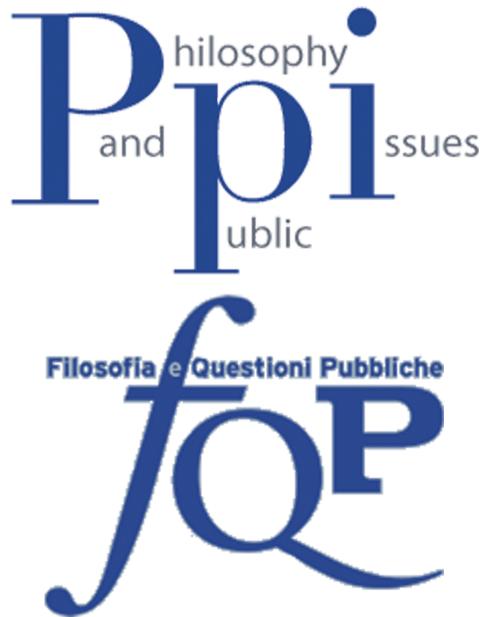


SYMPOSIUM  
PUBLIC ETHICS



Punishment and the Rebalancing of Status

BY

GERALD LANG

[THIS PAGE INTENTIONALLY LEFT BLANK]

# Punishment and the Rebalancing of Status

Gerald Lang

Jonathan Wolff's *Ethics and Public Policy: A Philosophical Inquiry* offers both an insightful and lucid primer on several problems in applied ethics and political philosophy—the regulation of drugs and gambling, safety standards, the distribution of health-care, the treatment of animals, the proper operation of the free market, and so on—as well as the salutary reminder that there is more than just philosophical theorizing which must be taken into consideration if philosophical arguments, and the philosophers who advance them, are going to stand any chance of influencing public policy.<sup>1</sup>

The division between pure theory and troublesome practice actually appears to play a smaller role, however, in Chapter 5, where Wolff considers crime and punishment. As ever in moral theorizing about punishment, problems as to the effectiveness of criminal sentencing are acknowledged, particularly in respect of ‘forward-looking’ theories of punishment (118-24). Along this particular dimension, the main questions Wolff pursues are these. How and when does punishment deter? How effectively does it deter? What assumptions must we make about the motivational profile of those who are likely to be deterred by the prospect of punishment, and can these assumptions be squared with what we take to be the proportionality conditions on punishment? Perhaps theorists of punishment have sometimes been content to

<sup>1</sup> Jonathan Wolff, *Ethics and Public Policy: A Philosophical Inquiry* (London: Routledge, 2011). In-text page references will be to this book.

speculate about such matters from the armchair, rather than through the consultation of empirical data. Be that as it may, the significance of empirical questions about the sources of motivation and the effectiveness of punishment as a deterrent were always staring such theorists in the face.

Turn now to ‘backward-looking’ theories of punishment, to the ranks of which Wolff adds his own theory. (The appraisal of this theory will occupy the bulk of my attention in this article.) How can such a backward-looking account of punishment avoid the charge of being nasty or obnoxious? This is also a wholly familiar problem facing backward-looking theorists.

In sum, the questions about punishment which Wolff raises and grapples with are these largely familiar ones. What deserves a much closer look is Wolff’s way of tackling those questions. As already indicated, I will be dealing mainly with the details of Wolff’s backward-looking account.

This article will unfold as follows. Section I will outline Wolff’s backward-looking account of punishment. Some critical questions for this account are raised in section II. The remaining part of the article will then attempt to show that Wolff is nonetheless indirectly getting at something that deserves to be taken seriously. In section III, I suggest that the prospects for an account broadly sympathetic to Wolff’s aims will be brighter if we pay greater attention to the basis of an offender’s liability and the connections between self-defence and punishment. Some brief conclusions are drawn in section IV.

