

Free Will and the Asymmetric Justifiability of Holding Morally Responsible

Abstract

Even if we do not follow Susan Wolf in holding that praise and blame have different control conditions, there is still an important praise/blame asymmetry which must be recognized. That is, even if praise and blame have the same control condition, we must have stronger reasons for believing that it is fulfilled to treat someone as blameworthy than to treat someone as praiseworthy. Blaming behavior which involves serious harm can only be justified if the claim *that the target of blame acted freely* cannot be reasonably doubted. But harmless praise can be justified so long as the claim *that the candidate for praise did not act freely* can be reasonably doubted. Anyone who thinks the debate about whether we have free will is truth-conducive has to acknowledge that reasonable doubt is possible in both these cases.

Introduction

Among Susan Wolf's contributions to the free will debate is to highlight the intuition that we must meet a higher standard of justification to blame than to praise.¹ Wolf points out this intuition in the course of arguing that the control condition of moral responsibility is different in the contexts of praise and blame. (On her view, people can be justifiably blamed only if they could have done otherwise, but can be justifiably praised even if they could not have done otherwise.) But apart from Wolf's work and responses to it, there seems to have been little discussion of this intuition. This is surprising, because it seems to have interesting implications for free will theory all on its own, even if one does not follow Wolf in holding that the control condition of moral responsibility is different in the contexts of praise and blame. The purpose of this paper is to develop this point.

In this paper, the term "free will" is used to refer to whatever satisfies the control condition of moral responsibility, and "acting freely" means acting with free will.² Also, the terms "blame" and "praise" will be used to refer to actions used to express the negative and positive reactive attitudes. As I am using these terms, "blame" and "praise" involve not just *believing* that someone is morally responsible, but acting in such a way as to *hold* him morally responsible. (It is not claimed that blaming and praising are essentially matters of acting in particular ways—this usage is only meant to simplify exposition.)

The main claim of the paper is twofold: first, that seriously harmful blame is only justifiable if it cannot be reasonably doubted that the target of the blame had free will; and second, that costless praise is justifiable so long as it can be reasonably doubted that the candidate for praise did not have free will.

1. The Asymmetry Intuition

The intuition that praise and blame are asymmetrically justifiable can be explained as part of a more general intuition that harms and benefits are asymmetrically justifiable. Justice demands that arguments for harming people be to be held to a higher standard than arguments for refraining from harming them or benefiting them. All philosophers must acknowledge that this asymmetry exists, though disagreement is to be expected when it comes to giving a detailed explanation of why it exists. All will probably agree that, one way or another, the primary purpose of morality has to do with getting people to benefit and refrain from harming each other. If this is right, then arguments for harming people run against the primary purpose of morality (in at least a *prima facie* way), while arguments for benefiting people run with it. There is therefore a moral presumption in favor of refraining when it comes to harm, and in favor of

acting when it comes to benefit. If everyone agrees that some action under consideration would be a harm, the burden of proof is on whomever claims that this action would be legitimate. But if everyone agrees that some action under consideration would be a benefit that could be provided at no significant cost to anyone, the burden of proof would be on anyone who claims that it would be wrong to act in this way.

When praise and blame are successful, that is, when they have their intended effects, they are kinds of benefit and harm. That is, even in cases where praise and blame do not include the obvious benefit and harm of reward and punishment, they are intended to cause valuable or harmful emotional responses in the people who are their recipients. So the asymmetry in the justificatory standards for harm and benefit seems to imply an asymmetry in the justificatory standards for praise and blame.

Praise and blame are special kinds of benefit and harm, since they can only be legitimate if the people who are their recipients deserve them are morally responsible for the actions to which they are responses. Some might wonder whether this makes praise and blame make praise and blame so different from other kinds of harm and benefit that the justificatory asymmetry relevant for other kinds of harm and benefit has no bearing on the justifications for praise and blame. But I cannot see why this should be so. It is a straightforward matter to include claims about moral responsibility within the scope of the justificatory asymmetry. We can hold justifications for blame to a higher standard than justifications for praise by holding all the claims that play roles in justifications of blame to a higher standard than the claims that play roles in justifications of praise. Since claims about moral responsibility play roles in justifications of both praise and blame, those claims must be held to a higher standard when they appear in justifications for blame than when they appear in justifications for praise.

