

# The Intrinsic Good of Justice

BRIAN ROSEBURY

## Abstract

Some retributivists claim that when we punish wrongdoers we achieve a good: justice. The paper argues that the idea of justice, though rhetorically freighted with positive value, contains only a small core of universally-agreed meaning; and its development in a variety of competing conceptions simply recapitulates, without resolving, debates within the theory of punishment. If, to break this deadlock, we stipulate an expressly retributivist conception of justice, then we should concede that punishment which is just (in the stipulated sense) may be morally wrong.

## I

Advocates of a retributive theory of punishment sometimes appeal to *the intrinsic good of justice*. This good provides a reason why serious wrongdoers should be punished, irrespective of whether such aims as an increased sum of human happiness are well served by the decision to punish. As Michael S. Moore puts it,

Retributivism purports to be a theory of justice, and as such claims that punishing the guilty *achieves something good –namely justice-* and that therefore reference to any other good consequences is beside the point. [Emphasis added.] (1997: 11)

In this paper I will ask whether the idea of the intrinsic good of justice makes any distinctive contribution to the case for “pure” retributivism: the view that the offender’s guilt is not only a necessary, but a sufficient condition for punishment to be *pro tanto* justified, assuming that a legitimate authority to punish exists and that fair procedure is maintained. (When I refer later to “retributivism” without further qualification, I am referring to this view.)

An important reason for focusing on the idea of the good of justice as a justifying aim of punishment is that it is invoked every day in real-life situations, less often by philosophers than by a wider expressive community: judges, politicians, religious leaders, crime victims and the general public. A common formula is to advocate “justice [but] not revenge”. The distinction has great rhetorical power, since the attribution of justice tends to disable moral criticism: “just, yet morally wrong” strikes most people as puzzling if not contradictory. But what, in the context of punishment and as contrasted with revenge, is justice? Is it simply revenge, made orderly and proportionate by a system of law? While a few writers would accept that description,<sup>1</sup> I believe most people would reject it in both its aspects. They would claim that justice is quite different from revenge, and that there is more that marks out justice than the sovereignty of law: rather, law is the means whereby this important good, justice, is achieved. Following the 9/11 attacks, Amnesty International issued a position statement, *Pursuing Justice not Revenge* (2001). The statement makes no attempt to explain the distinction between justice and revenge. It characterises justice, by implication, largely in terms of internationally-recognized standards of procedural fair dealing. Yet it also speaks, as is usual in such cases, of “bringing to justice” those who committed the 9/11 attacks; and this cannot simply mean “bringing within the fair procedures” since everyone, including the innocent, is brought within those. (Or at best, Amnesty is equivocating between this

---

<sup>1</sup> An example is Barton 1999.

procedural sense and a retributive sense whereby “justice” would mean the desirable state of affairs in which the proven attackers receive judicial punishment.)

The idea of the good of justice is not reducible to assertions that actually use the word “justice”: arguments supporting or paraphrasing it may use other expressions. For example, a retributivist saying (as Moore does) that “punishment achieves the good of justice” may also say that “the good that punishment achieves is that someone who deserves it gets it.” (Moore, 87) But these claims are not logically identical, or interdependent. If someone produces a compelling argument to show that desert does not exist, that no-one deserves anything, that argument would not automatically invalidate any claim that punishment achieves the good of justice. The good of justice might be characterised without reference to desert: for example, as a state of harmony that must be restored whenever a crime has disturbed the social order. Conversely, the fact that the idea we are examining is sometimes supported or paraphrased without using the word “justice” does not render it unintelligible, any more than the fact that revenge is sometimes expressed without using the word “revenge” renders incoherent the idea of revenge. The crucial condition is that in both cases the canonical term denoting the idea must be evidently appropriate once the argument has been fully stated, even when it has not, or rarely, been employed in the course of the exposition. The upshot of the argument must be capable of being stated as: “here, then, we have the good of justice” (or “the evil of revenge”).

One more preliminary distinction needs to be made. In this field of theory it is conventional to distinguish “backward-looking” (retributivist) justifications of punishment from “forward-looking” (consequentialist) ones. Moore gives us a third possibility, however: a “present-looking” justification, in which the aim is the good state of affairs (GSA1) caused to exist when just punishment is enacted, prior to any consequent states of affairs (GSA2, 3 etc.) that may be aimed at. A justification on the basis of these further states of affairs would not be a retributivist justification: it would justify punishment instrumentally. I take GSA1 to

be the locus of the good of justice justification of retributivism. The GSA1 theory is not, of course, the only retributivist, non-instrumentalist theory of punishment. It is possible to conceive retributive justice, not as a state of affairs brought into existence by punishment, but as an attribute of the *act* of punishment. We can call this the strictly deontological, or non-state-of-affairs conception. On this view, it is right that we act to punish the guilty, irrespective of any state of affairs that may thereby come to obtain.<sup>2</sup> The strict deontological theory is simply that the authorised punisher acts rightly in responding to the offence with the correct intention-in-action. The justification lies not in any good that may inhere in GSA1 (though a retributivist of this kind might believe that GSA1 was also inherently good), but in the rightness of the intention-in action itself. *The just man justices*, as the poet Hopkins puts it, encapsulating this very idea by converting “justice” from a noun to a verb. (Hopkins, 1986, 129)

This justice-as-a-verb view might be defended in a number of ways: for example, by appeal to intuition, or to divine command. Justice would mean doing what we know in our hearts we ought to do, or what God tells us to do. Alternatively, it might mean doing what is required by the desert of the offender, or by his forfeiture of a right within some sacrosanct system of rights, neither of which requirements need strictly depend on any achieved state of affairs. All these widely-held non-instrumentalist justifications of retributive punishment I set

