

Article

The Justice of Punitive Wars

Benjamin Robert Koons^{1,*}

¹ Center for Thomistic Studies and Department of Philosophy, University of St. Thomas, 3800 Montrose Blvd., Houston, TX 77006, USA

* Corresponding author: benjamin.koons@stthom.edu

Submitted: 28 October 2024, accepted: 20 April 2026, published: 10 May 2026

Abstract: Many recent defenders of just war theory have denied that punishment is a just cause for war. Against this consensus, I argue that punishment is a just cause for war. To defend this claim, I appeal to recent work in social ontology and social epistemology (especially the work of Christian List and Philip Pettit) that shows that groups and not just individuals can be responsible for their actions. For this paper, I defend the thesis that an international treaty organization may initiate a war against one of its member-states for the purposes of punishing that member-state. Many critics of punitive wars have taken the war itself to be the punishment, but I consider the war to be merely the means by which one brings an offending state to accept punitive terms. I show that punitive terms can achieve the three aims of punishment: retribution, deterrence, and reform. Finally, I respond to two objections: that no punitive war could be just because victims of aggression are never impartial in judging their aggressors and that widespread acceptance of the justice of punitive wars would return us to a more warlike and barbaric stage of human history.

Keywords: just war theory; war; punishment; punitive war; collective intentionality; ethics of war; retribution; deterrence

How to cite: Koons, B. R. The Justice of Punitive Wars. *Journal of Controversial Ideas* 2026, 6(1), 11; doi:[10.63466/jci06010011](https://doi.org/10.63466/jci06010011).

©2026 Copyright by the author. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.



1. Introduction

Contemporary just war theory has largely abandoned punishment as one of the just causes for war, but I intend to show that if one accepts the justice of defensive wars, then punitive wars are plausibly justified.¹ I defend this thesis:

¹ For objections by just war theorists to the justice of punitive wars, see Anthony Lang (2005), David Luban (2011), and Cécile Fabre (2014). Classical just war theorists like Augustine ([419] 1865, *Quaestiones in*

Punishment as Just Cause (PJC): It is a just cause for international treaty organization X to initiate a war with member-state Y so as to punish Y for an injustice against state Z.

I have limited my thesis so that only international treaty organizations (e.g., the United Nations) can punish injustices by member-states. This thesis says nothing about the form of punishment and when it should occur. I do not discuss other conditions for justly initiating a war in this paper, but I presume there are others such as necessity, proportionality, and reasonable prospects of success. So, one should not interpret *PJC* as a thesis about what is sufficient for a state to justly initiate a war. Generally, “just cause” is understood as just one necessary but insufficient condition for justly initiating a war. There are no reasons the other conditions for a just war could not be met for punitive wars so long as the offense is sufficiently grave (e.g., genocide or economically debilitating cyberattacks) and all reasonable diplomatic means have been used to try to bring the offending country to accept suitable punishment for its offense. I also assume that even if one thought that *PJC* was true, one might still think it was imprudent to change international law such that one could legally prosecute a war for a punitive cause. In particular, one might think that while punitive wars could be just in principle, there’s still good reason to think that having just punitive wars would increase the number of wars, and that it’s better to have fewer wars even if these wars would be just wars. However, I respond to this worry in Section 5, so in fact I think that given *PJC* there would be good reason to change international law to allow punishment as a just cause of war.

When objecting to the possibility of any justified punitive wars, David Luban wrongly equates this possibility with the following claim about just cause (Luban, 2011, p. 299):

War as Punishment: It is a just cause for international treaty organization X to initiate a war with member-state Y so as to punish Y *with the destruction of war* for an injustice against state Z.

Another possibility is that the punishment is supposed to happen after the war’s conclusion by imposing punitive terms in a peace treaty:

Punitive Terms: It is a just cause for international treaty organization X to initiate a war with member-state Y so as to punish Y *with punitive terms* after the war’s termination for an injustice against state Z.²

An example of this sort of punishment is the Treaty of Versailles in which the victorious Allied Powers punished Germany through reparations payments and territorial

Heptateuchum VI.10), Thomas Aquinas ([1274] 1888, *Summa Theologiae* IIaIIae 40.1), and Hugo Grotius ([1625] 2005, *The Rights of War and Peace*, Book ii, chapter 1, section ii) claim that the punishment of an injustice is a just cause of war. Some notable contemporary just war theorists that still accept punitive wars as justifiable are Jean Bethke Elshtain (2003), Michael Walzer (2015), and Jeff McMahan (2008). Only McMahan gives a sustained defense of this position, and his defense substantially differs from mine. First, he doesn’t distinguish *War as Punishment* from *PJC* but refers to them interchangeably (McMahan, 2008, pp. 83–84). Second, he criticizes any social ontology that accepts groups as agents (83). Third and consequently, he denies that punitive wars can be retributive.

² Walzer makes a similar distinction between *War as Punishment* and *Punitive Terms* (Walzer, 2015, pp. 296–97). My thesis differs from Fabre’s (2016) argument, since she defends the imposition of punitive terms after a war without defending this as just cause for initiating a war in the first place.

concessions. In this essay defending *PJC*, I will refer solely to *Punitive Terms* and not to *War as Punishment*.

As a preliminary, I want to justify limiting my thesis to international treaty organizations. By making international treaty organizations the ones to mete out punishment, I avoid the objections Luban (2011) and Rodin (2002) raise: no victim can be impartial in punishing an offender, and so no state can be impartial in fighting a war against an aggressor state so as to punish it. The objections are similar in that both object to the possibility of a victim's being impartial in punishing an offender, but while Luban still apparently considers this to be punishment, albeit an unjust one, Rodin thinks this case would lack something constitutive of punishment as such (Rodin, 2002, pp. 174–79). Rodin plausibly takes impartiality to be necessary for one to be an authority imposing a punishment, but he wrongly takes it to be impossible for a victim to be impartial.³ In fact, I think that it is ultimately possible for a state that's the victim of injustice to be impartial in judging an offender state. Hence, I think the following version of *Punitive Terms* is also true:

Punitive Terms (State Version): It is a just cause for state X to initiate a war with another state Y in order to be able to punish Y *with punitive terms* after the war's termination for an injustice against state Z (where Z may or may not be identical to X).