How much higher is the justificatory standard for claims about moral responsibility in justifications for blame than in justifications for praise? If the praise/blame asymmetry exists because of the harm/benefit asymmetry, then it seems natural to suppose that the standard for praise is a slope that rises as the cost of that praise increases, and that the standard for blame is a slope that rises as the harmfulness of blame increases. But it is beyond the reach of this paper to survey this whole continuum. Here the argument will focus on the standards for two kinds of cases at the extremes of the continuum: seriously harmful blame, and praise that can be given at no significant cost to anyone.

It is probably clear enough how blame can be seriously harmful. Blame can include retributive bodily violence. Blame can also include the retributive infliction of great emotional pain as punishment for bad actions, and this emotional violence can sometimes be even more harmful than bodily violence.³ Other kinds of seriously harmful blame include retributive punishment by execution, and by imprisonment under dreadful conditions that make rehabilitation all but impossible (like those that prevail in contemporary prisons).⁴

The idea of praise which can be given at no significant cost to anyone is probably less familiar. What I have in mind is praise that benefits the recipient and doesn't do any significant harm to anyone. For example, it doesn't make the person giving the praise miss a significant opportunity to do something else worthwhile. Also, there are no third parties who are harmed in any significant way by the praise. That is, if we select one person from a group of people to praise her for doing something good, this can cause the other people in the group to feel sorrowful about not having done well enough to merit recognition. The valuable experience of the person singled out comes at the expense of the painful experiences of the others. But there are surely possible instances where costs like these could be avoided. There are probably some

actual instances too. We can sometimes single out individuals privately so that others' feelings are not hurt. On other occasions, we can be egalitarian with praise. Suppose everyone in the group has tried to do good things. We might praise them all for trying.⁵ It may well be that everyone who can be hurt by being excluded from praise sometimes tries to do good things. (That is, it may well be that everyone sometimes tries to do good things except sociopaths who do not care whether or not they are praised.) If this is right, then we might maximize the amount of costless praise we give by distributing praise in an egalitarian way among everyone we meet who sometimes tries to do good things.

In what follows, for the sake of brevity, seriously harmful blame will be referred to as "serious blame", and praise that can be given to someone at no significant cost to anyone will be referred to as "costless praise".

2. Serious Blame and Reasonable Doubt

In this section, it will be argued that serious blame is only justified if it cannot be reasonably doubted that the target acted freely.⁶ If serious blame is not justified, then there is an obligation to refrain from it. Call this the *serious blame principle*.

The serious blame principle should be accepted because of its close kinship to another "reasonable doubt" principle which is widely recognized to be a requirement of justice, that is, the requirement in criminal legal proceedings that the accused can only be convicted of a crime if it is proven beyond reasonable doubt that he acted criminally.

The conviction standard and the serious blame principle are both grounded on the same basic intuition about justice. The intuition is really just a further specification of the intuition described earlier about the asymmetrical justifiability of harm and benefit. Justice requires

arguments for harming people to be held to a higher standard than arguments which are not for harming anyone, and it requires arguments for seriously harming people to be held to an especially high standard: there must be no room for reasonable doubt about their soundness. This holds whether the harm at issue is blame or of some other kind.

In courts of law, an argument that someone has committed a crime is typically part of a larger argument that that person is to be given a punishment which will cause serious harm such as imprisonment (or even death, in some countries). This is why it is taken to be a requirement of justice to hold arguments that someone has committed a criminal act to the "reasonable doubt" standard. When a claim about free will serves as a premise in a justification for serious blame, it must be held to the same standard, for the same reason. That is, in this context, the claim that someone has free will plays a role in an argument for serious harm, just as the claim that someone has committed a crime typically does. For this reason, it must be held to a standard based on reasonable doubt, just as the claim that someone has committed a crime must be. So the serious blame principle has the same justification as the criminal conviction standard.