---

<sup>2</sup> I am far from sure (though the implications of this doubt lie beyond the scope of the present paper) that this distinction between an intentional action and the state of affairs that thereby comes to obtain can work for purposes of moral evaluation. Although performing an intentional action is indeed logically distinct from the state of affairs that obtains in virtue of its performance, it is difficult adequately to characterise, and *a fortiori* to evaluate, any intentional action without referring to what is intended by it. Moreover, the moral characterisation of an action as “just” (in contrast to, for example, “benevolent”) seems particularly to require attention to the state of affairs intended to be created by the action. (I am grateful to Terry Hopton for discussion of this point.)

aside for discussion elsewhere (with the exception of the divine command justification, which I would not know how to evaluate).<sup>3</sup> In the present paper I will focus on justice-as-a-noun: what Moore, Amnesty International, and countless victims and commentators, seem to be asking for when they say that offenders must be *brought to justice*: justice as the distinctively good state of affairs that we achieve through the intentional lawful punishment of offenders.

## II

According to the intrinsic good of justice claim, the state of affairs which obtains when we punish those who deserve it is good in itself: that is, it is good that it *obtains*, just as it is good that a calculation should be correct, or an argument valid, whether or not it is put to any instrumental good purpose. Justice is not good in virtue of its consequences for human happiness or alleviation of suffering, even if these consequences are often in fact good.

If justice is good, we need to consider when it would be right to create a just state of affairs at the expense of otherwise preferable consequences for human beings. Retributivists think the good of justice is a morally significant reason for action, and reasons for action must be capable of being weighed against one another, even if one is an intrinsic state-of-affairs reason and the other an expected-consequent-benefits reason. There can only be three possibilities:

**J1** The good of justice is of such moral weight that we should achieve it even if, by renouncing it, we could bring about an infinite increase in happiness or a complete alleviation of all suffering.

---

<sup>3</sup> See Rosebury 2011, 2015.

**J2** The good of justice has moral weight, but it only counts in favour of a course of action if all considerations of happiness and suffering are equally balanced.

**J3** The good of justice has a lesser moral weight than stated in J1, but a greater moral weight than stated in J2. Up to a point, it has sufficient weight to outweigh increased suffering or diminished happiness; but there is a threshold at which, in a given case, the situation in which there is not justice becomes the one we should choose, in preference to accepting the harms, or lost benefits, of the situation in which justice is preserved.

**J1** recalls a remark by J. H. Newman, to the effect that any amount of suffering is to be preferred to any wrongdoing.<sup>4</sup> It is also, I take it, Kant's reason for requiring murderers to be executed, irrespective of consequences. As he puts it, in a characteristically absolute formulation: "if justice and righteousness perish, human life would no longer have any value in the world." (1972: 104) It is not that justice has a higher arithmetic value than other morally significant goods such as happiness; rather, its holding is a moral requirement lexically prior to any calculation of such other goods. We are not permitted by the moral law to weigh happiness and suffering until we have established a just world. Whether Kant consistently held this view may be left to his commentators, but it is a possible view. Hardly anyone among modern retributivists holds it, however: Moore, for example, is careful to avoid it, acknowledging that "it is always relevant to ask 'at what cost do we attain some form of justice?'" (151)

---

<sup>4</sup> "The Catholic Church holds it better for... the earth to fail, and for all the many millions on it to die of starvation in extremest agony... that the one soul... should commit one single venial sin, should tell one wilful untruth, or should steal one poor farthing without excuse." (Newman 1967: 221)

**J2**, for all practical purposes, yields the wider argument to utilitarians. Almost all retributivists will therefore accept **J3**, and I will assume this view in later discussion.

Moore would, I believe, accept **J3**. Moore defines an intrinsic good as one the value of which does not depend on its being instrumental to another good (161). Intrinsic goods, such as retributive justice, or “pluralism, autonomy and tolerance,” are desirable states of affairs to bring about. What, then, is it about retributive justice that makes it *desirable that it obtain*, if it is not instrumental to those other goods, such as the reduction of suffering, which form part of the standard value-repertoire of utilitarianism and its ‘instrumentalist’ heirs? A devout theist once explained to me that when a crime is committed, the world becomes sick: the fitting punishment restores it to health.<sup>5</sup> This claim prompts the question: why does the affliction of the wrongdoer, who has himself previously afflicted others, make the world healthy again, rather than *more sick*? The defender of retributive justice must persuade us that some non-instrumental good inheres in the punishment of an offender, and that this good is sufficiently important that some “cost” in the form of other goods ought to be incurred to achieve it.

I shall try to show that there is no sense of an intrinsic good of justice that can count, without circularity, in favour of a retributivist position. Those who, on the strength of other arguments, already believe that it is intrinsically good that the deserving suffer punishment may use the commendatory expression “the good of justice” to characterise this event, but the

---

<sup>5</sup> Moore uses a similar metaphor. “For those like myself who are not theists, that cleansing function must be performed by the state, not God.” (152) It is possible that Moore means merely that the dispassionate agency of the state “cleanses” retribution of any suspicion of vengeful *ressentiment*, but the implication is that retribution aims at a state of affairs which can metaphorically be called cleaner, rather than even less clean, than that previously brought about by the crime.

invocation of justice adds only rhetorical dignity. In sections **III** to **V**, I will argue that the idea of justice, from Plato to the present, has only a small core of universally-agreed content, and that the construction around this common core of a variety of fuller conceptions gives no decisive advantage to retributivism: instrumentalist theories can as plausibly inherit the dignified rhetorical mantle of “justice” as retributivist ones, and indeed the former are widely-represented in the philosophical discourse that seeks to characterise justice. In section **VI**, I will argue that if, to avoid this stalemate, we simply stipulate “justice” as the term for punishment of the guilty as an intrinsic, non-instrumental objective, the price for our stipulation is to empty “justice” in this context of its rhetorical presumption of positive moral value. We must be prepared to conclude, counter-intuitive though this may sound, that what is just, in the stipulated sense, may be morally wrong.