Given that I think a punitive war waged by an individual state against an offending state is in principle possible, I give a more substantial response to Rodin's and Luban's objections in Section 4. However, since I think that it is extremely difficult in practice for any state to be sufficiently impartial in meting out punishment for an offense against itself, I focus in this paper on the original version of *Punitive Terms* which refers to international treaty organizations because such bodies are more suited to impartial judgment. Rodin thinks a sufficiently impartial body for such judgment would have to be a universal state (albeit an "ultra-minimal" one that only maintains a military and adjudicates international disputes), but presumably all that's really required is an organization that has (i) sufficient resources to wage war against offending states, (ii) sufficient interest to wage war for such offenses, and (iii) sufficient impartiality to punish justly.

Any international organization could satisfy iii for offenses between states so long as no single state in the organization dominated the organization so much as to be able to effectively decide any case. Yet while international courts might satisfy iii, they would fail to satisfy i and ii, which is why international treaty organizations which consist of member-states with armies would be better. In fact, even more temporary groupings of states would be sufficient for the sake of impartiality. Rodin (2002) requires some contractual submission of the offending party to the judge's judgment in order for the punishment to be just, so presumably an international treaty could include such submission and hence would be sufficient for this requirement. However, if Fabre (2016) is correct that the right to punish properly belongs to the victim, then this contractual submission to the judge's judgment couldn't be a universal requirement for just punishment.

³ For an opposed view to which I am sympathetic, see Fabre (2016). Fabre contends that in fact the right to punish properly belongs to the victim but that victims quite reasonably transfer this right to states both because they have more resources for juridical proceedings and also because the state is better capable of giving a dispassionate condemnation of the wrong action (Fabre, 2016, pp. 178–79).

One might worry that no international treaty organization would have sufficient interest to wage war against its own member-states. However, an international treaty organization would meet condition ii because it would not want its member-states committing injustices against each other, since this would weaken the whole organization. Hence, it would be willing to wage war for such offenses so long as they were sufficiently serious. Of course, there would still be the possibility of one international treaty organization's wronging a state not in that organization or another international treaty organization, but the ultra-minimal universal state would also be capable of acting unjustly against its member-states, which would leave no impartial arbiter. To my mind, the only solution is not to simplify but to add more and more overlapping levels whereby there will always be either some group of peers or some higher level of organization to appeal to.

2. Group Agency and Responsibility

Imagine the following case: the country Freedonia consists of three ethnic groups: Blue, Violet, and Green, which are each perfectly represented by an ephor, while a king presents the resolutions to be voted on for any judgment. Its founders framed the Freedonian constitution with the following rules for judicial proceedings:

1. A simple majority of the ephors is sufficient for any resolution to pass.
2. Each resolution consists of a single clause (to avoid riders).
3. The executive must enact the ephors' resolutions without inquiry.⁴

A case arises where a Sylvania settlement in Freedonia (Sylvania is a neighboring country) is accused of flying a Sylvania flag. This causes a great stir, and the king proposes the following three resolutions:

- A. The Sylvanians flew a Sylvania flag.
- B. Flying a foreign flag is a treasonous act.
- C. Treasonous acts are to be punished by execution.⁵

The following resolution, which represents A&B&C, won't be voted on though:

- D. The Sylvanians are to be executed.

The three ephors vote in the following way on these three issues, and I also show how they would vote on D (A&B&C) if they could:

⁴ This stipulation is unrealistic insofar as few governments would forbid a government official from further inquiries into the information backing some order, but it's realistic insofar as many governments don't give sufficient resources to officials to inquire into much of the information they have to act on or only make this information available after a lengthy bureaucratic process. There's also the classified information an official might not be given access to, which would be necessary for him to actually substantiate the information he's been given.

⁵ I assume that capital punishment is morally permissible, but if it's not, then one can modify the case by replacing execution with whatever is the most severe but also morally permissible form of punishment (e.g., lifelong imprisonment, impoverishing fines, and forced labor).

	A. Flew Flag?	B. Treason?	C. Execution?	D. Execute Sylvaniaans?
Blue	Yes	Yes	No	No
Violet	Yes	No	Yes	No
Green	No	Yes	Yes	No

Each of A, B, and C pass by a two-thirds majority, and so the executive enacts the ephors' resolutions by executing the Sylvaniaans. None of the ephors or their respective ethnic groups is individually responsible for this act, since none of them supported D. The ephors' votes only cause the execution given the constitution, but the constitution itself isn't intrinsically wicked (even if it's simplistic). The executive is not guilty because it is only given commands it must automatically enact—hence why it is unimportant whether the executive is a conscious agent or merely some kind of robot.⁶ Moreover, since by rule 3 the executive cannot inquire into the facts of the case, it has no reasonable grounds in any particular instance for refusing to obey a command given by the majority of the ephors. Hence, one can't just object that the executive does wrong by failing to disobey an immoral order. So, when the Sylvaniaans are wrongly executed, Freedonia has done an injustice, but no individual Freedonian is clearly responsible. If the Sylvaniaans seek to punish the wrong, whom will they punish? Neither each ephor and his ethnic group nor the king nor the executive nor the constitution's framers are individually responsible, and so it seems that only one entity is liable to punishment: the whole country of Freedonia.

