Against Moral Taint

Yitzhak Benbaji & Daniel Statman

Ethical Theory and Moral Practice
An International Forum

ISSN 1386-2820

Ethic Theory Moral Prac
DOI 10.1007/s10677-020-10130-y

Springer

Ethical Theory and Moral Practice
https://doi.org/10.1007/s10677-020-10130-y

Against Moral Taint

Yitzhak Benbaji[1] · Daniel Statman[2]

Accepted: 12 October 2020/Published online: 19 October 2020

Abstract
One motivation for adopting a justice-based view of the right to self-defense is that it seems to solve the puzzle of how a victim may kill her attacker even when doing so is not predicted to protect her from the threat imposed upon her. The paper shows (a) that this view leads to unacceptable results and (b) that its solution to cases of futile self-defense is unsatisfactory. This failure makes the interest-based theory of self-defense look more attractive, both in the context of futile self-defense and in general. To understand how a victim might use force in this context, one need only point to some interest of hers that is threatened, and the best candidate for such interest in cases of futile self-defense is her honor.

Keywords Moral taint · Self-defense · Øverland · McMahan

> And Cain said to the Lord: 'My punishment is too great to bear. Now that you have driven me this day from the soil and I must hide from your presence, I shall be a restless wanderer on the earth and whoever finds me will kill me'.
> (Genesis 4, 13–14)

1 Introduction

In situations of (legitimate) self-defense, a potential victim ('Victim') carries out an otherwise immoral act against her attacker ('Aggressor') in order to block an unjust attack against her.

✉ Daniel Statman
dstatman@research.haifa.ac.il

Yitzhak Benbaji
ybenbaji@gmail.com

[1] Tel Aviv University Law School, Tel Aviv, Israel

[2] Department of Philosophy, University of Haifa, Haifa, Israel

Thus, one necessary condition for the justification of a self-defensive act is its success in achieving this goal, namely, in preventing the relevant assault. What follows is –

> *Prevention:* Self-defensive harm is permissible only if it can prevent (or can reasonably be believed by Victim to prevent[1]) some unjust threat.

Prevention seems hard to deny, but it gives rise to a disturbing puzzle which was discussed, separately, by each of us (Benbaji 2005; Statman 2008). The puzzle grows out of examples in which Victim seems clearly justified in X-ing Aggressor (when 'X-ing' refers to any otherwise immoral act against Aggressor) though X-ing does not achieve defense (and cannot reasonably be believed to achieve it). John Wayne is permitted to shoot at his hunters even if they outnumber him and he is doomed anyway. His shooting might eliminate the threat from some hunters, but harm would be wreaked upon him by others, hence, on the face of it, he seems to gain nothing by the shooting. Similarly, a woman is permitted to wound her assailant even if doing so will not save her from being raped by him.[2]

The two cases are slightly different: In the rape case, Victim can neither prevent the harm, nor prevent the threat from the particular aggressor. Her X-ing seems completely pointless, in clear opposition to *prevention*. By contrast, in a narrow sense, Wayne's action would satisfy *prevention* – he would be blocking the threat posed to him by some specific aggressors, but given that he would fail to prevent the *harm* (he will be killed anyway by one of the other aggressors), it is hard to see how his otherwise immoral action could be justified.

In the papers mentioned above, we both argued that the key to understanding this permission lies in recognizing the importance of honor to the person whose honor the Aggressor threatens. When Aggressor tries to rape or kill Victim, he humiliates her by treating her as a means to his ends. It is her interest not to be so treated that Victim defends when she carries out seemingly futile acts of self-defense against Aggressor. While her defensive act is futile in protecting other interests (in life or in bodily integrity), it is not futile in protecting her honor. In particular, although Victim cannot save her life, she can at least avoid a *disrespectful death*.[3]

In response to the line taken by us, it has been argued that in the cases described above, the protection of honor cannot explain how Victim has a right to kill Aggressor.[4] As an alternative explanation, some have suggested anchoring this right not in facts about Victim (the threat to her honor), but in facts about Aggressor (his reduced moral status).[5] The debate is important beyond the context of futile self-defense as it brings to the fore a fundamental disagreement about the right to self-defense. Some, like Thomson (1991), believe it is based on Victim's right to protect her interests simply because they are unjustly threatened, while others, notably McMahan (1994a), believe it is based on the permission (and sometimes the duty) to enforce

[1] These two formulations refer respectively to an objective vs. a subjective view of self-defense. For the purpose of the present paper we need not decide between these views.

[2] In more technical terms, the puzzle concerns what Statman calls "the success condition" for justified self-defense. The condition is familiar from just war theory, i.e. from the requirement to wage war only if it has a "reasonable hope of success." The puzzle, then, is how to justify acts of self-defense which seem to violate the success condition, acts which fail – and are expected (in the descriptive sense of the term) by Victim to fail – in providing defense.

[3] Benbaji (2005), 586. The honor solution has attracted quite a lot of philosophical attention. See, for instance, Frowe (2014), 109 ff.; Bowen (2016); Robillard (2017); Ferzan (2018a, 2018b); and Husak (2018).

[4] See Uniacke (2014); Ferzan (2018a, 2018b); Husak (2018); Arneson (2018).

