to privacy regarding, for example, their voting record. If it is permissible to release an individual's private information, therefore, it will be restricted to information to which that individual has forfeited their right. More specifically, it will be information the release of which is necessary (or reasonably regarded as necessary) to prevent the harm threatened by the kidnapper.

[25] In order to avoid violating an individual's right to privacy by releasing their private information, it must also be true that the release of that information would be proportionate (in the narrow sense) to the unjust harm they pose. In other words, the harm posed to them via public exposure must be proportionate to the unjust harm they pose.

Agent has a right against (i.e., an unjust harm). Permissibility, as we have seen, favors what justice favors, and thus it is permissible for Migrant to defensively harm Agent and presumptively permissible for third parties to defensively aid Migrant. Thus, the right to safety is irrelevant in cases where Agent is already liable to defensive harm.

There is yet another reason to believe that doxxing Agent is permissible: If defensive *violence* is permissible against Agent, then defensive *doxxing* is also permissible. The inference we implicitly rely upon here is what we shall call the Lesser Harm Liability Principle: If Agent would be liable to some serious harm if there were no likely effective lesser harm, and if there is some lesser harm that will likely succeed in warding off Agent's threat, then Agent is liable to the lesser harm. Various self-defense theorists, such as Helen Frowe appear to rely upon similar inferences:

> Assuming that lethal force is a proportionate response to rape, harms up to lethal force—for example, permanent paralysis or blindness—will be proportionate for defending the parasitic threat to her honor.[26]

We agree with this conclusion, which we term the Lesser Harm Liability Principle.

Our reliance on the Lesser Harm Liability Principle is intended to support our claims about the permissibility of threatening and doxxing. If successful, however, it will do more than that. It will also show that other defensive measures intended to protect victims like Migrant are permissible, such as lying or concealment. In defense of the Lesser Harm Liability Principle, imagine a case like *Malfunctioning Airplane* but imagine that, in this revised case, Migrant needs only a moment of distraction to evade Agent and can accomplish this by pinching Agent on the arm, causing Agent very mild pain and affording Migrant the necessary distraction. Migrant knows that Agent is liable to being knocked over the head and infers that Agent is also liable to being pinched. If we suppose that the Lesser Harm Liability Principle is false, we must also suppose that Migrant should not rely on it during moral deliberation. Yet that implies that if Migrant must make a quick decision, they are not permitted.

If some lesser harm will do the trick for Migrant (so to speak), how is it that Agent is liable to a greater harm? Here it is essential to see that liability and permissibility are not the same. To say that Agent is liable to some harm is to say that the infliction of that harm on Agent does not wrong Agent. It might, however, be wrong for other reasons. For example, consider again the case of *Malfunctioning Airplane*. In the example, Agent unjustly threatens Migrant with death. Were Migrant to make a lethal threat to Agent in self-defense, Migrant would not thereby wrong Agent. However, if some lesser harm would do the trick, then inflicting a greater harm on Agent would be either unnecessary or disproportionate (or both) since the *degree* of harm suffered by Agent would be excessive. Inflicting a greater harm is permissible only if it is necessary and proportionate, and thus if Migrant intends to self-defend using one of these means, Migrant is *obligated* to inflict the lesser harm. If a lesser harm will do the trick, then it is permissibility, and not liability, that is affected.

This analysis further strengthens our account. Even if we assume that defensive violence in the form of assault or killing would result in the successful defense of refugees and other unprotected migrants, other, less harmful avenues of defensive resistance are also likely to

[26] Frowe (2014: 110). While Frowe's claim is about proportionality, there is reason to believe she also has liability in mind. Her concern is about whether it is *permissible* to inflict lesser harms when greater harms would nonetheless be (narrowly) proportionate. Thus, while her language appears restricted to proportionality, it is not. Note that not all theorists accept this implication.

succeed. Threats of harm, doxxing, lying, and concealment are ways of protecting refugees like Migrant and Malala, and they are typically far more successful and less harmful to state agents than overt physical violence. Because of this, these forms of resistance will more often meet necessity and proportionality requirements and are thus morally preferable to their more harmful alternatives.