In fact, one can defend the conclusion that the whole country of Freedonia is liable to punishment even if one thinks that certain individuals in the case are also liable to punishment. All that is really necessary to show is that there is a responsibility "deficit" for the wrongdoing once all the responsibility of individuals is accounted for. So, even if one thinks that each of the Freedonian ephors has done something wrong in engaging in the vote at all when this substantially risks the unjust executions of innocent Sylvaniaans, it is still the case that the Sylvaniaans suffer the wrong of being unjustly executed and

⁶ One might object that it was wrong in the first place to have the executive act in such an unthinking way. However, there may be good reasons for this based on a history of tyrannical executives. If the resolutions of the ephors are almost always just, then it seems quite reasonable for the executive merely to accept their group testimony in favor of the resolution. For an account of group testimony as something over and above the testimony of the individuals in the group, see [Tollefsen \(2007\)](#). For two accounts of justified group belief on which the beliefs of the ephors would be justified but false in this particular case, see [Dunn \(2019\)](#) and [Koons \(2026\)](#). Moreover, it is possible to modify the case and dispense with the executive altogether such that not only is the decision to punish collective but the actual action of punishing is also collective. For instance, imagine that the punishment the Freedonians decide on for treason is that traitors be denied all access to the market such that no Freedonian would give any goods to a condemned traitor. In this case, any given instance of a Freedonian refusing to sell goods to a Sylvaniaan living in Freedonia on its own would not be especially burdensome to the Sylvaniaans, but the cumulative effect of this denial of market access could be devastating even to the point of effectively starving these Sylvaniaan immigrants. Who starves the Sylvaniaans then? No individual Freedonian seems to by refusing to give them food, but then the Sylvaniaans' starving is no accident. Hence, it must be the Freedonians as a group that enact this punishment. This case is different from the Harmless Torturers case Derek Parfit presents ([Parfit, 1984](#), p. 80). Parfit's case involves a thousand individuals who each inflict an imperceptibly small amount of pain on a thousand victims, which together causes each victim to suffer agony. An individual Freedonian's refusal to sell goods to a Sylvaniaan need not cause any harm—even imperceptible harm—to a Sylvaniaan. Indeed, even if all but one Freedonian refuses to sell goods to Sylvaniaans, but one sufficiently well-supplied Freedonian does sell goods to Sylvaniaans, this could be sufficient to prevent any harm's coming to the Sylvaniaans. Hence, it is essential to the harm caused by the Freedonians that they inflict it collectively.

not merely the wrong of being subjected to a process that substantially risks their being unjustly executed. Who is the agent responsible for this graver wrong?⁷ It is not the ephors or any other individual, and so it must be Freedonia collectively.

One objection to punishing the whole country of Freedonia is that if no individual Freedonian has done anything wrong, then it would be unjust to punish all Freedonians, since then one would be punishing an individual that was not guilty of any wrongdoing. In response, I argue that this objection wrongly supposes that in punishing some whole, one necessarily punishes each of its parts. One can punish a guilty person without punishing any of his body-parts even if it's the case that in punishing a guilty person one does harm to parts of his body. Similarly, one can punish a guilty collective without punishing any of its individual members even if it's the case that in punishing a guilty collective one does harm to its members.⁸ In both cases, however, the part only suffers harm insofar as it is part of some whole which is guilty. So, for instance, one might punish a baseball team for its practice of sign stealing by depriving it of its right to compete for the rest of the season, stripping it of its championships or depriving it of its first-round picks for new players. In such a case, individual players are only deprived of rights they have in virtue of being part of the team.

A second related objection is that if no individual Freedonian has done wrong and hence no individual is liable to punishment, then Sylvania cannot legitimately prosecute a punitive war against Freedonia, since no individual Freedonian would be liable to the harm he would suffer in war. This objection comes at a high cost however. As we will see in Section 4, there can be cases in which a country must prosecute a punitive war or else it will be the victim of further injustices that culminate in its destruction. It's implausible that a country could be put in a position where through no fault of its own but merely in virtue of its being the victim of injustice it cannot legitimately act to prevent its own destruction, at least when in so acting it has a reasonable prospect of preventing this destruction and the harms foreseen in the war are proportional to the harms it is acting to prevent.

Finally, one might object that even if it is true that Freedonia as a country is liable to punishment, the case I've presented is dissimilar to any real-world case. In real-world examples, there will almost always be individuals who bear some individual responsibility for the harm. For instance, imagine an altered case where the Blues and Violets vote the same way but the Greens vote for A, B, and C. One could say that now the Greens bear a special individual responsibility for the unjust execution. In response, I argue that even if in more realistic examples there always happen to be guilty individuals, this doesn't entail that now Freedonia as a group is absolved of its collective responsibility for the injustice. [Lawford-Smith \(2018\)](#) defends this point at more length, but in this section I will

⁷ I assume that one cannot suffer wrongdoing without there being an agent responsible for this wrongdoing.

⁸ However, in punishing the collective and harming individuals, it is necessary neither that one cause overall harms to these individuals nor that one intend any harm to them. One need not cause overall harms because it's possible in punishing a guilty collective to compensate innocent individuals for any harm they might sustain through this punishment. So, for instance, if the public treasury is fined, the punishing treaty organization could use these fines to give out payments to individuals in the punished state who are innocent of participation in the collective's guilty actions. If the doctrine of double effect is true, then one need not intend to harm individuals at all even if in punishing a guilty collective one foresees that in punishing the guilty collective individuals will sustain some harm.

go through some reasons for punishing a guilty collective rather than merely punishing guilty individuals.