[5] See Øverland (2011a), hereafter 'DF', and Øverland (2011b), hereafter 'MT'.

justice in the distribution of harm. The former will be friendly to the honor solution to the problem of futile self-defense while the latter less so.

The main purpose of our paper is to revisit the philosophical debate between the interest-based and justice-based conceptions of self-defense from the angle of the problem of futile self-defense. We will analyze Gerhard Øverland's attempt to solve the problem of futile self-defense by developing a version of a justice-based view of the right to self-defense. According to this view, a culpable aggressor deserves less protection than others, hence it is permissible to harm him even for the sake of a negligible benefit. Our rejection of this view owes a lot to Rawls's idea that conceptions of desert and moral character should not be employed in developing the principles of justice. It is this Rawlsian idea that Øverland is implicitly denying when he assumes that, in exercising his right to self-defense, Victim can and should "play God"; compare the overall moral character of Aggressor with that of Victim. In Rawls's view, only God can carry out such comparisons and only He has the moral authority to do so. We will argue that *Justice* is to be rejected in favor of the alternative view, to wit, *the interest-based view of self-defense*.

We shall proceed as follows. In Section 2, we shall present two arguments against the honor solution which have led Øverland to opt for an alternative solution based on the idea of *moral taint*. We shall present this solution in Section 3 and show that it leads to unacceptable implications. Section 4 will continue the critical discussion of the taint solution, arguing that it involves the sin of "playing God." In Section 5, we return to the honor solution and defend it from the arguments raised in Section 2, thus strengthening the interest-based view of (the right to) self-defense against its rival, justice-based view. In Section 6, we end with a few concluding remarks.

2 Objections to the Honor Solution

In Øverland's view, honor cannot play the role assigned to it in our accounts. This is because honor fails to explain how Victim might be permitted to kill Aggressor in cases in which Victim herself does not care about her honor, or doesn't believe that killing Aggressor will redeem it. Consider the case of Clint. Like Wayne, Clint is hunted by bad guys wishing to kill him, and he cannot save his life by shooting at them. However, unlike Wayne, "Clint does not believe in the idea of honor, and is considering not killing anyone in vain. Eventually succumbing to boredom and hunger, Clint draws his gun, takes aim, and fires, killing three attackers" (DF, 245). Since Clint does not care about his honor, he cannot be said to have acted *in order to defend it*. He similarly cannot be said to have acted *in order to save his life* because in the circumstance this is not a realistic possibility. Nevertheless, Øverland argues, surely Clint does have a right to fire at his attackers. Hence, the justification of this act has nothing to do with the defense of honor, and probably *not with any other interest of Clint's*. Rather, it has to do with facts about his attackers.

Another argument against pointless self-defense is based on the assumption that killing Aggressor in such circumstances is not the least harmful way of effectively defending Victim's honor and is therefore ruled out by the necessity condition. If Clint wants to avoid a disrespectful death at the hands of his hunters, or to reduce the attack on his honor, he could do so "by turning to face the bunch and firing his gun in the air, without killing anybody." This, so Øverland believes, "would be a simultaneous show of generosity and contempt" (DF, 243). To those who argue that such a reaction would be pathetic and would just reinforce

Clint's helplessness, Øverland says that "what matters for saving one's honor is the attitude you have toward what's facing you, not whether you are able to get in a few futile shots" (DF, 244, n.18).

Although Øverland concedes that Victim has a right to attack Aggressor even if that will not save her from the threat posed by him, he thinks that the honor solution fails in substantiating this right. And this failure is telling because – in Øverland's view – it suggests a more general lesson about the right to self-defense, namely, that to justify self-defense we must look beyond the interests of victims to facts about the morality of their attackers. What facts should be looked at and how they figure in an alternative theory of self-defense will be discussed in the next section.

3 Futile Self-Defense and the Idea of Moral Taint

If X-ing Aggressor cannot defend Victim from the harm that Aggressor culpably imposes upon her, how could Victim be justified in X-ing him? One answer might be that X-ing Aggressor is *punishing* him. Øverland rejects this answer claiming that one has no right to go around and "dish out punishment" on whoever deserves it (DF, 254). Instead, he proposes what he calls the "lesser-claim-for-protection-view." Aggressor may be harmed by Victim not because he deserves the harm (though he might), but because, as a result of his culpable attack, his claim for protection is weakened. While, on retributive accounts, there is intrinsic value in making Aggressor suffer for his wrongful behavior, hence a moral gain in doing so, on the lesser-claim-for-protection-view, such gain need not be assumed, and, in any case, is not the basis for the permission to X him. The proper goal of X-ing Aggressor is rather "to realize a more just distribution of costs" (DF, 254). Since Aggressor culpably created a situation in which some cost must be borne, justice requires that *he* ought to bear it rather than Victim. Judgments regarding self-defense against culpable aggressors are essentially comparative. When it is said that Aggressor is liable to bear the cost of preventing the attack on Victim, what is meant is not that he is liable *simpliciter*, but that he is liable to bear this cost more than Victim is.[6] We shall refer to this view as *justice*. In unpacking the right to self-defense in terms of distributive justice, Øverland follows McMahan who is the chief defender of this understanding.[7]