At this point, it would be useful to review the overall structure of our argument. We have argued that the PNR identifies a central moral norm, on which we are morally prohibited from using coercive force to return to people in which their basic rights are at risk. We have further argued that one who is facing a risk of such a return is licensed to use means that are ordinarily impermissible in response to that risk – including, in at least some cases, the use of violence. We have further argued that these permissions to use ordinarily prohibited means extend not simply to the one facing the rights-deprivation, but to third parties as well. We have argued, finally, that the permission to use violent means implies a similar permission to use non-violent, but presumptively wrong, means as well. These materials, we argue, can ground a moral defense of the sanctuary city – and it is to that defense that we now turn.

3.2 The Sanctuary City

The modern conception of the sanctuary city traces its lineage back to the sanctuary movement of the 1980s. The Reagan Administration, which found it politically awkward to acknowledge human rights abuses in the allied states of El Salvador and Guatemala, refused to recognize asylum-seekers from these countries as having legitimate claims, and sought widespread deportation. The sanctuary movement leveraged the Christian concept of sanctuary, and provided space – often within the sacred spaces of Churches – in which the asylum seekers from those countries would be protected. In Tucson's Southside Presbyterian Church, between fifty and a hundred people slept on the floor each night, protected – to a limited degree – by the federal government's unwillingness to send heavily armed police into a sacred space to arrest clergy and their followers.[27] John Fife, the pastor of Southside Presbyterian, connected this in 2008 to more recent conflicts over migration:

> When government violates human rights, the church needs to be a safe space. In the 1980s, people were fleeing death squads and massacres in Guatemala and El Salvador . . . Now the government is threatening human rights and family integrity, and parents are the disappeared from the workplace. U. S. policy is to use death in the desert as a deterrent to coming here, and that is a violation of international law and human rights. Churches ought to stand up for the right of people to work and feed their families.[28]

What is new in the recent debate over sanctuary is the contention that not simply the church, but the city itself, might be a site of resistance against those who propose the violation of human rights. The modern sanctuary movement understands itself as the inheritor of the earlier movement's refusal to be complicit in injustice. Where that early movement sought to resist immigration law through civil society and individual agency, the modern movement begins with the city itself. It argues that, when the federal government violates the rights of the

[27] Alex Kotlowitz, "The Limits of Sanctuary Cities," *The New Yorker,* November 26, 2016.
[28] Interview available at http://reflections.yale.edu/article/who-my-neighbor-facing-immigration/no-more-deaths-interview-john-fife

vulnerable, the city can do right by refusing those forms of participation that are generally incumbent upon subsidiary political units within a federal system.

There are, of course, any number of moral defenses of the sanctuary city. One of these might begin with civil disobedience, and the legitimacy of civil disobedience by city officials.[29] On this analysis, the sanctuary city – like the sanctuary movement before it – is intended to cause a new form of political discourse, by holding public attention to particular forms of injustice through public disobedience. Martin Luther King's analysis of civil disobedience, for instance, explicitly justifies non-compliance with law as permissible only when ordinary politics has failed – and it is justified only as a means of returning to ordinary politics: negotiation, King writes, is the "purpose of direct action," in that disobedience must continue until a political community "that has consistently refused to negotiate is forced to confront the issue."[30] Another possible basis might be the simple refusal to participate in injustice. John Fife described his church as facing a choice between "complicity and resistance," in which the only rightful choice was resistance.[31] The former basis begins with the thought that we might have a limited permission to do what is ordinarily wrong, when doing so calls attention to an evil and brings about a new conversation about political responses to that evil. The latter basis begins with the more simple thought that we are called upon to avoid *doing* wrong, even when told to do so by an agent with presumptive authority over our actions.[32]

We do not want to dispute the power of either of these rationales. What we want to do, instead, is offer an alternative, which we take to have some advantages over these two moral accounts. The argument from civil disobedience, after all, makes the rightness of our actions depend upon the effects of those acts upon political discourse – whereas we wish to defend the sanctuary city, even in those contexts in which its value as a political statement is minimal. In particular, we want to insist that the sanctuary city might be justified, based on the moral principle undergirding the PNR, even when the discourse surrounding migrant rights is not improved by one's practice of sanctuary; the argument we make – unlike the argument made by King – does not depend upon any return to normal political discourse, as a part of its justification for disobedience. The argument from complicity, moreover, depends upon the sanctuary city's acts being understood as omissions – as refusal to help, refusal to cooperate, and so forth. While many sanctuary city policies can be understood in these terms, we wish to hold out as potentially legitimate even those sorts of sanctuary policies that are best understood in active terms – including, perhaps, the active refusal to allow ICE agents to enter city buildings, the active deletion of information demanded by ICE policy, and so on.