## I

### **Wolff’s Rebalancing of Status Account**

Wolff offers a striking and ambitious theory of punishment,

blending elements of a ‘communicative’ theory (117, 125) and retributivism, and I will now quote from it at length.<sup>2</sup> It begins to be outlined in this passage:

[W]hat is so bad about crime, or at least some crimes for some people, is the fact of being made a victim. It is not so much that others attempt to treat you with contempt, but rather that they manage to do so. This is why... there is such a psychological difference between failed attempts and successful crimes. In succeeding in their crime against you, perhaps, they implicitly announce themselves as in some respect your superior. They have victimized you, and left you in a lower status. Even when there is no identifiable victim—as in the case of vandalism of public property—the successful criminal implies that in some sense he or she is above the norm, or, at least, above the rules. Crime communicates a message (124-5).

Wolff continues, in the paragraph below:

If [that is] so, then punishment appears in a new light. For at least part of the purpose of punishment then becomes to re-establish some sort of proper status between all the parties. If a criminal is caught and adequately punished he has no longer got away with something. He can no longer implicitly claim to be of higher status, and those who were victims may feel that their victim is expunged, and they have their previous status restored to them (125).

As for crimes without living identifiable victims, either because the crimes were so-called ‘victimless’ crimes or because they were homicidal:

The analysis [...] still applies in modified form. In the standard case, where there is an identifiable surviving victim, punishment ‘rebalances’ status by raising the standard of the victim and lowering that of the perpetrator. In cases where the victim is dead, nevertheless the punishment can still show that we as a society still take that life very seriously... In cases of victimless

<sup>2</sup> For a detailed example of a communicative theory, see, in particular, Anthony Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001).

crime, all that can be done is lower the standing of the perpetrator, but this remains a significant matter. The echo of retributivism is, presumably, the greater the crime, the more is needed to restore the moral balance. (125)

As we can see, then, Wolff's account is focused on the status of the offender and the victim: in a criminal act, the offender's status is heightened, and the victim's status is lowered. Punishment then rebalances the offender's objectionably heightened status, and the victim's objectionably lowered status, so that there is, once more, parity of status between them. I will call Wolff's account the *Rebalancing of Status Account*, or the *Rebalancing Account* for short.

## II Problems with the Rebalancing Account

By Wolff's own admission, the Rebalancing Account is 'speculative rather than research based' (127). But that feature seems unavoidable in any serious backward-looking account of punishment, and is nothing to apologize for, at least as I see it. Punishment involves the intentional infliction of suffering, or at least hard treatment, on offenders, and that feature raises profound moral problems that cannot be settled simply by empirical research.<sup>3</sup> The basic problem here is that, even in the unconstrained speculative space in which philosophers feel most at home, there is nothing approaching a secure consensus among them as to why it is morally permissible to inflict such hard treatment on offenders. So Wolff's argument for the Rebalancing Account will not be taken to task for that reason. That leaves us, however, with a few other problems.

<sup>3</sup> I take the useful phrase 'hard treatment' from T. M. Scanlon, 'Giving Desert Its Due', *Philosophical Explorations* 16 (2013): 101-16.

In what follows, assume that  $X$  is the offender, and that  $Y$  is the victim. Consider the following claims:

- (1)  $X$  criminally offends against  $Y$  at  $t_1$ .
- (2) It is appropriate to punish  $X$ , at  $t_2$ , for his criminal offence against  $Y$ .

Claim (1) is true by assumption.<sup>4</sup> Claim (2) might strike us, or many of us, as intuitively attractive, but we clearly need an argument for it. In particular, what is the argument for (2) submitted by the Rebalancing Account?

Our initial task is to get a bit clearer about Wolff's notion of 'status'. The preferred interpretation of Wolffian status must be answerable to several different constraints, or features. First, we need an interpretation of it which accommodates its alterability: status can go up, and it can go down. Call this the *alterability feature*. The alterability feature thus excludes a purely normative understanding of status which simply records an individual's worth as a moral agent and moral patient. (Even if (1) is true,  $Y$ 's worth surely does not decline, and  $X$ 's worth surely does not increase.) Second, our interpretation of Wolffian status must provide for the *interconnectedness* between the alterations in  $X$ 's status and  $Y$ 's status: *because*  $X$ 's status increases,  $Y$ 's status declines. Call this the *interconnectedness feature*. Third, it would be pointless to settle for an interpretation of status which simply *re-described* (2) in slightly different language. We need claims which can genuinely *illuminate* or *explain* (2). Call this the *explanation feature*.