According to *justice*, the more culpable Aggressor is, the greater the share he ought to shoulder in order for his threat to be neutralized. If he is *fully* culpable, he loses (almost) all the moral protection to which he would otherwise be entitled, and is vulnerable to (almost) any defensive act by Victim, however small the benefit to her. In Øverland's view, this is what explains the permissibility of killing the aggressors in the Wayne and in the rape cases mentioned at the outset. Even when the expected gain for Victim from such killing is negligible – there must, after all, be *some* gain for him/her! – the killing is nonetheless permissible because by culpably attacking their victims, both the rapist and Wayne's hunters lose their claim for protection. In a morally ordered world – a world regulated by *justice* – those with a lesser claim for protection will enjoy a weaker protection in dilemmas concerning life and death than those with a stronger claim for protection. Thus, a world in which the rape victim injures the rapist is more just that a world in which she simply surrenders to him even if

[6] On the distinction between deserving X and being liable to X, see McMahan (2009), 8–9.

[7] McMahan (1994a). This is how McMahan justifies killing in war as well. See McMahan (1994b) and McMahan (2005). For a critique of this approach, see Lazar (2009).

the hurt she causes fails to prevent the rape. And a world in which Wayne kills some of his hunters is more just that a world in which he doesn't because his response would be fitting to their very weak claim to protection resulting from their unjust attack against him.

The common understanding of the right to self-defense contends that the permission to X a culpable aggressor has a fixed direction: X-ing is allowed only when it comes from Victim (or a third party acting on her behalf) and is directed against Aggressor. Following Øverland (who, as we'll see immediately, rejects this principle), let's call it *directionality*. According to *directionality*, if Aggressor culpably poses an unjust threat to Victim, then whereas Victim may harm Aggressor in order to prevent *this* threat from materializing, she is not allowed to harm Aggressor in order to defend herself from *other* threats, to which Aggressor has no causal connection. A fortiori, *other* (potential) victims are not allowed to harm Aggressor in order to save *themselves* from whatever misfortunes might befall them (unrelated to Aggressor) if they don't do so. The permission granted to third parties is consistent with *directionality* because they act on Victim's behalf. Thus, self-defensive measures go in only one direction – from a specific victim (and those who act on her behalf) to a specific aggressor, namely, the person who created the threat to her and whose threat can be blocked by X-ing him.

The problem is that, as far as *justice* is concerned, the causal restriction which underlies *directionality* seems ad hoc. It is hard to understand why culpable attackers are liable to defensive harm *before* committing a murder, whereas once they succeed in committing it they are no longer liable and regain their immunity. Why limit the relevant comparison to that obtaining between Victim and her direct aggressor? Why doesn't the culpability of some bystander (who might be a bystander with regard to the current threat but culpable with regard to a different threat to a different victim) tip the otherwise balanced scales of one life versus another? *Justice* seems to have no resources to explain such limitations on defensive force.

This tension between *justice* and *directionality* has not gone unnoticed. After arguing that the right to self-defense is based on the idea of a just distribution of burdens between Aggressor and Victim, McMahan goes on to suggest the following:

> There are, however, ways in which this core might be extended. It is possible, for example, that it can be permissible to attack someone who unsuccessfully attempts to create an unjust threat if the same threat, or a similar threat, then arises independently of this person's action, and the threat can be averted only by attacking him. Or perhaps, in order to avert an unjust threat, it can be permissible to attack someone who is not responsible for the threat but who would have created the threat, or would now create a similar threat, if he could. There is in fact a spectrum of possible bases for liability ranging from the possibilities just noted to responsibility for a different unjust threat of the same type in the past or present, responsibility for a different type of threat in the past or present, being willing or disposed to create an unjust threat, possession of a bad moral character, and so on. (McMahan 2004, 722)

All these extensions grow naturally from the assumed moral asymmetry between Victim and Aggressor, an asymmetry which is supposed to explain why Victim's life may be preferred to Aggressor's when a forced choice between their lives occurs. If the fundamental criterion for distributing burdens between Victim and Aggressor is their relative moral records, why should this apply only to the standard case of self-defense and not to any situation in which we need to decide which of two (or more) people would suffer some misfortune? Once the right to self-defense is "moralized" in the sense of being grounded in the aim of realizing a more just

distribution of burdens, and once this distribution is based on the moral quality of the relevant parties, *directionality* looks arbitrary.

Øverland too acknowledges the incompatibility of *justice* with *directionality* and since he is committed to the former, he rejects the latter and offers in its stead *transferability* (MT, 127) which can be formulated as follows:

> *Transferability*: The lesser claim to protection that a person is entitled to as a result of his culpability for an unjust attack transfers to any situation in which his interests conflict with those of others, in particular his interest in not being harmed.

In this vein he concludes that Victim may kill Aggressor if necessary to save her life even if the threat that Aggressor posed no longer exists (*Incapacitated Aggressor1*, DF 254). Moreover, Victim may kill Aggressor if necessary to save her life if it turns out that Aggressor culpably killed some *other* person (*Window*, MT 127). *Transferability* also implies that Aggressor has a lesser claim to be *saved* when in trouble. If Victim survives a failed attempt by Aggressor to kill her and later realizes that Aggressor is drowning in a pond and she can save either him or some innocent third party, then she ought to save the latter rather than Aggressor.