Group agency and responsibility are regaining currency in political and social theory, especially thanks to the work of Philip Pettit and Christian List. It's a pervasive belief that groups may act and be responsible for their actions, and our linguistic and social practice presumes the existence of group entities. For the purposes of this paper, I will assume that states and other group entities are the sorts of things that can act and can be held responsible for their actions, but I remain neutral on whether group entities are ultimately reducible to their individual members. One worry about whether group entities can be held responsible for their actions is whether they are autonomous agents (Pettit, 2007, p. 175). According to List and Pettit's proof using the discursive paradox, some group agents are autonomous agents, since the judgments (and desires) of these group agents on any given proposition don't supervene on their individual members' judgments on that same proposition. The judgments of the three Freedonian ephors on the three propositions and their judgments on the conclusion that would follow from these propositions is an example of a discursive paradox, since what collective attitude Freedonia has towards the proposition that the Sylvania's ought to be executed will either (i) fail to be responsive to the attitudes of the individual ephors on the three first propositions or (ii) fail to be responsive to the attitudes of the individual ephors on the fourth and concluding proposition or (iii) be irrational since Freedonia inconsistently believes the Sylvania's ought not to be executed but believes propositions that entail that they should. More generally, the discursive paradox can arise in any case in which the attitudes of a group are responsive to the attitudes of the individuals in a group and where the group forms attitudes towards different propositions that are rationally connected (e.g., where two propositions entail some third proposition). In their proof, List and Pettit show that if one assumes that groups may have members with any number of sets of consistent beliefs, that the group's beliefs and other attitudes are themselves consistent and complete, and that no individual's attitudes have a disproportionate impact on the group attitude, then it follows that there is no function aggregating the set of attitudes of individuals to a given proposition to derive the group attitude towards the proposition (List & Pettit, 2002, pp. 89–110).⁹ Let's explain how this works in the case of Freedonia. Freedonia can only achieve group rationality by limiting voting to the propositions A, B, and C because if the ephors could also vote on D, then the group would have an inconsistent belief: that the Sylvania's should and should not be executed. The way that the Freedonian constitution prevents this group irrationality is by privileging the individual attitudes on certain propositions (those chosen by the king) over other propositions. There is no function from the individual attitudes on proposition D through which one can derive the group attitude towards proposition D.¹⁰ The complicated way in which group attitudes

⁹ See also List and Pettit (2011) for a non-technical discussion of this theorem.

¹⁰ Since List and Pettit require that the judgment aggregation function gives a complete set of beliefs for the group, i.e., for every proposition the group either assents to that proposition or its negation, a function like unanimitarianism that entailed that the group believes all and only what all the individuals in a group believes would fail to be a judgment aggregation function. They defend completeness more generally, but they point out that were a group unable to believe what it lacked unanimity about, this would prove crippling for most groups (List & Pettit, 2002, pp. 105–6). Others have also given examples where all the individuals in a group have the same attitude towards a proposition, while the group does not have that same attitude. Alex Bird gives the example of the Emperor's New Clothes as a case in which all the

supervene on individual ones grants the group agent some autonomy from its individual members.

There are three reasons for holding groups responsible rather than just apportioning blame to individuals. First, there is the problem we saw in the Freedonia case where it may be that there is no individual (at least, *qua* individual) responsible for the injustice and that only the group may be held responsible.

Second, if one understands group actions as consisting in the attitudes and actions of individuals bearing *qua*-relations to various social roles (e.g., Texas' execution of a murderer consists in the performance of a lethal injection by George Smith *qua* state executioner), then it makes sense to punish individuals *qua* member of the group. So, instead of punishing individual citizens of an offending state *qua* individuals, one might punish these same individuals *qua* citizens by fining their public treasury, seizing public goods, and dismantling public infrastructure.¹¹ These are all benefits the individuals only enjoy insofar as they belong to a certain group, so they are liable to losing those benefits when the group is punished. This is my preferred response to the objection that one cannot justly harm innocent individuals when punishing a guilty group. However, one can also reply that it's not always necessary to leave such individuals any worse off when punishing a guilty group. For instance, in the aftermath of World War I, the Allied powers could have compelled the dissolution of Germany as a political unit but allowed individual Germans to form new states and divided all of the public assets of Germany among these new successor states (e.g., Bavaria and Hanover could again become their own sovereign countries). This would be a punishment of Germany as a group, but it wouldn't necessarily lead to any individual German's being worse off overall.

Third, there is the practical difficulty of apportioning blame to individuals in a complicated act of group malfeasance. For example, Chicago's police department plainly covered up damning details in a police brutality case, but it's nearly impossible to determine who knew what when. Moreover, if one has a framework that leads to intractable epistemic problems which an alternative framework doesn't lead to, this is a reason to prefer the alternative framework because it suggests that the original framework is blind to a feature of reality which makes it very difficult for it to make judgments that depend on this feature. In this case, not allowing the possibility of collective responsibility leads to intractable epistemic problems in assigning blame for wrongdoing. On the other hand, if one accepts the possibility of group responsibility, one can explain why it is more straightforward to assign blame to groups in such cases than individuals because this is the correct level for assigning blame.

3. Groups and Justifications of Punishment

There is no problem applying any of the traditional justifications of civil punishment (retribution, deterrence, and reform/rehabilitation) to the international setting. The analogy is clear for *Punitive Terms*, since such terms take the same form as any judge's sentence.

individuals in the kingdom know that the emperor has a certain birth-mark on his shoulder but in which this knowledge has no social effects and hence which is not known collectively (Bird, 2010, pp. 29–30).

¹¹ Group punishments may incidentally harm individuals *qua* individuals too, but many legitimate punishments of individuals also have significant adverse effects on people who aren't being punished (e.g., taxpayers must pay to maintain prisons, and the prisoners' families lose their income).

Retribution is the easiest to apply, since it consists in giving an offender his “just deserts,” which a peace treaty accomplishes by penalizing the offending country. Deterrence is similarly easy, since one state can punish another to deter other states from similar offenses. Wrongdoing by one state against another is generally a more public act than wrongdoing by one individual against another, and so a state’s doing wrong with impunity is more likely to embolden other states’ doing likewise.¹²

Reform may also be accomplished in a punitive peace treaty. An individual is reformed in punishment insofar as he forms dispositions not to engage in this kind of wrongdoing again. Peace treaties can accomplish this reform by deposing the leaders and imposing constitutional changes on an offending state, which make the state disposed not to engage in such wrongdoing again. For instance, the constitutional changes in Japan after World War II were directed towards this kind of reform.