With this slightly fussy stage-setting in mind, we can proceed. The following claims form the substance of the Rebalancing

<sup>4</sup> Further assume, to ease possible concerns, that the particular offence  $X$  commits against  $Y$  is one which any reasonable or sane jurisdiction would criminalize: violent unprovoked assault, for example.

Account, and will surely have a large role to play in how the Rebalancing Account delivers (2):

(A) Before  $t_1$ ,  $X$  and  $Y$  have the same status.

(B) After  $t_1$ , as a result of his criminal offence against  $Y$ ,  $X$  has a higher status.

(C) After  $t_1$ , as a result of her subjection to  $X$ 's criminal offence against her,  $Y$  has a lower status.

(D) After (proportional) punishment is imposed on  $X$ , at  $t_2$ ,  $X$ 's higher status is lowered to the status  $X$  possessed before  $t_1$ .

(E) After (proportional) punishment is imposed on  $X$ , at  $t_2$ ,  $Y$ 's lower status is raised to the status  $Y$  possessed before  $t_1$ .

Due to the interconnectedness feature, it makes sense to treat these claims primarily in pairs: (B) and (C) need to be jointly considered, as well as (D) and (E). We must also pay attention to the relationships between these pairs of claims.

How are (B) and (C) to be interpreted? It makes sense to isolate features of (B) and (C) which track, or reflect, the fact that  $X$  has wronged  $Y$ , since that will permit straightforward accommodation of the alterability and interconnectedness features; it will, furthermore, help us to see how (A) can be true; and it will help to pave the way for the explanation feature. But, to provide for the explanation feature, we also need to somehow *go beyond* these facts about wrongdoing, since the story about  $X$ 's wronging of  $Y$  does not, by itself, demonstrate why it is appropriate, permissible, or desirable to punish  $X$  for having wronged  $Y$ . This is just another way of saying that the truth of (1) does not give us any obvious basis for (2). We are appealing to the Rebalancing Account, after all, to plug the gap between (1) and (2).

Look again at what Wolff says in the passage above: offenders ‘implicitly announce themselves as in some respect [the victim’s] superior’ (125). In victimizing *Y*, *X* also implicitly conveys the attitude that *Y* is fit to be victimized, and that *X* is fit to victimize her. But even if this victimization story is true—even if it successfully captures *X*’s underlying attitudes to *Y*<sup>5</sup>—it is not immediately clear what role it can play in justifying punishment. That is, even if the victimization story takes care of (A), (B), and (C), it does not obviously take care of (D) and (E). To put the underlying point more sharply, we do not need punishment to tell us that *X*’s attitudes are morally flawed. But if we do not appeal to our practice of punishing murder and rape to reassure ourselves that murder and rape are wrong, then neither should we need to appeal to the practice of punishment to reassure ourselves that the attitudes which accompanied the original offences are morally flawed. Indeed—a stronger point still—it would be normatively back-to-front to appeal to the legitimacy of punishment to demonstrate that the offences which provoked punishment were wrong. So it seems odd to appeal to the practice of punishment to demonstrate that the attitudes which accompanied those acts were flawed. How could there be such a stark justificatory asymmetry at this point in the story? The prospects for the explanation feature now seem poor.

But perhaps Wolff’s deeper point is not that we need to appeal to punishment to vindicate the wrongness of the attitudes that accompanied the offences, but that punishment gives us an appropriate way of *publicly expressing* our conviction that these attitudes were flawed. If these attitudes are morally flawed, then it is not inappropriate to seek some form of expression for our

<sup>5</sup> Wolff does not pretend to have demonstrated that this account about the offender’s attitudes is necessarily manifested in his treatment of the victim (125).

condemnation. This is an important idea, and I shall be revisiting it in the next section. But it is not yet clear how the Rebalancing Account establishes that *punishment*—the imposition of hard treatment—is the proper expressive outlet. And it is far from clear how punishment can actually *rebalance* status. *Y* has been victimized, and this is morally objectionable; but how is *Y*'s status as a victim corrected, annulled, or cancelled out by *X*'s punishment? This is a significant lacuna for the Rebalancing Account. Wolff's story about rebalancing status is supposedly meant to tell us *why* punishment is the proper vehicle of that rebalancing; but it is none too clear how punishment achieves this rebalancing, because it is none too clear what, exactly, is being rebalanced.