Transferability further implies that years after Aggressor culpably posed an unjust threat to somebody, it might still be morally permissible to kill him in order to save the life of some innocent person whose life is under threat from a different aggressor, in circumstances in which the new (potential) victim is unrelated to Aggressor, to the original aggression, or to Victim. This looks rather far-fetched and Øverland's response is to limit *transferability* to cases occurring "more or less immediately" (MT, 123) after the initial aggression takes place. But this limitation seems ad hoc. If *justice* need not assume a causal connection between Aggressor and the current threat to Victim, it need not assume a temporal or spatial connection between them either.

Øverland offers two arguments to deal with this difficulty, yet neither is convincing. First, he says, "a wrongdoer may repent" (DF, 258), and repentance has the power to erase moral taint and restore the wrongdoer's moral status. Second, "time itself" has "a tendency to restore people's moral status" (DF, 258) because over time identity changes. A taint on a twenty year old aggressor no longer exists when the aggressor is fifty or sixty. Metaphysically speaking, the sixty year old is a different person, so the taint does not transfer.

However, if by 'repentance' one means a reform of moral character (ibid.), then Øverland must show that most wrongdoers undergo such a reform within ten years or so of their wrongdoing, which is a rather shaky empirical assumption. As for the argument from personal identity, even if it is accepted, it applies only to cases occurring long after the initial aggression. Whatever his conception of personal identity, Øverland would surely concede that a person might deserve punishment for a crime she committed ten years earlier, even though she is not, strictly speaking, the same person. By the same logic he should admit that, according to *justice*, a person might have a lesser claim for protection because of the way he behaved five or even ten years earlier.

We conclude that by denying *directionality* and endorsing *transferability*, Øverland is committed to the view that if once A posed a serious and unjust threat to B, that makes him morally tainted, which means that whenever he happens to be in circumstances in which killing him is necessary to save an innocent life – or even the limb of an innocent person – doing so would be permissible.

What sort of culpability might justify the use of lethal measures against Aggressor? Øverland assumes that culpability comes in degrees and he argues that the threshold above

which a person starts to lose his claim for protection is negligence. If the aggression is reckless, the claim is further weakened and if it is intentional, even more so (MT, 126). In standard cases of self-defense this seems pretty intuitive. Consider –

> *Reckless driver1*: If a pedestrian, Alice, faces a serious threat of injury by a reckless driver, Bill, and she can defend herself only by killing Bill, she has the right to do so.

However, once *transferability* enters the picture, we are led to rather counter-intuitive results. Let's start with –

> *Reckless driver2:* Bill's reckless driving endangers Alice's leg. She takes out her gun to shoot him, but, fortunately, his reckless behavior turns out to be inconsequential: Bill misses her, and Alice does not have to shoot him in self-defense. A year later, Alice faces a threat of leg injury by Cathy's reckless driving, a risk that can only be prevented by killing Bill. Bill is a passenger in Cathy's car and killing him is the only way to divert the car from its path.

Since "the culpability factor maintain[s] its significance in absence of [causal] contribution, as when there has been a mere attempt to do wrong" (MT, 131), Bill would have a lesser claim for protection even though he caused Alice no harm at all. Given *transferability*, there is no reason why Alice would be allowed to defend her leg with lethal measures in *reckless driver1* but not in *reckless driver2*. But why limit the right to kill Bill to the direct (potential) victims of his recklessness? Consider –

> *Reckless driver3*: Bill's reckless driving endangers Alice's leg. Fortunately, at the last minute, Bill's car misses her. A year later, David faces a threat of leg injury by Cathy's reckless driving that can only be prevented by killing Bill. Bill is a passenger in Cathy's car and killing him is the only way to divert the car from its path and save David's leg.

On the combination of *justice* and *transferability*, David – or anybody acting on his behalf – would seem to have a right to kill Bill if that's the only way to save David's leg or the leg of another.

> A further complication generated by *transferability* can be illustrated by –
> *Reckless driver4*: Bill's reckless driving endangers Alice's leg. Fortunately, at the last minute Bill's car misses her. A month later, it is Alice who drives recklessly and, by sheer coincidence, endangering Bill's leg. Fortunately, Bill is not hurt either. A year later, Alice's leg is in danger by some car speeding in her direction. Bill is a passenger in this car and killing him is the only way to divert the car and thereby save her leg.

Alice, who is the (potential) victim in one incident happens to be the aggressor in another, and the same with Bill who is the aggressor in the first incident and the victim in the second. In *Reckless driver4*, it seems, then, that *justice* would not permit Alice to save her leg in the third encounter by killing Bill because when deciding about the "just distribution of costs" (DF, 254), she needs to take into account not only the immediately previous incident but the one preceding it as well. And when the two past incidents are taken into account, they seem to cancel each other out, leaving Alice with no basis for arguing that her claim for protection is stronger than Bill's (and thus justifies killing him in order to save her own self).

But, on reflection, why limit such moral calculation to the few contingent past encounters on the road between Bill and Alice? And why must all factors that feed into this calculation refer to the same kind of threat (or harm), e.g. threat to injure another as a result of reckless

driving? If what needs to be determined is who has a stronger claim for protection, a much more ambitious principle seems to be called for. Consider –

> *Reckless driver5*: Bill's reckless driving endangers Alice's leg. Fortunately, at the last minute, Bill's car misses her. Although Bill was clearly reckless in this incident, his recklessness on this occasion is the exception to his extraordinary positive moral character and record of good deeds. By contrast, Alice is a nasty and aggressive woman who volunteers on a regular basis for a local neo-Nazi group. A year after the original incident, Alice's leg is again in danger, this time from Cathy's car. Bill is a passenger in her car and killing him is the only way open to Alice to divert the car and thereby save her leg.