Our own alternative, instead, begins with the thought that those who violate the PNR not only enable wrongdoing; they do, in fact, wrong those against whom the violation is brought to bear. People are wronged, to return to our earlier discussion, when they are tortured; and those who use violence or coercion to bring people to the torturer commit a wrong as well. If this is so, though, then defensive violence seems at least potentially legitimate, as a response to serious forms of wrongdoing – not simply on the part of the victim, but also on the part of those who are mere bystanders to the wrong in question. If this latter fact is true, however, then

[29] On official disobedience, see Arthur Applbaum, *Ethics for Adversaries* (Cambridge: Harvard University Press).

[30] Martin Luther King, Jr., "Letter from the Birmingham Jail," available at http://okra.stanford.edu/transcription/document_images/undecided/630416-019.pdf

[31] Kotlowitz.

[32] Recall, here, that Socrates – who famously defended obedience to unjust law in the *Crito* – refused to obey legal orders to perform unjust acts himself; see *Apology* 32d.

bystanders would seem to have a limited permission to use not simply violent or coercive means, but also lesser tools, as a response to serious forms of wrongdoing. These thoughts seem to provide us with the tools needed to justify the sanctuary city. Imagine that the leaders of such a city become convinced that the federal government is going to fail to live up to the demands placed upon it by the PNR. There exist some people, that is, who will be directly wronged by the federal government's practices of deportation, as that deportation will be to a country in which the most basic human rights of that person will be violated. The sanctuary city might, in response to these thoughts, argue that the city has the right to engage in means that are ordinarily prohibited, in response to these patterns of harm and wrong. If the city's rightful range of actions are ordinarily defined with reference to constitutional norms, and if that constitution marks out certain forms of act as wrongful, and if that constitution has some degree of moral power – then what the city proposes to do here is at least potentially wrongful. But what we argue is that the city is here permitted to what is presumptively wrong, when it has the possibility of working in defense of the most central rights of the human person.

These ideas are not, perhaps, new; the public response to the Fugitive Slave Acts – which demanded the participation of the Northern states in detaining and returning slaves to the South – had a similar echo of lawbreaking as a tool for the frustration of evil. (The Declaration of Causes of Seceding States complains loudly about such local disobedience: the states of Ohio and Iowa have "refused to surrender fugitives charged with . . . inciting servile insurrection," complains the document – which South Carolina took as justification for political divorce from Ohio and Iowa.)[33] What we want to insist upon, though, is that these ideas are useful for understanding the morality of the sanctuary city in the present context. The sanctuary city might be legitimate, we argue, even if cities are ordinarily bound to obey the commands of the federal government; and this legitimacy might emerge from the simple fact that those over whom that federal government proposes to act have the right to assistance from third parties, which can legitimately include means ordinarily prohibited.

Do our present circumstances make this sort of defense possible? It would be possible, of course, only if the federal government was in fact likely to violate its duties under the PNR. This is, sadly, not an unreasonable assumption for an individual city to make. In the first instance, the proposed executive order on migration brought out by the Trump administration in 2017 – which followed his defense, on the campaign trail, of a shutdown of immigration rights for Muslims – has been widely interpreted as a refusal to be bound by the norms of international law.[34] The complete bar on refugees from Syria, in particular, announces an abandonment of the principle that no-one shall be returned to a place in which their basic rights are at risk. Sadly, however, we do not even need to make the case here based upon the particular decisions of the Trump Administration. Recent history suggests that the federal government has a longer track record of poor compliance with the PNR. Return to the migration crisis of 2010–2014, in which migrants – including many children – sought admission to the United States, to escape persistent violence in Central America. Amnesty International has documented that 83 people have been murdered in Central America during this time period, after having been deported from the United States.[35] The city could, instead, simply rely upon the federal government's own legal arguments. The PNR has an exception, on which we may return someone guilty of a "particularly serious crime" to a situation of risk –