One worry about the possibility of collective punishment is whether such punishment can ever achieve one of the (putative) constitutive aims of punishment: the offender’s being penitent for his wrongdoing (Duff, 2003, pp. 106–15). There are two different kinds of reasons one might be skeptical of a group’s being capable of penance. First, penance involves believing that one has done wrong and that one’s punishment is fitting, but groups are not capable of having beliefs.¹³ Second, penance involves feeling pain and sorrow at one’s wrongdoing, but groups are not capable of having feelings.¹⁴

In fact, I think groups are capable of having beliefs and having feelings or at least capable of states analogous to beliefs and feelings, but one need not make these assumptions to respond to this objection. Instead, I would deny that penance is a constitutive aim of punishment at all. Otherwise, there would be no reason to punish a hardened criminal for whom there was no prospect of his being penitent about his crime.¹⁵ Worse, it would give one more reason to punish those who were more likely to be penitent than those who were less likely to be penitent. Even if one thinks that communication of some kind is integral to justifying punishment, one need not think either that the criminal need be persuaded by this communication or that it is the criminal alone to whom one is communicating through the act of punishment. On the denunciatory theory of punishment, the state communicates to the community in general its condemnation of the offender’s wrongdoing (Wringe, 2016).¹⁶ Such denunciation is clearly a part of international actions as well, since punitive terms imposed by an international treaty organization also communicate to the community of nations a denunciation of the offending country’s actions.

¹² Cyberattacks and state-sponsored terrorism complicate this generality.

¹³ In defense of groups’ having beliefs, see Gilbert (1989, pp. 288–314) and more recently Gilbert (2002a, 2004). If one denies that groups have beliefs but allows them to have “acceptances,” then one might allow groups to be penitent insofar as they accepted that their actions were wrong and deserved punishment. For this distinction applied to groups, see Mathiesen (2006).

¹⁴ In defense of groups’ feeling guilt, albeit in virtue of individuals’ having group-directed guilt, see Tollefsen (2006). Lawford-Smith (2018) proposes instead that one “reinterpret” collective guilt as merely the belief that the group is guilty, following Gilbert’s (2002b) account of collective guilt in terms of a joint commitment.

¹⁵ Duff refers to these as “defiant offenders” (Duff, 2003, pp. 121–25). His response to the criticism I raise here is that we owe it to the victim to attempt to cause the offender to repent even if we are “sure” to fail (2003, p. 123). I do not see how causing the offender to be penitent can be a constitutive aim of punishment if one can punish someone while knowing that one cannot achieve this aim.

¹⁶ Lawford-Smith (2018) also defends collective punishment on the basis of Wringe’s denunciatory theory.

While I have argued that one can wage a war to impose terms to punish a state so as to deter further wrongdoing, Jeff McMahan argues that one can accept deterrence as a just cause for war without taking this deterrence to be punitive.¹⁷ However, he imposes a constraint on this war aim: a state can't wage a war against another state for the purposes of deterrence unless some wrongdoing by the state has made some of its members liable to be harmed. In particular, McMahan claims that one state makes itself liable to be harmed by committing some wrong which lowers the level of deterrence another state might have. So, an attack by one state on another is an act of wrongdoing that makes the attacked state (as well as other states at risk of attack) more vulnerable to future attacks. For instance, in the attack on Pearl Harbor, Japan's unjust aggression made the United States more vulnerable to future attacks unless the United States waged a war against Japan to restore its previous level of deterrence.

To respond to this alternative, non-punitive way of accepting deterrence as a just cause for war, I will indicate three differences between our views and the implications for these differences.

First, on McMahan's view, for one's act of deterrence through war to be just, it is necessary that the one being warred upon have committed some wrong, but it is not essential to its being the act of deterrence that it is that it be a response to some perceived or actual wrongdoing. On my view, since the deterrence one seeks in war is punitive, it is part of what this kind of deterrence is that it be directed to respond to some perceived or actual wrongdoing. Given this difference, McMahan cannot explain on the basis of the nature of the deterrence which is sought why one can only war upon those who have made themselves liable to be harmed. Hence, this restriction is merely a side constraint on one's acting for the sake of deterrence. On my view, in acting for the sake of deterrence one can only war upon those who have done wrong because of the very nature of the kind of deterrence one is acting for. Punitive deterrence is essentially directed against one who has done wrong or seemed to have done wrong.

This has further implications for our two views. One implication is that on my view one's intention in punitive terms can be the general intention to deter further wrongdoing. Indeed, one could even have the intention of deterring all further wrongdoing (even if this goal ends up being infeasible). By contrast, on McMahan's view, the intention has to be to restore the level of deterrence one had before the wrongdoing. However, it's not clear why this should be a constraint on what a state seeks in deterrence. If a state could attain a higher level of deterrence than it had previously, this seems quite legitimate. In fact, I think that McMahan needs to include this constraint because he can't rely on the extent of the wrongdoing as the measure for what would be a proportionate act of deterrence.¹⁸ Punitive terms need to be proportionate to the wrong done, but it's perfectly fine if in imposing these terms the punishing state ends up with a higher level of

¹⁷ The comments I present here are based on editorial comments by McMahan on this paper, but one can compare them to his published remarks (McMahan, 2009, pp. 156–57, 189–92). Whenever I speak here of “McMahan's view,” I just mean the view suggested by his editorial comments.