Surely in these circumstances, *justice* should not permit Alice to kill Bill in order to save her leg. Notwithstanding his one-time reckless driving, Bill's moral record is much better than that of Alice and hence entitles him with a stronger claim for protection.

The same conclusion follows if, instead of Alice, a new actor enters the scene:

> *Reckless driver*6: Bill's reckless driving endangers Alice's leg. Fortunately, at the last minute Bill's car misses her. A year later, David can save his leg only by killing Bill. Notwithstanding Bill's one-time reckless driving and the threat it then posed to Alice's leg, his moral record is much better than David's.

If we aim at a just distribution of costs, we should not allow David to kill Bill. Moreover, it is *Bill*, the reckless driver in the first encounter, who might be allowed to kill David in the second encounter if circumstances occur in which this is the only way to save his (i.e. David's) leg.

Note further that this analysis applies to former victims no less than it does to former aggressors. Consider Øverland's *Incapacitated Aggressor1* (DF, 254–5). At t1, Aggressor tries though fails to kill Victim. Shortly afterwards, at t2, Victim can save herself from a new threat only by severely injuring Aggressor. According to *justice*, she's permitted to do so. But now suppose it turns out that at t0, it was Victim who tried but failed to kill some third person. Or that although she never tried to kill anybody, she led a life full of deceit, exploitation, cruelty and so on down the list of vices. Surely this should be relevant in determining who has a stronger claim for protection at t2.

Under *transferability*, then, a person who was a culpable aggressor at t1 has a lesser claim for protection at t2 (when a forced choice occurs between his life and that of others), whether or not his culpable action at t1 bears causal connection to the forced choice at t2. The basis for this reduced protection is Aggressor's inferior moral status vis-à-vis the other person *at t2*. But in determining moral status, it would be absurd to refer only to Aggressor's wrongful behavior at t1, or only to behavior that involved the same kind of harm that is about to befall the other person at t2, e.g. death. We must also refer to other wrongful behaviors of Aggressor, as well as to past wrongful behaviors of the other person(s). What follows is:

> *Entire Moral Record*: When deciding who, among *n* parties, has a stronger claim for protection, the entire moral records of all parties must be taken into consideration.

To appreciate just how radical *Entire Moral Record* is, consider the following:

> *Good aggressor vs. villainous victim*: Bill is a sweet man who never did any harm to anybody. He volunteers for many good causes and donates a lot of money to Oxfam. Most probably his donations along the years have significantly improved the lives of

many children in the third world. By contrast, Alice is a nasty and aggressive woman who volunteers on a regular basis for a local neo-Nazi group. One day, Bill sees Alice in some neo-Nazi march, loses his temper and attacks her, posing a real threat to her life. Alice is well-trained and carries a knife for self-defense.

Does Alice have a right to kill Bill in order to defend herself? According to *Entire Moral Record*, while Bill's record shines like a star, except for one stain resulting from his current attack on Alice, Alice's record is full of filthy stains of different sizes and shapes, with very few white areas. Thus, the most just distribution of harm would be realized by allocating the relevant cost to villainous Alice rather than to good Bill. Better that saintly Bill lives and villainous Alice dies than the other way round.

To recap. One source of motivation for accepting *justice* has been the hope that it can solve the puzzle of futile self-defense. According to this solution, the right to self-defense in such cases is not anchored in Victim's interest in saving her honor, but in the stronger claim for protection that she has in comparison to that enjoyed by Aggressor. *Justice* demands that if A has a stronger claim for protection than B and one of them must die, then it is B who should die rather than A. What we showed in this section was that *justice* together with *transferability*, lead to unacceptable results.

4 Playing God

In some of Øverland's cases, it is a past victim who is entrusted, so to say, with the task of promoting justice. It is she who acts in such a way that the person with a stronger claim for protection, namely herself, gets priority over the person with a lesser claim, namely, her past attacker. In other cases, it is some third party who was never touched by Aggressor, but who, for instance, happens to pass near a pool in which (past) Aggressor and (past) Victim are drowning. According to the normative story told by Øverland, aggressors are entitled to less protection from *any* human being, hence anyone is permitted – probably *required* – to make sure that they receive less protection than others, in cases of forced choices between their lives (or other vital interests) and those of others. Similarly, any person is permitted – again, probably *required* – to consult the full moral record of the relevant parties when considering which of them gets priority when their lives are in danger and resources are limited. Our strategy in the previous section was to show that this leads to unreasonable results. What we wish to do in the present section is to locate our disagreement with Øverland in the larger debate on the role of desert in the context of self-defense and in that of punishment.

Recall that, according to Øverland, the individual right to self-defense is close to the right (and duty) of states to punish wrongdoers, as the latter is usually understood in liberal societies. On this understanding, the fact that a person deserves a certain harm is necessary but insufficient to impose it upon him. What is also necessary is that this harm – this *punishment* – contributes to the prevention of some future (and undeserved) harm. Wrongdoers go to jail not only because they deserve it, but because imprisonment prevents further wrongdoing (by them and others). Similarly, in the series of *Reckless Driver* cases, Bill may be killed not only because he deserves such fate (or because he has a weaker claim for protection), but because doing so would prevent serious injury or loss of life from befalling Alice.