### **III**

Is there an intrinsic, non-instrumental, good of justice, and if so, what is it? It is common ground that “justice” is used with normative intent, and that it is used in the context of human relations. So we can begin from the understanding that “justice” must denote some asserted principle of right action, or of some set or system of right actions, by one or more persons towards one or more persons, or some state of affairs that obtains in so far as this principle of right action has been followed. Plato takes this as his starting point in the opening pages of the *Republic*, and the still extant senses of “just” and “justice” in which these terms are hardly distinguishable from “righteous, morally right” and “rightful action” respectively have their prototype there. (Plato’s *dykaiosyne*, generally translated as “justice,” contains an additional idea of fitness or appropriateness to a certain harmonious order, as we will note later.) Aristotle similarly characterises justice as “complete virtue... in relation to another person.” (2000: 83; V i 1129b 30-31)

It is also, I believe, common ground that “justice” incorporates the idea of non-arbitrariness (which is another way of saying that it denotes a *rule* or *principle* of right action); and that the non-arbitrary character of actions towards others which are guided by justice consists in their being justified by some relevant intentional action by, or relevant interpersonal relation to, the other person. Justice excludes acting on an individual whim. The kinds of interpersonal relations which would provide non-arbitrary reasons for action towards others would be characterised in different ways by believers in, for example, (a) a caste system, (b) a society of equals arising from a social contract, or (c) the loving community of all humankind. But in every case, either the intentional actions of the other person, or some demonstrable feature of the social relations between one’s self and the other person, must form part of the justification for actions towards the other person. So for Aristotle, for example, it might not be unjust to treat slaves different from free men, but arbitrariness of treatment of either would be unjust: it is in this, arbitrariness-excluding, sense that “all men think justice to be a sort of equality... it should be equal for equals.” (1996: 79; III xii, 1282b 18-21)

Justice need not imply beneficence or altruism. A God who consistently afflicted his creatures by reason of their sinful deeds could be a just God, even if the consequence of his acting on this principle was a world of widespread misery. But a God who dispensed suffering arbitrarily could not be a just God, since he would not be acting on any principle of justice. Nor could a God who acted on some such principle as “afflict every ninth person,” or “afflict someone when I happen to feel in a vindictive mood,” or “afflict anyone whose suffering I find particularly amusing.” These are not principles of justice, since they do not appeal for justification to some relevant intentional action by, or relevant interpersonal relation to, the other person.

So much, I believe, is uncontroversial. Note that this common core, comprising some conception of *right action towards others* with *absence of arbitrariness* in our interpersonal

dealings, comes fairly close to the “fair procedures” construal of Amnesty’s requirement. There could not be any institutions of justice if we had no notion of right action towards others, or thought ourselves morally free to act arbitrarily. Can we go further in explaining justice? Among recent writers, Schmitz claims bluntly that “justice concerns what people are due.” (2006: 7) But this claim controversially locates a determinant of justice in *rights* or *entitlements* supposed to be possessed by persons. We could, on the contrary, believe with Bentham or Nietzsche that no-one is due anything, that rights are nonsense on stilts, and yet have a conception of justice that structures our actions. Such a conception of justice would turn, not on what people are due, but on how we should act towards people. These alternatives are not different sides of the same coin. The claim that I ought to offer Isabel one of my sandwiches is quite different from the claim that Isabel is due (in the sense of having a right or entitlement to) the offer by me of one of my sandwiches. The first claim says that I have a morally decisive reason to perform an action; the second equips Isabel with a right which is binding on me. The first could be asserted and the second, without contradiction, denied.

Here is an example of a conception of justice based on how we should act towards people. We could be unconditional altruists, taking our cue from Gibbon’s remark that “benevolence is the foundation of justice, since we are forbid to injure those whom we are bound to assist.” (1988, V: 238)<sup>6</sup> At the level of individual moral choice, each person would aim to contribute to justice by beneficent actions. At the collective level, justice would be

---

<sup>6</sup> Other examples of unconditional altruists are Schopenhauer (see, e.g., Schopenhauer 1995: 120-198); Buddhists; and those Christian churches ironically mentioned by Melville who preach “unconditional inoffensiveness by all to all.” (Melville 1988: 321) Though unconditional altruism sometimes appeals, as in these cases, to a religious or metaphysical foundation, in the main text I have framed it in entirely secular terms.

conceived as a system of social relations appropriate to a community of the mutually beneficent. Since mutual beneficence is in some contexts self-defeating (you and I can share the costs of our children's education, but we cannot both be the last person to enter a lifeboat), justice as unconditional altruism would embrace those principles of action, and those social institutions, which were best adapted to the aim of injuring none and assisting all. And since altruistic objectives often conflict with one another, and altruistically-motivated projects can misfire because we make poor epistemic judgements, or because of our limited capacities and opportunities, we would need to develop secondary, epistemic and prioritising, principles to guide our decisions.