You don't have to be a free will skeptic to accept that it can be reasonably doubted that anyone ever has free will. I think compatibilists and libertarians who think the free will debate is truth-conducive must accept that free will skeptics' doubts are reasonable (though I won't argue this here). If this is right, and the serious blame principle is also right, then compatibilists and libertarians who think the free will debate is truth-conducive must accept that serious blame is not justifiable.

3. Costless Praise and Reasonable Doubt

In this section, it will be argued that costless praise is justified so long as it can be reasonably doubted that the candidate did not act freely. Call this the *costless praise principle*. The costless praise principle does not imply an obligation in the way the serious blame principle does, since the fact that costless praise is justified in some situation would not imply that it was obligatory.

The argument for the costless praise principle has to be quite different from the argument for the serious blame principle, because there is nothing in jurisprudence which might play a role parallel to the one played by the criminal conviction standard earlier.

As a first step, it may be worth pointing out that if reasonable doubt standards govern both serious blame and costless praise, then there is a symmetry within the asymmetry. That is, even though the standard is higher for blame than for praise, they are both reasonable doubt standards.

Pursuing this symmetry may seem to come at the price of setting a peculiarly low standard for costless praise. But recall the perspective on costless praise suggested earlier. Costless praise is a kind of costless benefit. When all parties agree that an action under consideration is a costless benefit, we usually expect no further justification for it. Costless praise cannot be supposed to be entirely typical in this regard, of course. Costless benefits are typically justified merely by the claim that they are costless benefits, so people who have no reason to doubt this claim have no reason to ask for any further justification. But costless praise must be seen to rest on a further claim, that is, a claim that the candidate for praise is morally responsible for the action at issue. If one doubts this further claim, it makes sense to ask for a

defense of it, and part of the defense must involve giving some reason to suppose that the candidate acted freely.

But when it is asked how high a justificatory standard to apply to the claim that the candidate had free will, it makes sense to look at the moral context in which the question is asked. It seems important that the context is an attempt to provide someone with a kind of costless benefit. If the primary purpose of morality is to get people to benefit and refrain from harming each other, then the context is an attempt to do something that helps fulfill the primary purpose of morality. So it seems unreasonable to impose a high standard. If the costless praise principle is correct, then the standard is very low, i.e. that the falsity of this claim's negation can be reasonably doubted.

There are, however, many degrees between high and very low. Why should we pick a very low standard? Consider the following scenario. Suppose a person with a neurological disorder that causes intermittent problems with intention formation and motor control knocks someone else out of the path of an oncoming truck and saves his life. Suppose it is not clear whether or not he acted intentionally, and there is no way to settle the question. (The disorder also sometimes causes confusion and memory difficulties, so we cannot take our potential hero to be authoritative on this point.) But he would not object to being given the benefit of the doubt and praised as an actual hero. He has few opportunities to be praised, so it would bring him no small happiness to be praised in this case. Suppose that there is no reason to doubt that he satisfied whatever other conditions of praiseworthiness there may be in addition to the control condition. Finally, suppose that praising him would be costless—no third parties would feel slighted, and praising him would not deprive us of an opportunity to do something else worthwhile.

Even if we put metaphysical worries about free will aside, there are good reasons to doubt that the potential hero acted freely, since even free will theorists who disdain metaphysics agree that one cannot have free will with respect to anything that is not an intentional action. Yet there seems to be a strong intuition in favor of the view that costless praise of the potential hero as an actual hero is justified, because it is not clear that he did not act with free will, and he deserves the benefit of the doubt.

Next consider whether this intuition would change if we had more detailed information about the probability that he acted freely. Suppose that we could know that there was a 10% chance that he acted freely. Would we be inclined to praise him then? I think so. How about a 1% chance? Since the praise is costless, and the potential hero benefits from being treated as an actual hero, it is not clear that anything could give us a reason to accept 10% as good enough, but not 1%. This line of thought pushes the standard downward until we reach the reasonable doubt threshold. No reasonable person could justifiably praise him if it cannot be reasonably doubted that he did not act freely. But since the praise is costless, it would seem that a reasonable person could justifiably praise him so long as it could be reasonably doubted that he did not act freely. If this is right, then the costless praise principle is correct.⁷

You don't have to believe in free will to accept that it can be reasonably doubted that someone didn't have free will. I think free will skeptics who take the free will debate to be truth-conducive must accept that free will believers' views are reasonable (but once again I won't argue for this here). If this is right, and the costless praise principle is also right, then skeptics who think the free will debate truth conducive should probably accept that costless praise is justifiable in at least some cases.