¹⁸ In correspondence, McMahan argues that this constraint is based on the extent to which a wrongdoer is liable to harm. In particular, “a wrongdoer cannot be *liable* to more harm than is necessary to restore deterrence to the level from which it has been reduced by the wrongdoer's wrongful action.” By contrast, I take it that it is the wrongness of the act itself rather than merely its actual effect on the level of deterrence which is relevant to determining what harm is proportionate for the wrongdoer to suffer.

deterrence. Another implication for McMahan's view is that a state in seeking to restore its level of deterrence need not be acting to respond to the wrongdoing insofar as it's wrong. Presumably, on his view, an amoral state that suffered some wrong at the hands of a neighboring state it had hoped to wage war with anyways but wanted a pretext could acknowledge the wrongness of the wrong, be indifferent to its wrongness, and wage war with the neighboring state. Even if this amoral state only waged a war devastating enough so as to restore its previous level of deterrence, it would still not be waging war for the sake of a just cause. One could compare the problem with McMahan's view here with the deviant causal chain problem for Donald Davidson's theory of action as what is caused by beliefs and desires. A state that wages war for the sake of restoring a level of deterrence that a wrongdoing lowered need not be acting in response to the wrongdoing insofar as it is a wrongdoing just as an action caused by a belief-desire pair need not be caused by these in virtue of the content of the belief or desire. However, if the state is acting so as to deter such wrongdoing insofar as it is wrong, then this is just punitive deterrence.

A second difference is that on McMahan's view it is only the war itself that acts as a non-punitive deterrent and not the terms imposed after the war because such terms would clearly be punitive. This would imply that in their conduct of the war itself, military commanders not only need to achieve strategic goals to bring about the capitulation of the enemy's forces but that they also need to make sure that the damage to the enemy caused by the war is sufficiently serious to act as a deterrent¹⁹. This is compatible with noncombatant immunity and can still condemn causing damage that is unnecessary for restoring the level of deterrence, but it complicates the role of military commanders in war. On my view, all that the military commanders need to achieve is the capitulation of the enemy to accepting punitive terms, and if they can achieve this by an entirely bloodless war that causes no damage to the enemy, this would be a mark of their success and not their failure as military commanders.

A third difference is that on McMahan's view he has to deny that retribution or reform would be sufficient for a just cause for war on their own without deterrence. Otherwise, one could accept that the deterrence one seeks in war is non-punitive, while thinking that the other ends of punishment are sufficient for just cause. In his comments, McMahan does think it unlikely that any war fought only for the sake of retribution or reform with no intended or expected deterrent effect could ever be proportionate given the vast expected harm modern war brings to noncombatants (e.g., to the many children and other civilians killed by Israel's strikes in Gaza). This might be generally true at present (although there have been recent wars like in the Falklands with almost no civilian casualties), but historically and possibly in the future there have been and could be highly targeted wars with minimal harm to noncombatants.²⁰ Some classical defenders of just war theory

¹⁹ Even if in achieving the capitulation of the enemy, a military commander also achieves a certain deterrent effect, this doesn't mean that the military commanders intend this deterrent effect. Nor do these two effects necessarily coincide: a nation with a much weaker military could initiate a series of brief wars of harassment against a more powerful nation with the intended effect of making it more susceptible to attack by other powerful rivals. Even if the military commanders of the more powerful nation successfully achieved the capitulation of the weaker, aggressor nation in each war, in itself this might not deter future attacks. It would be for the punitive terms imposed on the weaker, aggressor nation to actually punish this nation's aggression so as to deter it and the more powerful rivals from further unjust aggression.

²⁰ I mention this point about potentially bloodless future wars in the next section when I address a similar objection about proportionality.

have inclined towards a view like McMahan's when it comes to criminal punishment more generally, but this is typically for special theological reasons. For instance, following Augustine's earlier view, Aquinas thinks that the state ought to punish only those vices that tend to undermine social order by harming others rather than all vices (*Summa Theologiae* I-II 96.2). However, this is not because he thinks retribution is a relatively insignificant good but only because he believes on the basis of his Christian faith that God will ultimately ensure that all just retribution is doled out eventually (if not in this world, then at the Last Judgment). However, if one doubts the Last Judgment and thinks that retribution is an intrinsic good, it seems to me quite plausible that it would be among the greatest intrinsic goods. If retribution is something worth seeking for its own sake, then it is worth risking even a person's life for and possibly many people's lives to attain. I suspect that belief in the Last Judgment has shaped many historical thinkers' views to underestimate the number of lives worth risking for the sake of retribution, since such thinkers know that whether or not retribution is achieved now it will inevitably be achieved later, and so one need not risk much for the sake of achieving it now unless it achieves other ends.

On the basis of these three differences, it's quite clear that there are significant differences between taking punitive or non-punitive deterrence as a just cause for war. I do not think the implications of these three differences settle the matter once and for all, but they at least give some reasons for preferring Punitive Terms over an alternative theory based on non-punitive deterrence like the one suggested by McMahan.

4. The Problem of Biased Judgment

David Luban considers the problem of biased judgment a devastating objection to the possibility of justified punitive wars (Luban, 2011, pp. 317–18). The basic idea of the objection is that it is morally impossible for the injured party to be an impartial judge in its own case. There are three separate claims in this objection:

Vengeful are the Injured: An injured party will necessarily be vengeful in judging its injurer.

Proportional Punishment: In order for a punishment to be just, it must be proportionate.

Vengefulness and the Impossibility of Proportionality: A vengeful judge cannot give a proportionate punishment.

These three claims in the context of punitive wars entail that a state cannot give a just punishment to the state that has offended against it. Actually, even more radically, these three claims would entail that no state could give a just punishment to anyone that acted against it. So, a state could not punish traitors or rebels, since these act to injure the state, and it could not punish pirates who attack its citizens either, which is an absurd result. Moreover, even if these three claims were true, they would not entail the falsity of *PJC*. This objection does not rule out that a third party could act as the judge in such cases whether this third party is an international organization like the UN or a third state uninvolved in the original offense.

There is no reason to accept all three of these claims, though. *Proportional Punishment* should be accepted in some form, but the other two claims are too strong.