[33] See document at https://www.civilwar.org/learn/primary-sources/declaration-causes-seceding-states
[34] Jamil Dakwar, "All international laws Trump's Muslim ban is breaking," *Al Jazeera,* 2 February 2017.
[35] Report available at https://www.amnestyusa.org/files/central_american_refugees_-_report_eng_1-min.pdf

a clause which was originally intended to prevent Nazi leaders from taking advantage of the PNR for their own purposes.[36] The United States has begun to stretch this clause well beyond its original intention, arguing in a recent case that the purchase of five dollars' worth of crack cocaine was a "particularly serious crime" for migration purposes.[37] Under these circumstances, the city might well decide – and decide reasonably – that the federal government was engaging in actions that were likely to end in the violation of the PNR, and that such violations make legitimate forms of resistance ordinarily prohibited. This justification has, again, some intrinsic normative power, simply in virtue of the fact that this principle reflects a valid norm of moral analysis. But it gains some power, we think, from the fact that this principle also reflects a basic moral commitment *of the state itself.* Insofar as a state respects the basic norms of international law, within which states operate and insist upon their rights as regards other states, it is bound to respect the PNR, and the moral principle making that principle powerful. It is, therefore, perhaps that much more difficult – as a matter of rhetoric, and as a matter of philosophy – for the federal state to rightly condemn those who insist upon a principle endorsed by that state itself.

4 Conclusions

There are still a great many things to be said about this defense of the sanctuary city. We cannot hope to address all these issues in the present context. In what remains, we will instead examine only three potential objections to our conclusions here, which represent worries that frequently occur in the public discourse about sanctuary policies. We will address these, in increasing order of persuasiveness. We will end with a few thoughts about what it is that the argument development here entails.

We can begin with the worry most frequently cited by conservative opponents to the sanctuary city: that it licenses – indeed, supports – lawbreaking, which can only end in a more lawless (and violent) society. Thus, conservative opposition has frequently made reference to the shooting of Kathryn Steinle, who was murdered by an undocumented Mexican citizen released earlier by the sanctuary city policies of San Francisco.[38] Jeff Sessions made reference to Steinle's death in a recent condemnation of sanctuary city policies, arguing that criminals flock to sanctuary cities: "when cities and states refuse to help enforce our immigration laws, our nation is less safe. . . Countless Americans would be alive today – and countless loved ones would not be grieving – if the policies of these sanctuary cities were ended."[39]

There are two immediate rejoinders to these worries. The first is that the empirical contention may not, in fact, be true. Sanctuary cities seem to be, if anything, safer than non-sanctuary jurisdictions.[40] The more interesting issue is what we might say if Sessions' contention were empirically correct. If it were true that a city risked increasing crime rates

[36] See McAdam and Chong

[37] Human Rights Watch, "U. S. Drug Deportations Tearing Families Apart," 2015. Available at https://www.hrw.org/news/2015/06/16/us-drug-deportations-tearing-families-apart

[38] Christina Littlefield, "Sanctuary cities: how Kathryn Steinle's death intensified the immigration debate," *Los Angeles Times,* July 24, 2015.

[39] Jeff Sessions, remarks on sanctuary jurisdictions, March 27, 2017. Available at https://www.justice.gov/opa/speech/attorney-general-jeff-sessions-delivers-remarks-sanctuary-jurisdictions

[40] Christopher Ingraham, "Trump says sanctuary cities are hotbeds of crime. Data say the opposite." Available at https://www.americanprogress.org/issues/immigration/reports/2017/01/26/297366/the-effects-of-sanctuary-policies-on-crime-and-the-economy/

when accepting sanctuary status, would that give it a moral reason to refuse that status? The answer to that question, we believe, is: not generally, no. The fact remains that violations of the PNR involve directly placing people into circumstances in which violations of basic rights are either inevitable or highly likely. A marginal decrease in crime rate within a particular jurisdiction would be an insufficient justification for that violation – just as we could not legitimately engage in torture ourselves, even if that practice were effective at reducing overall crime. There are, to put it simply, some risks that must be taken in the name of justice; crime reduction is not a blank check with which to license harming others.

The second worry, however, is more complex and difficult. Return to the issue of federalism; the question of migration policy is one over which the federal government has exclusive jurisdiction. (The language of Constitutional law refers to the "plenary power" of Congress to make law in this area.) This is a legal fact; but it might be given a moral justification. We need, in disputed areas of policy, to know who is entitled to the last word. We might, that is, separate the substantive rightness of the federal policy, from the procedural benefit to knowing that it is the federal government whose word shall be law. A country in which every subsidiary jurisdiction is able to make migration law is, in the end, a country in which there is no migration law at all – and that fact might have enormous costs, for both migrants and native-born alike.