The anti-*justice* view that we defended here owes a lot to Rawls whose political liberalism makes one's desert irrelevant to the scope of one's right to self-defense. Rawls does not deny

the idea of moral desert in general. He merely denies that desert can "be incorporated into a *political* conception of justice" (Rawls 2001, 73). In his view, the right to self-defense, like other moral rights that a just society legalizes and secures, is independent of any reasonable conception of the good which addresses the issue of desert. Rawls offers two reasons for this independence. The first is contractarian: "Having conflicting conceptions of the good, citizens cannot agree on a comprehensive doctrine to specify an idea of moral desert for political purposes" (ibid.). Principles that apply to the basic structure of society must be endorsed by reasonable people who do not know what their conception of the good will be.

The other reason concerns practicality. Moral worth would be impracticable as a criterion when applied to questions about the nature and scope of the right to self-defense. This is because, "only God could make those judgments. In public life we need to avoid the idea of moral desert and to find a replacement that belongs to a reasonable political conception" (ibid.). If only God can respond to considerations of desert, then insofar as the right to self-defense is an element of the morality that states ought to build into their basic structure, it must be desert-independent. It is to be grounded in the interests that free and equal people have independently of their conception of the good.

Three distinct but related issues manifest themselves when one commits the sin of Playing God. The first is that human beings don't have the epistemic capacity to determine the full moral record of others – or even of themselves – and definitely don't have the capacity to make reliable comparisons between the full moral records of different people. *Entire Moral Record* is, therefore, based on the illusion that we can transcend our epistemic limitations. Even in the case of a fully culpable aggressor currently attacking us we can't know for sure that his entire moral record (one that takes into consideration all past intentions, attempts and actions) is worse than ours and hence that he has a lesser claim to protection and may be killed in self-defense. All the more so in less paradigmatic cases of the right to kill in self-defense.

Second, and mainly because of these epistemic shortcomings, the right to play God and determine who shall live and who shall die is almost sure to be abused. Either intentionally or – more likely – unintentionally, agents will be prone to judge their own lives and the lives of those they especially care for (their kin, their friends, members of their national/religious/racial group) as having a stronger claim for protection than the lives of others when choices must be made about who will live and who will die.

Third, even if we had the epistemic tools to determine that if we did not intervene then Bill, who enjoys a lesser claim for protection would end up alive, while Alice, who has a stronger claim, would end up dead, we simply wouldn't have the moral authority to intervene. *Entire Moral Record* assumes an authority we don't have, viz., an authority to advance justice by actively harming people who assumingly have a lesser claim for protection in order to benefit others whose claim for protection is thought to be stronger. The point is that the very fact that some harmful action X promotes justice is insufficient to justify an agent in X-ing; she must have *authority* to actively inflict such harm. Her decision to kill A in order to save B in cases like *Reckless driver* will rightly give rise to the kind of complaint leveled against Moses when he intervened, with no authority, in a conflict between the two Israelites: "Who made thee a prince and a judge over us?" (Exodus 2, 14).

Rawls's theory of justice encapsulates all these issues by claiming that the interests that the just society secures by conferring rights are those shared by all members of society, independently of their comprehensive doctrines. Since they all share the interest of not being unjustly killed, injured or humiliated, they endorse a right to self-defense against such threats. Since they do *not* share the same conception of desert, this notion is not part of the basic structure. In

Rawls's view, then, which is consistent with the one proposed here, individual desert (as cashed out by *Entire Moral Record*) is irrelevant to determining whether or not one is liable to defensive action.

Note how Rawls's view is echoed in Victor Tadros's objection to relying on desert and retribution in justifying punishment. According to Tadros, retributivists believe that "a person's well-being, or some component of it, over the course of her whole life must track the quality of the person or her actions over the course of her whole life" (Tadros 2011, 61) – a belief that underlies *Entire Moral Record* as well. Yet, he goes on to argue, "this view cannot vindicate a plausible set of state institutions" (ibid.). The current paper has sought to show that the same holds for the moral taint justification of the right to self-defense.

Interestingly, Øverland seems to assume something like our playing God argument when *retributive* justice is concerned. As we saw above, he explicitly rejects the idea that one has a right to go around and "dish out punishment" on whoever deserves it. Although he doesn't elaborate on the basis for this claim, we trust that it has to do with precisely the reasons just mentioned, namely, (a) as individuals, we rarely have enough evidence to warrant judgments regarding the hardships that people deserve due to their past wrongdoing, (b) a right to "dish out" punishment is sure to be abused, and (c) even if we had the required knowledge about desert and were immune from the danger of abuse, we would still lack the authority to go around dishing out punishment.