Principled altruism so defined could reasonably be called a theory of justice, yet it would assign no necessary role to the rights or entitlements of others. Rather, it would say that the just person helps everyone and harms no-one, to the best of his or her ability; and the just society takes the good of all as its aim. And it would be, in its implication for criminal punishment, a categorical denial of pure retributivism: in telling us to aim to benefit every person to whom we have a relation, it would entail that aiming at the affliction of an offender could never be intrinsically good, but would be, at best, the preferable choice among alternative evils, defensible as creating overriding benefit for the offender or for others. The corresponding theory of social justice would assign to the social institutions the objective of helping and not harming every member of the community, limiting such aid only when unavoidable, and only for non-arbitrary altruistic reasons, such as the prioritisation of those least advantaged, or the withholding of benefits from some people where these benefits would carry heavy costs in harms to others.

Actual unconditional altruists, such as Schopenhauer, or Gibbon in the sentence just quoted, generally distinguish acts of justice, in which we resist the motives that would lead us to give priority to our own interests over those of others, from acts of altruism or benevolence,

in which we always give priority to the interests of the other. (Schopenhauer 1995: 148-167)

But at the social level this distinction becomes less important. Because, as we have noted, individual altruistic objectives can conflict, what emerges at the social level from unconditional altruism is a theory of justice in which each member of the society counts equally, because all members unconditionally value the interests of every member. The injustice of actual societies begins whenever one citizen successfully lays claim to count more than another, or when the aid extended by the community to one citizen is limited for some arbitrary or non-altruistic reason.

Justice as unconditional altruism is not the only alternative to a rights-based conception such as that of Schmitz. Here are some further examples.

1. Plato rejects the idea that justice, as a personal virtue, consists in giving people “their due.” Rather, it consists, he argues, in being a just man, in doing what is good and wise. As Socrates remarks, “it is not the property of the just man to treat his friend or anyone else badly. It is the property of his opposite, the unjust man.... So if anyone says that it is just to give everyone what is due to him, and if he means by this that what is due from the just man is harm to his enemies, and help to his friends, then whoever said this was not a wise man. What he said was wrong, since we have clearly seen that it is not just to treat anyone badly under any circumstances.” (2000: 12; book 1, 335d)<sup>7</sup>

2. Hume’s conception of justice is quite different from Plato’s: he sees it as a useful social instrument necessitated by the non-ideal conditions of human life (finite resources, and imperfect benevolence). But his theory is equally innocent of the notion of an intrinsic value in giving people their due. Though he acknowledges the affliction inherent in the punishment of criminals, Hume does not characterise this as a necessity of justice: on the contrary, he

---

<sup>7</sup> I note Plato’s social conception of justice at 16 below.

applies to criminal punishment the same description –the “suspension of justice”– that he applies to the state of war:

When any man, even in political society, renders himself, by his crimes, obnoxious to the public, he is punished by the laws in his goods and person; that is, the ordinary rules of justice are, with regard to him, suspended for a moment, and it becomes equitable to inflict on him, for the *benefit* of society, what, otherwise, he could not suffer without wrong and injury. (1998, 86: emphasis in the original)

Hume’s justification for punishment is not, then, that it complies with justice, but that it compensates for its absence: it is an emergency corrective, in a situation which makes it inadvisable for us to persevere with justice as he conceptualises it, that is as a system of mutual aid and co-operation.

3. For Hobbes, justice consists in performing covenants: therefore there cannot be justice in the state of nature.

For where no Covenant hath preceded, there hath no Right been transferred, and every man has right to every thing; and consequently, no action can be Unjust. But when a Covenant is made, then to break it is *Unjust*: And the definition of INJUSTICE, is no other than *the not Performance of Covenant*. And whatsoever is not Unjust, is Just. (Hobbes 1996, 100)

For Hobbes, giving people what they “are due” in some directly moral sense cannot be a feature of justice at all. We should give them what we are covenanted to give; and the morality of such transactions lies not in their corresponding to the moral desert of the recipients, but

instrumentally, in the contribution made by a system of covenants to “the Peace of Mankind.”  
(100)

#### **IV**

We can begin to distinguish between those conceptions of justice, like Hume’s, in which it is possible to act at once justly and wrongly and those conceptions, like Plato’s, in which this combination is excluded, either because justice is simply another name for right action, or because, on a more eclectic view, other criteria for right action must necessarily yield to that of justice. Further, we can distinguish between conceptions of right action which enjoin or permit retributive punishment, and those which forbid it. If we combine these distinctions, there are various possibilities, some of which I state below in their most basic forms. (“Right action” means “the morally best action all things considered”.)

**A1** Justice simply is right action, and right action may or must include retributive punishment.

**A2** Justice simply is right action, and right action excludes retributive punishment.

**A3** Justice is the overriding criterion of right action, and may or must include retributive punishment.

**A4** Justice is the overriding criterion of right action, and excludes retributive punishment.

**A5** Justice can be at variance from right action, but both permit or require retributive punishment.

**A6** Justice can be at variance from right action, but while justice permits or requires retributive punishment, right action forbids it.

**A7** Justice can be at variance from right action, but while justice forbids retributive punishment, right action permits or requires it.

**A8** Justice can be at variance from right action, and neither justice nor right action permits or requires retributive punishment.