These are powerful ideas; something like them undergirds the democratic norm that provides us with moral reasons to obey even those laws we take to be substantively unjust.[41] As with this latter norm, though, there are reasons to think our obedience ought to be limited. Our moral reasons to obey even unjust law often reduce to two ideas: we ought to obey that law because there is moral value in the procedure that gave rise to that law, and our disobedience risks destroying that system; and, we have some reason to respect the divergent views about justice of our fellow citizens, given the inevitable diversity of views within any free society.[42] Both of these ideas, though, can be rebutted in cases where the most basic rights are at stake. Even if our disobedience theoretically risks the destruction of the system, we have an obligation to evaluate that risk against the very real risk that a vulnerable person will be subjected to torture. The risk might be less for the system than we think - in many cases, the system can survive quite well in face of disobedience.[43] Humans, however, cannot long survive deprivation of their most basic rights. Further, the respect for diversity entails only a respect for particular visions of how to weight the divergent elements of a conception of justice, or the relation of that conception to the idea of the good; it cannot rightly include the question of whether torture is a good thing or not. It is possible, in short, for even someone committed to federalism to think that a subsidiary jurisdiction is sometimes right to disobey the commands of a jurisdictional superior. To say otherwise risks turning federalism from institutional form into something very much like a fetish; federalism is valuable, we believe, but its value is not absolute, and we should prefer the preservation of humans to the preservation of the federal order. The Constitution is, in short, one source of moral guidance; it is not the only source – and when it legitimates federal authority that demands morally

[41] For a discussion of obedience, see A. John Simmons and Christopher Heath Wellman, *Is There an Obligation to Obey the Law?* (Oxford: Oxford University Press).

[42] Thus, Rawls argues that we ought to restrict the use of public reason to "matters of basic justice and Constitutional essentials"; for other topics, we are bound to respect the majoritarian procedures that may produce results we dislike. See John Rawls, *Political Liberalism* (New York: Columbia University Press, 1989).

[43] Paul Butler's advocacy for jury nullification involves similar ideas.

wrongful actions, we have some permission to refrain from doing what that Constitution demands.

The final worry, however, is the most vexing. It is the simple fact that – given the stipulations we have made above – not all deportation is actually wrongful. A federal government which is willing to violate the PNR is likely also deporting a great many people whose deportation does *not* violate that norm. Not all deportees, to put it bluntly, are facing the deprivation of their basic rights; the category include foreign students whose visas have expired, undocumented citizens from relatively well-off countries of origin, and so on. A sanctuary city, however, declares as a matter of policy that it will not assist the federal government in the task of deporting the undocumented – which includes, of necessity, that the city will no longer help to do a thing that is not, in itself, wrong for that city to do. Does the overbreadth of the policy here give us moral reason to reconsider that policy?

We do not believe so. While it is true that in many individual cases, the federal government would be asking for assistance in the performing of an action that is not wrongful to do, the simple fact is that there is no way to provide that assistance while refraining from offering it in a wider variety of contexts. This is the essential difference between policy and philosophy; while philosophers can offer a variety of nuanced and individualized analyses of moral permissibility, those who design policy must of necessity work with blunt and fairly clumsy distinctions. If those blunt and clumsy distinctions can be expected to include the use of violence against those who have the right to be free from that violence, then we might reply with a demand for more accurate distinctions – or with a refusal to participate in the administration of the policy guided so poorly.

It might be objected, here, that *any* policy – short, perhaps, of open borders – entails *some* risk of wrongfully deporting an individual to circumstances that are morally impermissible.[44] Any real policy of criminal punishment, after all, entails *some* risk of imprisoning the actually innocent. Our response is not to refuse to participate in the process of criminal adjudication – but to insist that we ought to make that process *better,* so as to minimize that risk. Why should we not do the same, in the context of migration, by demanding that the federal government do a better job of deporting – rather than, as we suggest, by withholding cooperation with that government?