We can concede to Wolff that a societal insult would be conveyed to *Y* if we failed to punish *X* for his offence against her, and yet continued to punish other tokens of that offence type against other victims. Here it is the *selectivity* of punishment which would indicate that we were taking *Y*'s victimhood less seriously than the victimhood of other individuals. In this connection, Wolff gives the example of victims from unpopular racial minorities, crimes against whom receive scant investigation or publicity (125). But that consideration alone will not serve to justify punishment *across the board*.

One further comment on the dialectic as it has unfolded so far. Some might think that *X*'s attitudes do not matter as much as *Y*'s attitudes. *Y* is the victim, and *Y* may have a very strong preference that *X* be punished for his offence against her. But that is not the story offered by Wolff, and for good reason. Unless we are committed, on independent grounds, to an industrial-strength version of an unrefined preference-satisfaction theory, it is implausible to hold that victims' preferences that

offenders be punished can justify punishment. We surely need a further account of *why* those preferences should be heeded.

Finally, a further problem concerns Wolff's treatment of the difference between attempted crimes and successful crimes. The challenge here is that a serious attempt at a criminal offence arguably reveals just as much contempt towards the would-be victim as a successfully completed crime. True, successful crime produces the injury to add to the insult, but Wolff's Rebalancing Account would seem, in any case, to be more concerned with the insult than with the injury.

Wolff happily admits that '[c]ontempt is shown even by an attempted crime' (115), but that '[a] successful crime cuts deeper' (115). He says:

In both cases one has been victimized, but there is an important difference. In the failed attempt, one may be rather shaken but there may also be a rather triumphal feeling 'I got the better of him!' When the crime is successful, there is no such comfort (115).

Furthermore:

[W]hen the attempt is successful perhaps one begins to harbour the thought that the contempt is deserved. If I am unable to protect myself, what sort of person am I? A successful crime seems, in at least some cases, to bring about a change in status and in self-respect (116).

These passages give expression to a variety of different ideas. Now it would be foolish to deny that the injury produced by successful attempts, combined with the contempt they express, can severely affect a victim's sense of self-respect and self-worth. This is truly a poisonous cocktail for victims to deal with. But there is also a hint of strain in the suggestion that victims of successful attempts will harbour 'the thought that the contempt is deserved'. If a victim does not deserve to be attacked in the first

place, it would be strange to think that the difference between an offender's failed attempt and his successful attempt could nonetheless reflect the victim's desert, and it would be surprising if victims standardly came to that distorted view of themselves. The difference between failed attempts and successful attempts is an important one, as we will see in the next section, but it seems to me that Wolff does not quite manage to exploit it in the right way.

So, to sum up. Wolff writes:

If crime sends a message, then so does punishment, standardly an attempt to send a counter-message, cancelling out the first message. (125-6)

Even if there something attractive about this claim, there are at least two major problems for the Rebalancing Account. First, it is not clear why a 'counter-message' needs to be sent out at all; and second, it is not clear how that counter-message can 'cancel out' the first message issued by the offender.

### III

#### **The Extended Liability Account**

I have suggested that there is no obvious interpretation of status in the Rebalancing Account that can serve Wolff's purposes. Even if we are strongly persuaded by the claim that it is appropriate to punish *X* for what he did to *Y*, we lack a convincing status-based explanation of that claim: we cannot plausibly secure (2) by saying that punishment reduces *X*'s objectionably heightened status and raises objectionably lowered *Y*'s status. We should therefore de-emphasize the notion of status in any putative justification of punishment.

What should we appeal to instead? The account I propose will retain Wolff's investment in the story about the communicative significance of punishment, but jettison his investment in the story about rebalancing status. However, we need to start yet further back. I propose that we start with an account that blends backward-looking and forward-looking considerations: namely, an account which seeks ultimate grounding for the permissibility of punishment in the permissibility of self-defence.<sup>6</sup> More specifically, I suggest that we begin with *X*'s liability to self-defensive action. Now there are many ways of grounding defensive permissions and attacker's liability. I will not be concerned here with these intramural debates among self-defence theorists; my aim is simply to argue that, however the attacker's liability is to be grounded, it is implausible to hold that his liability evaporates altogether when his offence is completed.