Plato's view, at least as suggested in the passage discussed above, is **A2**. The "absolute" Kantian view is **A3** or perhaps **A1**. (Which it is depends on whether the Kantian we have in mind wants to use "justice" as the universal term of commendation.) Moore's view, since he allows that the cost of attaining justice may on occasion give us pause, could be expressed by **A5**. He holds that retributive punishment is required by justice and, fairly often, represents the right action. Hume's position in the passage just cited is **A8**, since the punishment he endorses during what he calls the "suspension of justice" serves social utility, not retribution. Rawls's views seem closest to **A4**: he states that "justice is the first virtue of social institutions," (1971: 3) but also that "a conception of justice is but one part of a moral view" (512); and in discussing just punishment (241), he adheres not to pure retributivism but to the "fair warning" view also familiar from Hart (1968) and Parfit (2011). **A7** lacks defenders, since those who hold that pure retributivism is right are extremely unlikely to renounce its characterisation as "justice." My own position, depending on terminological nuances, is either **A8**, or **A6** with "or requires" deleted, since I will accept that retributive punishment may be said, in one informative sense of the term "just" –to be explained later– to be just, but wrong.

## V

Some further distinctions can now be made. It will help to distinguish "justice" as attributed to actions by individuals, and "justice" as attributed to features of societies.

The conception of "justice" as right action by a person towards another or others has the following common variants:

**P1** treating others as they deserve to be treated

**P2** treating others as we should do when impartial as between their reasonable claims and our own

**P3** treating others as prescribed by some accepted system for interpersonal reciprocity which serves the general good

**P4** treating others as is fitting in light of their distinctive qualities or roles

These conceptions as applied to societies become:

**S1** assigning rewards and punishment according to desert

**S2** establishing and maintaining institutions designed to respect the reasonable claims of all equally

**S3** establishing and maintaining a well-ordered system of interpersonal reciprocity which serves the general good

**S4** assigning roles as is fitting to each person, and supporting these in a well-ordered system of social life

The retributivist asserts the intrinsic good of **P1** and especially **S1**.

Plato's view is best captured by **P4** and **S4**, though it is worth adding that he also sketches a conception close to **P3** and **S3**, when Glaucon suggests that the idea of justice in a society emerges as a compromise position, as the citizens come to realise they are generally better off renouncing the option of mutual antagonism, and Socrates responds with a vision of social development based on the advantages of co-operation and agreed specialisation of roles. (2000: 39, 51-55; 359a, 369b -372c)

Other conceptions of morally right action exist that stand apart from "justice" as analysed above. For example, in place of **P1** and **P2**, we might argue respectively for

**P1A** treating others *well, whatever* they deserve, and

**P2A** treating others as we should do when we attach *greater* importance to their well-being than to our own.

These are the principles of an unconditional altruism. If these are the best moral principles, then justice as defined in **P1** or **P2** is not identical with morality, but may conflict with it. We would then have a choice: either to accept a divergence between what is “just” and what is morally right, or to adopt **P1A** and **P2A**, in preference to **P1** and **P2**, as stating our conception of “justice” as right action by a person towards another or others. The latter choice would *identify* justice with unconditional altruism.

We might expect this analysis to be replicated for the social conceptions of justice, but there is an important difference. While unconditional altruism would reject **P1** and **P2**, and **S1**, it would be compatible with **S2**, **S3** and **S4**. The reason for this is that, if **P1A** and **P2A** are the correct conceptions of morally right action, then they apply to everyone: every person should treat every other persons in this altruistic way. In the transaction between *I* and *you*, I am the subject and you are the object of altruism; but if the society as a whole is considered, then all its members are equally the objects and the subjects of altruism. Thus the egalitarian **S2** and **S3** are on the face of consistent with **P1A** and **P2A**, while **S4** is not evidently in contradiction to it.

Since attempts to act altruistically towards allow others will often conflict with one another, this ideal of mutual altruism, where each is supposed to prefer the well-being of all the others to his or her own well-being, would require in any system of social relations to be stabilised and codified, just as its exact opposite, the idea of a society in which each seeks personal advantage at the expense of others, naturally generates the idea of a system of fair and consistent rules to adjudicate among the competing egoists. The conceptions of justice in **S2** and **S3** can emerge from this superimposition of altruistic and egoistic ideals in actual social interaction.

## VI

We have seen that there are competing senses of the normative expression “justice,” and that, in some of these senses, the attempt to realise justice does not entail a commitment to retributivism, and may even disallow it. A retributivist could reasonably reply, “Well, I am talking about the intrinsic good of justice in *my* sense.” But, thanks to the near-universal acceptance of its common-core meaning (right action without arbitrariness), combined with its elaboration as an unfailingly positive expression within various normative theories, “justice” has a rhetorical prestige that may mislead us. If the sense of “justice” being invoked is one that proves dissociable from morality, that sense may lack any claim to rhetorical prestige, in so far as that prestige arises from an implication of moral rightness. There is thus logical space for the position that retributive punishment, even if just, is wrong.

The distinctive claim of the “intrinsic good of justice” theory is that it is good that people get what they deserve, and that (as stated in **J3**) this good has sufficient weight to counterbalance considerations of increased suffering or diminished happiness, at least up to a certain point. The claim as it relates to punishment can be stated in a number of different versions. Here is one.

**G1** A world in which the happiness and suffering of individuals is perfectly matched to their morally right and morally wrong actions respectively is intrinsically better than a world in which there is no such match, even if the second world contains much more happiness, or much less suffering. We should aim to create the first kind of world in preference to the second, and our criminal justice system should contribute to this aim by ensuring that those who perform certain wrong actions are, in consequence, made to suffer.