The answer, we think, might require us to note that – even in criminal punishment – we seek to introduce rules of bias, so that what errors we produce are the errors we are most willing to countenance. In criminal law, then, we have tended to defend something like Blackstone's ratio – on which it is better that ten guilty men go free, rather than one innocent man be punished – because of our shared belief that wrongful punishment is a worse evil than

[44] One reviewer has expressed the worry that the principle we defend here is incompatible with exclusion at the border. If we are told that a *risk* of wrongful deportation is sufficient to made deportation as a practice impermissible, then why can we not similarly assert that a *risk* of wrongful exclusion is sufficient to impugn the morality of exclusion? To this, we have two broad replies. The first is that this extension may be a legitimate one; we do not here make that extension, but neither do we wish to provide any argument against it, The second is that our focus in the present paper is on the morality of *assisting* federal practices of deportation, and emerges from the fact that the sanctuary city is a site of resistance to those practices; no similar mode of local resistance to unjust border exclusion has emerged, to our knowledge, but what we say here would likely defend such resistance were it to develop. (We might, to speculate, imagine modes by which local governments could resist unjust practices and policies at the border.). Our focus on the sanctuary city does not reflect a moral commitment that the sanctuary city is uniquely justified; rather, it reflects the empirical fact that such cities *exist,* and require moral analysis. We are not thereby committed to the claim that other modes of legitimate resistance might emerge. We are grateful to this reviewer for pressing us to be precise on this point.

wrongful freedom. Something similar might hold true in the context of migration – in which wrongful non-deportation is, for similar reasons, a much less grave evil than wrongful deportation to one's death. Where a sanctuary city is unable to affirm that the federal government is, in fact, obeying something like Blackstone's ratio in its deportation practices – where it is tending to over-deport, as it were, rather than under-deport – then that sanctuary city might be permitted to say, based upon the morality of the PNR, that it will not lend its affirmative energies to the task of deportation. Nothing here, of course, prevents that sanctuary city from urging the federal government to reform its practices of deportation; and, indeed, were that government to do so, then the rationale we defend here might cease to apply. It is our belief, however, that such a blessed event is unlikely to take place within our immediate political future.[45]

We will conclude by making explicit something that has been implicit in what has gone before: We defend the potential legitimacy of the sanctuary city, not simply as a practice of refusal to act, but as a way in which the city can *act* – and can act in defense of its most vulnerable residents' basic rights. What we offer here could be used to defend not simply the refusal to help the federal government in achieving its policy goals, but in the active frustration of those goals. There are, of course, moral limits on what sorts of policy might legitimately be deployed by a city in response to morally wrongful federal policy. The contours of these limits, however, we leave for another occasion. In the present context, we will close simply by noting that what we say here might conceivably ground not only what sanctuary cities currently do, but potential innovation about how such cities might respond to widespread injustice.

References

Allais J (2002) "The *Jus Cogens* Nature of Non-Refoulement." *International Journal of Refugee Law*, Vol. 13, No.4

Amnesty International (2016) "Home sweet home? Honduras, Guatemala, and El Savador's role in a Deeping refugee crisis." Available at:https://www.amnestyusa.org/files/central_american_refugees_-_report_eng_1-min.pdf

Derrida J (2001) On cosmopolitanism and forgiveness. Routledge, London

Doggett T (2011) Recent work on the ethics of self-defense. Philos Compass 6(4):220–233

Frowe H (2016) The ethics of war and peace: an introduction, second edition. Routledge, NY

Frowe H (2014) Defensive Killing. Oxford University Press, NY

Human Rights Watch (2015) "U. S. Drug Deportations Tearing Families Apart." Available at: https://www.hrw.org/news/2015/06/16/us-drug-deportations-tearing-families-apart

McAdam J, Chong F (2014) Refugees: why seeking asylum is legal and Australia's policies are not. UNSW Press, Sydney

McMahan J (2005) The basis of moral liability to defensive killing. Philosophical Issues 15(1):386–405

Rawls J (1971) A theory of justice. Harvard University Press, Cambridge, MA

Rawls J (1989) Political liberalism. Columbia University Press, New York

Statman D (2008) On the success condition for legitimate self-defense. Ethics 118(4):659–686

Thomson JJ (1991) Self-Defense. Philos Public Aff 20(4):283–310

[45] We would note, finally, that nothing we say here depends upon the threat to the migrant being man-made in origin. Deportation to a country that is unable to support human life is just as wrongful as deportation to one in which human life is not political respected. We are grateful to an anonymous reviewer for this journal for urging us to be explicit about our answer here.