Conclusion

To conclude, I want to point out that there may be other issues in free will theory where reasonable doubt standards play a role. Consider the "ought implies can" principle. This principle implies that I should only believe I ought to do something if I have reason to believe that I can do it. But how strong does this reason have to be? Do I need to establish that I cannot reasonably doubt that I can do it? Presumably not. I can reasonably doubt that my car will start tomorrow, and if my car does not start I cannot get to campus to teach my class, but presumably this does not conflict with my belief that I ought to teach my class tomorrow. Do I need to establish that I have better than a 50% probability of being able to teach my class to believe that I ought to do it? This standard also seems too high. Perhaps all that is required is that it is possible to reasonably doubt that I cannot do it.⁸

NOTES

¹ She does not speak in exactly these terms, but I take this to be one of her points. In "Asymmetrical Freedom" (*Journal of Philosophy* 77.3, 1980), for example, she writes the following:

we have stronger reasons for wanting acts of blame to be justified [than we have for wanting praise to be justified]. If we blame someone or punish him, we are likely to be causing him some pain. But if we praise someone or reward him, we will probably only add to his pleasures. To blame someone undeservedly is, in

any case, to do him an injustice. Whereas to praise someone undeservedly is apt to be just a harmless mistake. (p. 156)

I take it that Wolf's claim that we have stronger reasons for wanting acts of blame to be justified involves the idea that acts of blame require a stronger justification than acts of praise.

² These terms are used in a broad sense meant to cover all the various accounts of the control condition of moral responsibility, including strong libertarian notion such as dual control, and weaker compatibilist notions such as guidance control or hierarchical control. For guidance control, see John Martin Fischer and Mark Ravizza, *Responsibility and Control: a Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998). For hierarchical control, see Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68.1 (1971).

³ In my view, not all remorse is based on self-blame. In [author's paper 1], I argue that there is a kind of remorse which is based on suffering in sympathy with the person one has harmed rather than on self-blame.

⁴ Punishment does not have to be retributively justified, and the argument of this paper is only relevant for retributive justifications of punishment. For a recent alternative, see Pereboom's quarantine justification in his book *Living Without Free Will* (Cambridge: Cambridge University Press, 2001).

⁵ Egalitarian praise may not be costless when people who achieve more than others protest that they deserve to be singled out for praise in a way that excludes the others. I do not want to take a position on the legitimacy of this protest (though I am suspicious of the claim that egalitarian praise is ever unjust, and I am inclined to see sorrow caused by not being included in praise as

more ethically important than sorrow caused by having to share praise with others). For present purposes, I only wish to claim that, in cases where people who achieve more than others choose not to protest egalitarian praise (perhaps out of magnanimity), they have not been treated unjustly, and there is no obstacle here to costless praise. (Also see note 10 below.)

⁶ The argument of this section is based on a longer, more detailed argument which I present in [author's paper 2].

⁷ Suppose that someone else, who has no such disorder, was praised for rescuing a different person from certain death last week. He might be pained if the potential hero got the same kind of praise he got—he might object that there is more reason to doubt that the potential hero acted freely than there is to doubt that he acted freely, so it is unfair to give the potential hero the same kind of praise he got. I have my doubts that this would be unfair. But as far as the argument of this paper is concerned, I am willing to concede that praise would not be costless in this case. It is enough to point out that last week's rescuer might be magnanimous enough not to be pained or to object, and that in this case, it is possible to give our potential hero the same kind of praise last week's rescuer got without being unfair. (Also see note 7 above.)

⁸ I discuss the problem of understanding the "ought implies can" principle in the context of free will skepticism in more detail in [author's paper 3].