This is formulated to refer to right and wrong actions, but could be revised to refer instead to virtue and vice, or to patterns of conduct rather than individual actions. I will not discuss these as separate alternatives, since the following discussion applies to all these versions. Similarly, nothing crucial will turn on the choice of “happiness” rather than “benefit,” “utility,” “preference-satisfaction” etc. as the positive term.

A retributivist need not, and probably will not, accept **G1**. She need not hold that the good of justice takes lexical priority over other goods, such as happiness, only that it has some independent weight. She must, therefore, at least hold that

**G2** A world in which happiness and suffering are perfectly matched to morally right and morally wrong actions respectively is intrinsically better than a world in which there is no such match, even if the second world contains some more happiness. We should aim to create the first world.

She might argue for **G2** as follows. Suppose that the surplus of happiness in the second world is provided solely by the immunity of a group of sadists who all escape punishment for their crimes, thanks to some laxity in the criminal justice system. In the first world they would commit the same crimes, so their own enjoyment and the suffering of their victims would be the same, but they themselves would in addition suffer the affliction of punishment. Can it be preferable that they should enjoy impunity for their crimes? Intuitively, this impunity seems clearly wrong.

To evaluate this argument we need to review the context in which an intuition is generated. If we strip away any real-life contexts, and simply isolate and compare the fates of the cruel criminals in the two worlds, it is indeed better that they should not suffer than that they should, and therefore better that their suffering should not be inflicted than that it should. Compare

*Micro-world One:* 100 cruel people enjoy themselves inflicting pain on 100 victims who eventually die, then they themselves are caused to suffer, then the world ends, with

*Micro-world Two:* 100 cruel people enjoy themselves inflicting pain on 100 victims who eventually die, then the world ends.

If we were obliged to create one or the other, we should, I believe, create Micro-world Two. By creating Micro-world One instead, our sole achievement would be to cause human suffering. The 100 victims are not helped by the retribution in Micro-world One. The impunity of the 100 cruel people is simply an absence of human suffering.

But if this artificially isolated sequence of events is now re-inserted into the context of a criminal justice system, in any world sufficiently similar to ours to give point to the invocation of intuitions, then for reasons which will strike any reflective person, and guide that reflective person's intuitions, it will seem wrong that there should be no punishment. Impunity takes on an objectionable character as soon as there is a background of social possibilities which create reasons for imposing punishment and for our expecting it normally to be imposed. In addition to familiar and important considerations of deterrence, incapacitation, maintaining respect for the law and so forth, there is a consideration provided by a "fair initial distribution" version of distributive justice. If the fair *status quo* in the distribution of goods has been disrupted by a criminal act (that is, an act by which one citizen breaks the implied contract, or tacit agreement of mutual forbearance, which makes society possible) and no steps are taken by the public authorities to vindicate the original distribution, then other citizens will feel abandoned by them: they will feel that unfairness has been rewarded. It is true that punishment of the criminal

may not in most cases be equivalent to restoration of the *status quo ante*,<sup>8</sup> but this matters less in motivating the intuition that punishment is right than the message of vindication for the initial distribution, or for the maintenance of only lawful methods of changing it, that conviction and punishment give. Such symbolic or actual repudiation of unlawful advantage makes the world go better, by helping to shore up the threatened solidarity of the social order: it is natural that we should believe strongly that such solidarity is desirable, and feel indignant if it is neglected. These “consequentialist” considerations, reinforced by the emotion of indignation, are more than adequate to explain the intuition against withholding punishment.

The retributivist might reply, “It is possible that some contribution may be made to our intuition by the considerations you mention, which understandably come to mind in the context of the criminal justice system. But their inadequacy in explaining away the core of the intuition is shown by another example, in which we remove that context and vindicate the intuition in its most direct form.”

*The terrorist's wife*

Acting secretly and extra-legally, you need to extract information from a terrorist in order to save a million lives imperilled by a nuclear device. Two strategies will work equally well: you can torture the terrorist, or torture his wholly innocent wife, to whom he is devoted. Because his wife is in better health than he is, she will recover more quickly from the effects of the torture. If you deny the intrinsic good of retributive justice, you are committed to

---

<sup>8</sup> See Hershenov for an attempt to assimilate the two. Some philosophers argue the weaker claim that the punishment for a given offence should be sufficient to provide a counter-motive against the advantages expected from the crime: see, for example, Beccaria 1995: 19-21, and Schopenhauer 1960: 102-103.

torturing the terrorist's wife rather than himself, since this will cause less total suffering in the long run, taking no account of the retributive consideration that he is guilty and she is innocent.

Examples of this kind are open to the objection that they exclude crucial real-world considerations, despite the fact that their usefulness in clarifying our intuitions depends precisely on our willingness to imagine that the situation is really happening. In this case, we should note that torture cannot be assumed to be a reliable way of extracting correct information, that we cannot really "know" that both options for torture will be equally effective, that we cannot really predict the longer-term physical and psychological after-effects of torture simply from observing a person's apparent state of health, that a man who loves his wife so much that he would abandon his commitment to terrorism to save her suffering would have been unlikely to undertake terrorism in the first place, etc. The example could be rewritten to eliminate these objections, but only at the cost of increasingly far-fetched scene-setting. By assuming secrecy and extra-legality, we have already abandoned the context of criminal justice policy: we are coming close to a position which is so far removed from the sphere of possible legislation and criminal justice process in a civilized society that its capacity to contribute to our understanding of it is very limited. Instead of being poised in reflective equilibrium between our intuitions and our laws and policies, we will have nothing to go on but our intuition itself.

Nevertheless, if for the sake of argument we take the example at face value, its implication can be accepted.