Imagine, in some very nearby possible world to the world that was described by (1) and (2), that *X* is in the course of criminally offending against *Y*. (To be more specific, imagine that *X* is about to assault *Y*.) On the basis of *X*'s attack on *Y*, we seem able to say the following:

(3) *Y* is permitted to take necessary and proportionate violent self-defensive action against *X*.

And, as an extension of (3), we can also say:

<sup>66</sup> See, for example, Warren Quinn, 'The Right to Threaten and the Right to Punish', *Philosophy & Public Affairs* 14 (1985): 327-73, and Daniel Farrell, 'The Justification of Deterrent Violence', *Ethics* 100 (1990): 301-17. David Boonin, *The Problem of Punishment* (Cambridge: Cambridge University Press, 2008), 192-211, provides a detailed critique of this 'self-defence solution' to punishment. I lack the room here to engage with the full details of Boonin's critique; I can only set the ball rolling in the right direction, by arguing that *X*'s liability to defensive attack establishes a presumption that *X* is also liable to punishment.

(4)  $Z$  (a third party) is permitted to take necessary and proportionate other-defensive action on behalf of  $Y$  against  $X$ .

And, as a way of explaining how these permissions arise, we can say:

(5) In virtue of his attack on  $Y$ ,  $X$  is liable to be attacked in self-defence (by  $Y$ ) or in other-defence (by  $Z$ ), subject to the necessity and proportionality conditions on self-defence and other-defence.

Of course, the world that was described by (1) and (2)—the ‘punishment world’—is not the ‘defensive world’ jointly characterized by (3), (4) and (5). In the punishment world, the imminent threat to  $Y$  has been and gone. It is therefore too late to apply self-defensive or other-defensive violent action to avert  $X$ ’s attack on  $Y$ .<sup>7</sup> The point is, by now, a fairly familiar one in the literature, but it bears blunt restatement:

When Moe harms Larry in self-defense, he harms Larry in order to *prevent* Larry from wrongfully harming him. But when the state punishes an offender, it punishes him precisely because he has already *succeeded* in wrongfully harming someone. It is easy to see how the notion of self-defense can justify harm to prevent a particular wrong from taking place. But it is far more difficult to see how an appeal to self-defense could be used to justify inflicting harm in response to a particular wrong when it is already too late to prevent that wrong from taking place and thus too late to provide a defense against it.<sup>8</sup>

Can the moral content of (5) nonetheless cast useful light on the permissibility of punishment? I believe it can.

Imagine that  $X$ ’s attack on  $Y$  begins at  $t_0$ , and is successfully completed at  $t_1$ , as (1) stipulates. Between  $t_0$  and  $t_1$ ,  $X$  is liable to

<sup>7</sup> To save words from now on, the phrase ‘self-defensive or other-defensive’ will be compressed into, simply, ‘defensive’.

<sup>8</sup> Boonin, *The Problem of Punishment*, 193-4; original emphases.

defensive attack. From  $t_1$  onwards,  $X$  can be no longer be liable to defensive attack, for the simple reason that the conditions applying to defensive attack no longer obtain. But it is intuitively odd to think that  $X$ 's moral standing is no longer impaired, just because  $X$  has now succeeded in an attempt which he was morally forbidden from making in the first place.<sup>9</sup> Plausibly,  $X$ 's overall moral standing is even worse after  $t_1$ , since the successful completion of his criminal attempt to  $F$  is morally worse than his mere attempt to  $F$ . That is:

(6)  $X$ 's successfully completed offence against  $Y$  is a morally worse action than  $X$ 's mere attempt to commit an offence against  $Y$ .