Weaker forms of these claims, which only refer to what's generally the case, might be acceptable, but weaker claims would not rule out *PJC*, although they would give any country reason for pause and self-reflection lest they be carried away by vengefulness into imposing disproportionately punitive terms. *Vengeful are the Injured* is only plausible when retribution is taken as punishment's justification, since retribution and revenge might seem conceptually related. When deterrence is taken as punishment's justification, any connection between being injured and seeking vengeance would appear merely accidental. Just as a teacher might be reluctant to punish a student in especially difficult circumstances for plagiarism but see that if she does not her other students will follow his example, so a state might loathe to punish a neighbor for its injuries but worry that its other neighbors will rise up against it if it does not. Punishment is not always something greeted with relish, and as often as not it is seen as an entirely disagreeable business. Moreover, the analogy between an injured individual and an injured group fails when it comes to vengefulness. It's likely that those citizens injured by the offending state are little cared for by the rulers of their own country. For example, perhaps a much-maligned minority in one country is attacked by another country; the leaders whose minority have been attacked may loathe the minority but still feel that for the sake of retribution or deterrence they must punish this injury.

Another problem for *Vengefulness and the Impossibility of Proportionality* is that it relies on a more general conception of vengefulness, which is articulated by the following claim:

Vengefulness and the Impossibility of Restraint: An injured avenger cannot restrain his anger.

If this claim is true, then there is a problem for accepting self-defense as a just cause for war as well, since it seems that in a case of self-defense it will be impossible for anyone to restrain himself from going beyond what is strictly necessary. Luban cites the American desire for vengeance on the Japanese after World War II as a case of a punitive campaign's shading into vengefulness, but it is rather a case of a defensive war's turning vengeful (Luban, 2011 p. 300). One might respond that self-defense is different, since even though there is a risk of its turning vengeful the injured state must defend itself or else face destruction. Yet this is a question-begging response, since many punitive wars justified on the grounds of deterrence (i.e., state X must punish state Y for its crimes or else state Y will do it again with impunity or state Z will also do it) are also necessary for the sake of a country's survival. A country that never punished injuries done to it would be at the mercy of all-comers.

Then again, one might object that these punitive wars of deterrence are more properly described as wars of self-defense, since it is ultimately the country's survival that is at stake in either case. The following case shows that the two justifications can come apart. Remember that Freedonia has three ethnic groups: Green, Blue, and Violet. Let us say that all of the Greens go for a holiday in Sylvania. It turns out that the Sylvanians are Blue and Violet supremacists who are racist against Green people, and so they murder all of the Green Freedonians. Thus, Freedonia no longer has any more Green people, and indeed they never again will have any more Green people. Freedonia cannot go to war with Sylvania for the sake of self-defense, since there is no threat that the Sylvanians will invade Freedonia (after all, Sylvania no longer has any reason to kill more Freedonians). Imagine, though, that there is another country Atlantis where all of the Blue Freedonians

are on holiday, and it turns out that the Atlantans are notoriously racist against Blue people. Freedonia knows that if they do not punish Sylvania for their genocide of the Green Freedonians, then the Atlantans will be emboldened to murder the Blue Freedonians. Hence, it is a matter of survival for Freedonia that they punish Sylvania even though Sylvania no longer poses a threat to them. Thus, deterrence and self-defense come apart in this example, and it is possible that a war be necessary for a state's survival even if it is an entirely punitive war.

Hence, this case also shows that it's entirely possible for a punitive war to meet the proportionality requirement for initiating a just war. However, perhaps one's concern is that punitive wars cannot be justified because the good they might attain is not proportionate to the risk modern war incurs because there is always the possibility of nuclear escalation. If that's the worry, however, then it will be difficult to justify wars of self-defense as well. How can the territorial integrity of one's own country compare to the threat of an all-consuming nuclear winter? Anyway, methods of warfare are always changing in unpredictable ways, so it's conceivable that future wars will be relatively bloodless affairs largely consisting in unmanned drones firing upon each other or low-casualty proxy wars in space.

Finally, *Vengefulness and the Impossibility of Proportionality* ignores the possibility of a virtuous state of character in regards to anger at insults and injuries. To apply an Aristotelian analysis of character virtues, there is a mean state of character between vengefulness and being a "doormat." This mean state is a habit that moderates one's emotional responses, and I will call it "indignation." When injured, the indignant person becomes angry but not so angry that he loses control of his reason. Rather he becomes irate and then reflects on whether an injustice has been done. On the other hand, a vengeful person becomes so angry about his injury that he cannot consider whether it was done unjustly. His emotions impair his ability to reason about the injury. A "doormat" suffers injury without becoming at all angry. Indeed, he may even see his friends injured without any kind of anger. Neither the vengeful person nor the doormat is virtuous because one feels more anger than is reasonable and the other less. The indignant person has developed a habit of feeling just the right amount of anger for the situation without allowing this anger to impair his reasoning. Now when the indignant person is injured, he does feel some sentiments of vengeful anger, but these do not impair his ability to consider what a proportionate response is. The indignant person does desire vengeance for the injury, but a virtuous habit governs this desire and moderates and channels it into a just response. If vengefulness is something a state can be accused of, then a state can also achieve an indignant habit. Such a state would be stirred by injuries and insults to a group sense of anger, and this anger would only subside if it decided that the injury was justified. This indignation would ensure that the state is not deterred by complacency or fear from punishing the affront. The indignant habit is completely compatible with proportionality, and it would motivate the state to take action in many cases when inhibiting emotions would prevent it.

5. The Barbarism Objection

Before concluding, I would like to respond to a more general concern about *Punishment as Just Cause*. One might worry that widespread acceptance of *PJC* would send us back to a more warlike stage of human history, since there would be another type of justification for wars besides self-defense. Surely, the last thing we need is more war,

and any philosopher who goes about defending more wars as just is really doing all of humanity a disservice.

In response, I contend that widespread acceptance of a new type of just war does not entail that there will be more wars. First, many punitive wars would have already been justified on the grounds of self-defense. These wars would either be genuine wars of self-defense that also intend to bring the aggressor to punishment or they would be purely punitive wars that were justified with the label 'self-defense.'