If we simply isolate and compare the suffering of the two parties, it is, I believe, better that the lesser suffering of the wife should be caused than the greater suffering of the husband. (It may be said, though this is also a consequentialist argument, that her suffering will really be greater because she knows she is innocent. But this is doubtful. Conceivably, she will be fortified by that knowledge.) Compare

*Micro-world Three:* Person X suffers greatly,  
with

*Micro-world Four:* Person Y suffers greatly, and for longer.

If we were obliged to create one or the other, we should create Micro-world Three. By creating Micro-world Four, our sole achievement would be to increase human suffering.

Of course, once we return from this stripped-down version to the imagined context of the example, our intuition is almost certainly the opposite. Any decent person, assuming such a person is capable of contemplating torture at all, will regard it as worse to torture the innocent wife. (We may ask whether there exists such a decent person. It may be that anyone who was not a sadist would need to be exceptionally cold-blooded to inflict torture, in which case the same cold-bloodedness might allow a calculation in favour of torturing the wife. Be that as it may, this is another indication of the unreliability of the example: the state of mind of the person supposed to be making the decision is one of which few of us have experience, and our intuitions are correspondingly untrustworthy).

There is a good moral reason, however, why, without being retributivists, we might conclude that it is better to torture the terrorist. The terrorist's wife has *grounds to complain* at such treatment, whereas he does not, since he has gravely afflicted others (or intends to do so), infringing the norms of reciprocal courtesy and peacefulness which help to hold society together, and so cannot consistently appeal to those norms against our threat of gravely afflicting him. From the fact that the offender has no grounds to complain if we afflict him, it does not follow that it would be right for us to do so. (It is precisely the challenge for the retributivist to explain why our emulating his willingness to afflict could be right, rather than a further wrong). But if we do decide to afflict the terrorist, we are not initiating a breach of reciprocal good social relations with *him*, whereas we *are* initiating such a breach if we afflict his wife.

In addition, there is a reason which is not moral but deserves a certain respect: we are likely to feel vengeful towards the terrorist, and not at all vengeful towards his wife. This may, in reality, be the dominant motive for our preference for torturing him rather than her, but it is not what makes our preference for sparing his wife the morally correct one.

A forward-looking, egalitarian distributive principle might next be invoked to support the good of justice claim. It might be argued, in the spirit of **S2d**, that the state should attempt, so far as possible, to equalise the happiness or welfare of citizens. Where A has afflicted B, therefore, the state should afflict A and compensate B, in order, so far as possible, to make their lives equal again.

**G3** A world in which happiness (or welfare, etc.) is equally distributed among individual human beings is intrinsically better than a world in which there is no such equality, even if the second world contains a little more happiness overall. We should aim to create the first kind of world in preference to the second, and our criminal justice system should contribute to this aim by afflicting those whose crimes have disrupted equality in that they have raised their own happiness while lowering someone else's.

Equality is a distinct value, yet so closely allied to many of our conceptions of justice that it may plausibly be argued to be the more basic value underpinning them. (Kymlicka 2002: 3-5) However, many crimes do not make the distribution of such good things less equal: they may incidentally, as when a poor man robs a rich man, make them more equal, but we do not for that reason refrain from punishment. Moreover, there are countless avoidable sources of inequality in the world, in addition to those created by crime on the one hand or by free unconstrained choices on the other. It needs to be explained why the state should intervene to correct inequalities caused by crime and not, say, by misfortune, particularly since the state action which requires defence in the context of punishment is the ugly business of inflicting

suffering, whereas state interventions to correct inequalities of another kind generally take the form of conferring benefits (for example, welfare payments to the unemployed).

The main precedent for disadvantaging some citizens in order to correct inequalities arising from misfortune in modern societies is provided by redistributive taxation, but such taxation does not (joking apart) constitute the infliction of suffering: withholding second helpings from the already well-fed is an act different in kind, not just in degree, from criminal punishments, and is not *prima facie* a challenge to the principle that afflicting another person is a bad thing to do.<sup>9</sup>

In response to these objections, the “good of justice” theorist, drawing now on **S2a**, may offer a revised claim.

**G4** Whatever the moral status of disparities of happiness among human beings in general, there is a special kind of wrong in those disparities that are caused by criminal actions by one or more persons to the detriment of others or against the community. Because they are infringements of the law, these are unjust disparities in a distinctive sense: over and above any material detriments that they cause, they involve the denial to some by others of the rights that all share as members of a law-governed society. Equal possession of rights is the value that constitutes the good of justice. The law must therefore vindicate the outraged rights of the victims, and restore equality among all citizens by proportionate penalties against the offenders.

---

<sup>9</sup> It is true that some fines imposed as criminal penalties may seem insignificant in comparison to some taxation rates. The distinction holds, nevertheless, because the fine takes something from the offender in addition to what he must expect to yield as a result of normal political processes. Where the fine is imperceptibly small given the offender’s wealth, it is significant that we may then plausibly protest that it is not really a punishment at all.

The problem with **G4** is that the anti-retributivist could accept it, provided that the vindication of rights, and indeed the institution of legal rights itself, are understood as being justified by the contribution they make to human happiness and well-being. The retributivist need not deny that there is such a contribution, but would need to add the claim that

**G4.1** The institution and vindication of legal rights are not to be justified solely by their good consequences, but as intrinsically just irrespective of consequences.