In response, Øverland might have argued that while we don't have the authority to go around and impose punishment on whoever deserves it, we do have the authority to make sure that, in the domain of life and death decisions, those who have a stronger claim for protection get priority over those who have a lesser claim for protection. However, this distinction between (the right to realize) retributive and (the right to realize) distributive justice looks ad hoc. The thought that individuals have a right to act in order to realize a more just world is vulnerable, in both spheres of justice, to the three difficulties listed above.[8]

A different response to the Rawlsian objections to *Entire Moral Record* would rely on the familiar distinction between the question of whether some ethical theory is successful in identifying the ultimate right- and wrong-making properties of actions and the question of whether it provides the best decision procedure to do the right thing. It is often argued, for instance, that while act utilitarianism might offer the correct account of what makes actions right or wrong, it is hopelessly difficult to apply in real world dilemmas. In a similar vein, one might suggest that the difficulties in applying *Entire Moral Record* in the real world given our cognitive limitations are insufficient by themselves to discredit it.

Although there is nothing logically incoherent about this answer, it leads to the odd result that when faced with self-defense dilemmas in the real world, for example in circumstances of futile self-defense, the practical advice given by McMahan, Øverland and their followers would be no different than that given by supporters of the interest-based view of self-defense.

[8] One might suggest that only the state has authority to punish, which explains why although individuals may not dish out punishment on wrongdoers they are allowed to rely on justice-based considerations in forced choices between lives. This view of punishment has a long history; see, for example, the Biblical law that forbids revenge by family members and demands instead that the killer escape to one of the cities of refuge "until he stand before the congregation in judgment" (Numbers, 35:12, KJV translation). But, first, Øverland does not indicate that he assumes this view of punishment. Second, the spirit of his argument does not seem to welcome it. And third, thinkers close to Øverland, like McMahan, explicitly concede that in some contexts individuals may refer to punishment as one of their considerations to justify harming others. See Ferzan (2018a, 2018b), esp. pp. 274–276. Ferzan herself shares this view, saying: "Like Steinhoff and McMahan, I believe individuals can inflict punishment in some cases" (276).

In other words, if, in real-life situations, Øverland does not expect Victim to consult *Entire Moral Record* in order to decide whether or not to harm Aggressor, it remains unclear what he *does* expect her to do. In the utilitarian context, it is sometimes assumed that the best strategy to maximize utility is for people to rely on *non*-utilitarian reasoning, including various deontological constraints. The analogy in the present context would be an admission on the part of *justice* supporters that, for practical matters, Victim need not concern herself with the entire moral record of the relevant parties. Rather, she may act against Aggressor to defend her interests – including her interest in maintaining her honor. But at this point it is no longer clear that *justice* offers anything much different from the honor solution to the problem of futile self-defense.[9]

5 An Alternative to *justice*

The rejection of *justice* lends support to its main alternative which is an interest-based theory of (the right to) self-defense a la Thomson (1991). This is not the place to offer a full defense of it. Let us just say that, according to this theory, an agent who exercises the right to self-defense does not act as an agent of justice. Her permission to harm Aggressor follows from the brute fact that Aggressor violated Victim's (Hohfeldian) claim against him not to threaten her fundamental interests. It is Aggressor's violation of this claim that generates a permission (a Hohfeldian liberty) for Victim to harm Aggressor, at times even to kill him. Third parties intervening to harm Attacker in Victim's defense cannot be described as making the world a more just place, but simply as acting on behalf of Victim to defend her interests. Thus, the interest theory of rights neatly explains why Victim is allowed to kill Aggressor in the cases mentioned at the outset. What seems as futile self-defense is not really futile but a promising way of protecting Victim's honor. When Wayne shoots at his hunters, he thereby defends his interest in a respectful death.

The Rawlsian test that the honor view should pass is simple. Does the right to futile self-defense, as the honor view interprets it, protect an interest that all reasonable people share, whatever their comprehensive doctrine? Since Øverland's objections to the honor view (mentioned in Section 2) suggest a negative answer, this is a fitting time to return to them. As you recall, these objections were supposed to motivate acceptance of *justice* with its notion of moral taint which were supposed to ground a better solution to the puzzle of pointless self-defense.

The first objection relied on the case of Clint who – just like Wayne – is hunted by bad guys, but who doesn't believe in the idea of honor and, therefore, cannot be said to act in order to defend his honor. Nonetheless, Clint, like Wayne and the rape-victim, has a right to shoot at his hunters. Therefore, argues Øverland, this right can't be based on Clint's interest in his honor. In Rawlsian terms, what Øverland assumes in this objection is that people's interest in honor is subject to reasonable disagreement and therefore does not belong to the basic structure of society. Our view is different though we cannot fully argue for it here. In our view, honor belongs to the "social bases of self-respect, understood as those aspects of basic institutions normally essential if citizens are to have a lively sense of their worth as persons and to be able to advance their ends with self-confidence" (Rawls 2001, 58–59). Thus, if Clint does not

[9] The analogous argument against act-utilitarianism is that if act-utilitarianism does not provide a unique decision procedure, it "no longer defines a distinctive political position" (Kymlicka 1990, 47).

appreciate the value of honor, he's mistaken. Or, in Rawls's more technical terms, the interest in honor falls within the category of those things regarding which there can be no reasonable disagreement.