Second, one of the justifications for the punishment involved in punitive wars is deterrence, and wars of deterrence are intended to prevent future injustices and acts of aggression. Moreover, the punitive terms may also have a reformatory effect that will prevent the defeated aggressor from initiating future wars. The Treaty of Versailles' punitive terms for Germany tried (but sadly failed) to demilitarize the country and prevent future aggression against its neighbors. Indeed, arguably, the reason the treaty failed was that it wasn't enforced, and the Allied powers did nothing as Hitler violated its restrictions on German remilitarization.

Third, wars initiated for the sake of punishing wrongs won't trigger mutual defense treaties, which means that more wars will be localized. For example, if in 2015 Russia were to have initiated a war to punish Turkey for shooting down one of its fighter planes, then Turkey's NATO allies could have extricated themselves from involvement in such a war by appealing to the punitive nature of Russia's actions. Thus, the result of the acceptance of *PJC* might be to lessen the frequency and severity of wars.

There is a related objection one might raise against *PJC*: if punishment was accepted as a legitimate cause for war, then governments would use punishment as a pretext to justify what were really unjust wars of aggression. In response, I argue that if one thinks that governments fight wars solely for their own advantage and without any regard for justice and that such governments would call unjust wars of conquest "punitive wars" to give themselves a pretext, then it's unclear to me why accepting *PJC* would affect the number of wars. The number of wars would be entirely determined by the strategic advantages of wars. Perhaps one thinks that such governments still need a pretext to launch such wars even if this is not their true reason for waging war. Yet in that case, "self-defense" seems to be just as suitable a pretext for wars of aggression. In his declaration of war against Ukraine, Vladimir Putin framed the whole war as an act of Russian self-defense. As he said, "I repeat, our actions are self-defence against the threats posed to us and from an even greater disaster than what is happening today." Hence, if one thinks that advantage and not justice is what motivates states to begin wars and that states appeal to the justice of their *casus belli* as mere pretexts, then there's no reason to think that widespread acceptance of *PJC* would affect the number of wars.

6. Conclusion

This article's main point has been to show that if one country wrongs another and refuses to accept an international authority's punishments, then that authority may bring the offending country to accept punitive terms through military force. My argument relied on a moderate claim about social ontology: that there are group agents even if these are reducible to individual agents and that these group agents may be held responsible for their actions. I also dissociated the notion of punitive wars from any specific account of

the justifying aims of punishment indicating that it's compatible with retribution, deterrence, and reform as justifications.

Funding: This research received no external funding.

Acknowledgments: I thank the two anonymous referees, the editor Jeff McMahan, the anonymous BPhil examiners, and Rob Koons for their many helpful comments on previous versions of this article.

Conflicts of Interest: The author declares no conflict of interest.

References

- Aquinas, T. (1888). *Summa theologiae*. Leonine Edition. (Original work published 1274). [link to the article](#).
- Augustine, A. (1865). *Quaestiones in heptateuchum*. Patrologia Latina 34. (Original work published 419).
- Bird, A. (2010). Social knowing. *Philosophical Perspectives*, 24, 23–56. [[CrossRef](#)]
- Duff, R. A. (2003). *Punishment, communication, and community*. Oxford University Press.
- Dunn, J. (2019). Reliable group belief. *Synthese*, 198(S23), 5653–77. [[CrossRef](#)]
- Elshtain, J. B. (2003). *Just war against terror*. Basic Books.
- Fabre, C. (2014). Cosmopolitanism and wars of self-defence. In C. Fabre, & S. Lazar (Eds.), *The morality of defensive war* (pp. 90–114). Oxford University Press.
- Fabre, C. (2016). *Cosmopolitan peace*. Oxford University Press.
- Gilbert, M. (1989). *On social facts*. Princeton University Press.
- Gilbert, M. (2002a). Belief and acceptance as features of groups. *Protosociology*, 16, 35–69. [[CrossRef](#)]
- Gilbert, M. (2002b). Collective guilt and collective guilt feelings. *Journal of Ethics*, 6(2), 115–43. [[CrossRef](#)]
- Gilbert, M. (2004). Collective epistemology. *Episteme*, 1(2), 95–107. [[CrossRef](#)]
- Grotius, H. (2005). *The rights of war and peace* (R. Tuck, Ed.; Vol. ii). Liberty Fund. (Original work published 1625). [link to the article](#).
- Koons, B. R. (2026). Warranted group belief. *Erkenntnis*, 91, 1909–29. [[CrossRef](#)]
- Lang, A. F. (2005). Punitive intervention. In M. Evans (Ed.), *Just war theory: A reappraisal* (pp. 50–70). Edinburgh University Press.
- Lawford-Smith, H. (2018). What's wrong with collective punishment? *Proceedings of the Aristotelian Society*, 118(3), 327–45. [[CrossRef](#)]
- List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18, 89–110. [[CrossRef](#)]
- List, C., & Pettit, P. (2011). *Group agency*. Oxford University Press.
- Luban, D. (2011). War as punishment. *Philosophy & Public Affairs*, 39(4), 299–330.
- Mathiesen, K. (2006). The epistemic features of group belief. *Episteme*, 2(3), 161–75. [[CrossRef](#)]
- McMahan, J. (2008). Aggression and punishment. In L. May (Ed.), *War: Essays in political philosophy* (pp. 67–84). Cambridge University Press.
- McMahan, J. (2009). *Killing in war*. Oxford University Press.
- Parfit, D. (1984). *Reasons and persons*. Oxford University Press.
- Pettit, P. (2007). Responsibility incorporated. *Ethics*, 117(2), 171–201. [[CrossRef](#)]
- Rodin, D. (2002). *War and self-defense*. Oxford University Press.
- Tollefsen, D. (2006). The rationality of collective guilt. *Midwest Studies in Philosophy*, 30(1), 222–39. [[CrossRef](#)]
- Tollefsen, D. (2007). Group testimony. *Social Epistemology*, 21(3), 299–311. [[CrossRef](#)]
- Walzer, M. (2015). *Just and unjust wars* (5th ed.). Basic Books.
- Wringe, B. (2016). *An expressive theory of punishment*. Palgrave Macmillan.