But what should we conclude from (6)? Even if it is true of  $X$ , after  $t_1$ , that  $X$  was liable to defensive violence between  $t_0$  and  $t_1$ , what further relevance does that fact have? On one view, that fact has no further relevance:  $X$ 's liability to hard treatment is exhausted by his liability to defensive attack, which obtains only in the interval between  $t_0$  and  $t_1$ . After  $t_1$ , in the world in which  $X$ 's attempt is successfully completed,  $X$  is simply not liable to hard treatment. Call this the *No Further Implications View*.

The No Further Implications View strikes me as intuitively implausible. But it will not do, in this context, to rest the argument squarely on confidence in that verdict. Friends of the No Further Implications View may reply that, even though  $X$  is not liable to any hard treatment upon expiry of the conditions for defensive attack,  $X$ 's moral standing is still impaired, *simply because*

<sup>9</sup> See Gerald Lang, "Why Not Forfeiture?," in *How We Fight: Ethics in War* (Oxford: Oxford University Press, 2014), ed. H. Frowe and G. Lang, at 43. My particular concern, in that discussion, was with the prospects for a forfeiture account of self-defence; now I am trying, on the assumption that there is some or other secure basis for the liability of an individual to defensive attack, to enlarge the basic story to encompass the individual's liability to punishment.

(6) *is true*. If (6) is true, then *X* can still be *blamed* for what he did, even if he cannot, on that basis, be *punished* for what he did.

I am not persuaded by this counter-response, for reasons that are partly connected with something that is laboured by Wolff: the communicative significance of punishment, or (to put it more neutrally) the communicative significance of having *some* effective form of social condemnation of criminal wrongdoing. On the No Further Implications View, *X* can be blamed, but not punished. However ‘hot’ blame can be—however ardently or forcefully it can be conveyed—it cannot take the same form as the hard treatment (in this case, defensive violence) that would be appropriate had *X* not successfully completed his attempt to wrong *Y*. But then it will seem that, by comparison with the sanctions available for dealing with the normative situation between  $t_0$  and  $t_1$ , *X* is actually being rewarded for the successful completion of his criminal attempt against *Y*. If this ‘sends out a message’, then the message it sends out is that *X*’s position is improved if he does something that is morally worse.

To recapitulate, there are at least three problems to worry about if the No Further Implications View is taken to be true. First, public condemnation of *X*’s offence will be largely drained of force, conviction, or efficacy. The condemnation will seem relatively toothless. Second, the relative toothlessness of the public condemnation of *X*’s offence may contribute to, and help to sustain, *Y*’s feelings of psychological and social fragility. Third, this feature also seems apt to generate a perverse personal incentive for *X* to do something that is morally worse.

To escape these implications, I propose that we embrace the following:

(7) The facts which make *X* liable to defensive attack between  $t_0$  and  $t_1$  also make *X* liable to punishment after  $t_1$ .

We can refer to the account I have sketched, in admittedly embryonic form, as the *Extended Liability Account*. The Extended Liability Account plainly differs from the Rebalancing Account. But, as I have already hinted, I think there may be underlying similarities between them. I try to identify what they are in the concluding section.

#### IV

#### **Conclusion: Rebalancing the Rebalancing Account**

I have argued that Wolff's Rebalancing Account fails to deliver what he is after: namely, a secure backward-looking account of punishment. My view is that, if an offender is liable to the hard treatment constituted by defensive force, then he is also liable to the hard treatment constituted by punishment. As I see it, punishment does not rebalance the respective statuses of the offender and the victim. That is an essential element of Wolff's story, and it is no real part of mine.

However, the Extended Liability Account also leans, at certain points, on similar intuitions as Wolff's about the communicative or expressive significance of punishment; and it aims to respect and uphold the victim's status by helping to confirm that the offender's completed offence leaves him in a worse moral and personal position than he was in when he had not yet completed that offence. As different in crucial aspects as they may be, then, the Rebalancing Account and the Extended Liability Account still draw water from the same well.<sup>10</sup>

*University of Leeds*

<sup>10</sup> I thank Corine Besson, Jimmy Lenman, and Fiona Woollard for helpful conversations about these issues.

If you need to cite this article, please use the following format:

Lang, Gerald, "Punishment and the Rebalancing of Status," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 3 (2014), 53-67, edited by S. Maffettone, G. Pellegrino and M. Bocchiola