If this is to be more than an oblique rephrasing of the retributivist claim that punishment is justified, irrespective of consequences, by the fact that the offender deserves it, the idea of rights must function to support that claim in some new way. In many rights theorists, such as Hobbes, Locke and Rousseau, legal rights derive their moral authority from the supposed surrender of pre-existing natural rights by citizens entering into society. In that tradition, it might be argued that the social contract would have been broken if, after I had surrendered to the state my right to seek redress against another who harms me, the state were to fail to grant me that redress in the form of punishment. But this argument does not advance the retributivist's case: it simply transfers her difficulty to the state of nature. If punishment of an offender by the law is just because it is the vindication of a right held in nature by an individual person and then yielded to the state through the social contract, then that natural right must have been a right to punish. This is, in fact, Locke's view.

In transgressing the Law of Nature, the Offender declares himself to live by another Rule, than that of reason and common Equity, which is that measure God has set to the actions of Men, for their mutual security: and so he becomes dangerous to Mankind, the tie, which is to secure them from injury and violence, being slighted and broken by him. Which being a trespass against the whole Species, and the Peace and Safety of it, provided for by the Law of Nature, every man upon this score, by the Right he hath to preserve Mankind in general, may... bring

such evil on any one, who hath transgressed that Law, as may make him repent the doing of it, and thereby deter him, and by his Example others, from doing the like mischief. (1998, II, ss. 8-9, 272)

It is hardly necessary to point out that this passage is bursting at the seams with “consequentialist” reasons. The retributivist whose theory draws on a natural right to punish needs to explain why it exists independently of such reasons.

Locke himself may be thought to have provided such an explanation, in his suggestion that God has established the norm the breach of which justifies punishment. It would be possible to decouple our obligation to do God’s bidding from the consequentialist reasons which Locke clearly believes He had in mind. The retributivist who holds that in punishing the offender we are acting as God’s obedient servants cannot easily be refuted, but I will assume that this claim will not any longer satisfy philosophers, especially since the belief in a retributively just God can equally well motivate the view that human beings should not arrogate retribution to themselves. The adherent of such a view would still need to address the main challenge for a retributivist, by explaining why retribution by, or on behalf of, God is consistent with His supposed benevolence.

## **VII**

In conclusion: I do not believe there is a plausible conception of the intrinsic good of justice that independently strengthens, and does not merely restate, the claim that “the good that punishment achieves is that someone who deserves it gets it.” One can stipulate that by “justice” one means “the intrinsic, non-instrumental good that consists in the punishment of the guilty.” But that would add nothing new to our understanding. Justice, on that definition, is

simply a resoundingly honorific synonym, not a justification, for legal retribution against those who have afflicted others. Justice, on that definition, might be morally wrong.

*University of Central Lancashire*  
*School of Humanities and Social Sciences*  
*Preston PR1 2HE*  
*England*

## **REFERENCES**

Amnesty International. 2001. *Pursuing Justice not Revenge*. Index number: ACT 30/068/2001. <https://www.amnesty.org/en/documents/act30/068/2001/en/>, accessed 10 September 2018.

Aristotle. 2000. *Nicomachean Ethics*. Ed. and Trans. R. Crisp. Cambridge: Cambridge University Press.

Barton, C. K. B. 1999. *Getting Even: Revenge as a Form of Justice*. Chicago and La Salle, Illinois: Open Court.

Beccaria, C. 1995. *On Crimes and Punishment*. Ed. R. Bellamy. Trans. R. Davies. Cambridge: Cambridge University Press.

Gibbon, E. 1988. *The History of the Decline and Fall of the Roman Empire*. London: The Folio Society.

Hart, H. L. A. 1968. *Punishment and Responsibility*. Oxford: Oxford University Press.

Hobbes, T. 1996. *Leviathan*. Ed. R. Tuck. Cambridge: Cambridge University Press.

Hopkins, G. M. 1986. *The Oxford Authors: Gerard Manley Hopkins*. Ed. C. Phillips. Oxford: Oxford University Press.

Hume, D. 1998. *An Enquiry Concerning the Principles of Morals*. Edited by Tom L. Beauchamp. Oxford: Oxford University Press.

Kant, I. 1972. Justice and Punishment. Trans. W. Hastie. In *Philosophical Perspectives on Punishment*. Ed G. Ezorsky. Albany: State University of New York Press.

Kymlicka, W. 2002. *Contemporary Political Philosophy: an Introduction*. New York: Oxford University Press.

Locke, J. 1998. *Two Treatises of Government*. Ed. P. Laslett. Cambridge: Cambridge University Press.

Melville, H. 1988. *Moby Dick*. Ed. T. Tanner. New York: Oxford University Press.

Moore, M. S. 1997. *Placing Blame*. Oxford: Clarendon Press.

Newman, J. H. 1967. *Apologia Pro Vita Sua*. Oxford: Clarendon Press.

Parfit, D. 2011. *On What Matters*. Oxford: Oxford University Press.

Plato. 2000. *The Republic*. Ed. G. R. F. Ferrari. Trans. T. Griffith. Cambridge: Cambridge University Press.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Rosebury, Brian. 2015. Moore's Moral Facts and the Gap in the Retributive Theory. *Criminal Law and Philosophy* 5: 361-376.

Rosebury, Brian. 2015. The Theory of the Offenders Forfeited Right. *Criminal Justice Ethics* 34: 259-283.

Schmidtz, David. 2006. *Elements of Justice*. New York: Cambridge University Press.

Schopenhauer, Arthur. 1960. *Essay on the Freedom of the Will*. Trans. K. Kolenda. New York: Bobbs-Merrill.

Schopenhauer, Arthur. 1995. *On the Basis of Morality*. Trans. E. F. J. Payne. Oxford: Berghahn Books.