We should add that Øverland's example is in any case a bit hard to grasp. If Clint doesn't care about his honor, and if he assesses that he is doomed, why does he bother at all to fire at his attackers, killing no less than three of them? Øverland's rather odd answer is that Clint does so as a result of "succumbing to boredom and hunger." But what would Clint answer if asked "Hey Clint, why did you kill those guys, they are human beings, are they not?" He couldn't say, "They attacked me and I had no other way of defending myself from them" because, ex hypothesi, he did not believe that he could defend his life. Neither could he say that he killed them because he couldn't let them treat him as if he were a wild animal. Øverland's Clint is both irrational and morally flawed, is prepared to kill fellow human beings out of *boredom*.

Another problem with this objection has to do with the thought that because Clint does not believe in the idea of honor he has no interest in protecting it, hence he cannot be said to kill his hunters in order to defend his honor. This thought assumes that what people have an interest in is the same as what they *believe* they have an interest in, which is surely false. People have an interest in not being treated as mere means, whether or not they acknowledge it; they have an interest in not being killed (save extreme circumstances) whether or not they acknowledge it, and – to the present discussion – have an interest in maintaining their honor even if they don't recognize it.

The second objection against the honor solution argued that its application to cases like John Wayne violated the necessity condition for self-defense. Wayne could defend his honor (or significantly reduce the offense to it) without shooting at his enemies. He could instead face the bunch and fire his gun, which would be a simultaneous manifestation of generosity and contempt. According to this objection, what matters for saving one's honor is the *attitude* one has toward one's attackers, not whether one is able to get in a few futile shots.

However, when one thinks of paradigmatic examples of disrespectful treatment such as being thrown to the ground, trod upon or spat at, it's really hard to see how loss of honor can be avoided simply by taking a different *attitude* towards one's attackers. This might have been an option for one or two Stoics, but not for the rest of humanity. For ordinary people, what is required in order to deal with degrading messages of the kind just mentioned is a counter-message, one that presents Victim (both in her own eyes as well as in the eyes of others) as a proud agent who has the will, the power, and the courage to fight in defense of her vital interests. In cases like the rape-victim and Wayne, violent resistance therefore seems necessary for the protection of honor.

6 Conclusion

The problem of futile self-defense provides an opportunity to revisit the philosophical debate about the ultimate basis for the right to self-defense. According to the interest-based theory of self-defense, this right is based on Victim's right to defend her interests. When harming Aggressor cannot neutralize the primary threat posed to Victim, it is not her interest in her life or in her bodily integrity that is protected but her interest in her honor. By contrast, on the justice-based theory of self-defense, Victim's right to harm Aggressor is grounded in her right to realize a just distribution of risks and burdens. Victim is allowed to harm Aggressor because

that would make the world a more just place; Aggressor, who was tainted by his act of aggression, is entitled to a lesser claim to protection than Victim.

The main purpose of our paper has been to show to what such a justice-based view is committed and how it entails unacceptable results. If we're right, then this theory not only falls short of solving the problem of futile self-defense, but fails more generally as a theory of self-defense. This failure makes the alternative theory look more attractive, again both in the context of futile self-defense and more generally. To understand how Victim might use force in this context, we must point to some interest of hers that is threatened, and the most likely one is her honor.

References

Arneson R (2018) Self-defense and culpability: fault forfeits first. San Diego Law Rev 55:259–260
Benbaji Y (2005) Culpable bystanders, innocent threats and the ethics of self-defense. Can J Philos 35:585–522
Bowen J (2016) Necessity and liability: on an honour-based justification for defensive harming. J Pract Ethics 4: 79–93
Ferzan K (2018a) defending honor and beyond: reconsidering the relationship between seemingly futile defense and permissible harming. J Moral Philos 15:683–705
Ferzan K (2018b) Defense and desert: when reasons don't share. San Diego Law Rev 55:265–289
Frowe H (2014) Defensive killing: An essay on war and self-defense. Oxford University Press
Husak D (2018) The vindication of good over evil: futile self-defense. San Diego Law Rev 55:291–314
Kymlicka W (1990) Contemporary political philosophy. Oxford University Press, New York
Lazar S (2009) Responsibility, risk, and killing in self-defense. Ethics 119:699–728
McMahan J (1994a) Self-defense and the problem of the innocent attacker. Ethics 104:252–290
McMahan J (1994b) Innocence, self-defense and killing in war. J Polit Philos 2:193–221
McMahan J (2004) The ethics of killing in war. Ethics 114:693–733
McMahan J (2005) The basis of moral liability to defensive killing. Philos Issues 15:386–405
McMahan J (2009) Killing in war. Oxford University Press, New York
Øverland G (2011a) On disproportionate force and fighting in vain. Can J Philos 41:235–262
Øverland G (2011b) Moral taint: on the transfer of the implications of moral culpability. J Appl Philos 28:122–136
Rawls J (2001) Justice as fairness; a restatement. Harvard University Press, Cambridge
Robillard M (2017) Fighting for one's self. In: Jenkins R, Robillard M, Strawser BG (eds) who should die?: the ethics of killing in war, Oxford University press, pp 102-117
Statman D (2008) The success condition for legitimate self-defense. Ethics 118:659–686
Tadros V (2011) The end of harm: the moral foundations of criminal law. Oxford University Press, New York
Thomson J (1991) Self-defense. Philos Public Aff 20:283–310
Uniacke S (2014) Reasonable prospect of success. In: Frowe H, Gerald L (eds), How we fight: ethics in war, Oxford University press, Oxford, ch. 4