# Free Will and the Asymmetrical Justifiability of Holding Morally Responsible

*The Philosophical Quarterly*, October 2015, Vol. 65, No. 261, pp. 772-789.

Benjamin Vilhauer, City College of New York

## Abstract

This paper is about an asymmetry in the justification of praising and blaming behavior which free will theorists should acknowledge even if they do not follow Wolf and Nelkin in holding that praise and blame have different control conditions. That is, even if praise and blame have the same control condition, we must have stronger reasons for believing that it is satisfied to treat someone as blameworthy than we require 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 a debate about whether someone acted freely is truth-conducive has to acknowledge that reasonable doubt is possible in both these cases.

## Introduction[1]

One of Susan Wolf's contributions to the free will literature is to highlight the intuition that we must meet a higher standard of justification to blame than to praise (Wolf 1980).[2] Wolf points out this justificatory asymmetry in the course of arguing for what might be called an ontological asymmetry in the control conditions of praise and blame: on her

[1] Thanks to Jeff Blustein, Gregg Caruso, Neil Levy, Daniel Miller, Jennifer Morton, Derk Pereboom, Nick Pappas, Saul Smilansky, Eric Steinhart, and Bruce Waller for helpful discussion of the issues I address in this paper.

[2] Wolf does not speak in exactly these terms, but I take this to be one of her points.

view, people can be blameworthy only if they had alternative possibilities, but can be praiseworthy even if they did not have alternative possibilities. There has been relatively little discussion of the justificatory asymmetry. Conversations have tended to focus on the claim that there is an ontological asymmetry (a claim which has been further defended by Dana Nelkin (2008 and 2009) and the justificatory asymmetry has largely been treated as a side-topic in those conversations.[3] But the claim that there is a justificatory asymmetry is largely independent of the claim that there is an ontological asymmetry, and it has important implications for free will theory even if we do not accept the claim that there is an ontological asymmetry. (In this paper, 'free will' means whatever satisfies the control condition of moral responsibility.[4]) For example, the claim that there is a justificatory asymmetry seems to imply that even if we accept a theory on which the control condition is the same in the contexts of praise and blame, we need better reasons to believe that the condition is satisfied to justify treating someone as blameworthy than we need to justify treating someone as praiseworthy. It also seems to imply that we need better reasons to believe that our theory of free will is true to use it to justify treating someone as blameworthy than to justify treating someone as praiseworthy. This would seem to hold whatever theory of the control condition we endorse, whether it is a metaphysically parsimonious compatibilism or an elaborate libertarianism. My goal in this paper is to develop this point, and then to draw some

---

[3] One exception is Watson (1996), but he considers different issues than I do here.

[4] 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 notions such as dual control, and weaker compatibilist notions such as guidance control or hierarchical control. For guidance control, see Fischer and Ravizza (1998). For hierarchical control, see Frankfurt (1971).

conclusions about when we can justify holding people morally responsible. The idea is to see how far this project can be carried without making independent commitments on the issues that typically occupy free will theorists, such as whether compatibilism is true, and whether free will requires alternative possibilities or agent causation.

This paper has four sections. In section 1, I discuss the intuition about the asymmetric justifiability of praise and blame in more detail. I then ask: how much higher is the standard for blame? In section 2 I offer a partial answer. Seriously harmful blaming behavior can only be justified if the claim *that the target of blame acted freely* cannot be reasonably doubted. On the other hand, harmless praising behavior can be justified as long as the claim *that the candidate for praise did not act freely* can be reasonably doubted. In section 3, I argue that anyone who thinks a debate about whether someone acted freely is truth-conducive must acknowledge that both these claims can be reasonably doubted. In section 4, I consider two potential objections and reply to them.

## 1. The Asymmetry Intuition

First, a remark on terminology: to be concise, I will often use 'blame' to refer to blaming behavior (actions used to express the Strawsonian negative reactive attitudes), and 'praise' to refer to praising behavior (actions used to express the positive reactive attitudes). In other words, as used here, 'blame' and 'praise' involve not just *believing* that someone is morally responsible, but acting in such a way as to *hold* him morally responsible. This should not be construed as implying that blame and praise are essentially matters of behaving in certain ways—it is merely an expository aid.

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 held to a higher standard than

arguments for refraining from harming them or benefiting them. All philosophers should acknowledge that this asymmetry exists, though disagreement is to be expected when it comes to giving a detailed explanation of why it exists. Everyone should agree that, in one sense or another, the primary orientation of morality is to give people reasons to benefit and avoid harming other people. If this is right, then arguments for harming people oppose the orientation of morality (in at least a *prima facie* way), while arguments for benefiting people align with it. In my own view, this asymmetry exists because of personhood-based desert claims which can be made by or on behalf of all people. In other words, it exists because of facts about how people deserve to be treated just because they are people. People deserve to be given the benefit of the doubt. They deserve not to be harmed unless there is a very strong justification for that harm. But if someone is inclined to benefit someone else, and the benefit can be conferred without harming anyone, the burden of justification is light: that is, if all parties to an interaction agree that something is a benefit which can be provided without harm to anyone, no further justification is typically expected.

Praise and blame are often kinds of benefit and harm. That is, praise and blame often produce valuable or harmful experiences for the people who are their recipients (and this effect is often intended by the people doing the praising or blaming). This is the case even when praise and blame do not involve the obvious benefit and harm of tangible reward and punishment. That is, praise obviously involves benefits when it includes a tangible reward, as when a rich rescuee lavishes not just words of gratitude but money upon a rescuer. But when a poor rescuee only offers words of gratitude, he may still benefit the rescuer, by causing the rescuer to have a valuable emotional experience. Something similar holds for blame: when blame includes retributive punishment by imprisonment or death, for example, it obviously includes harm, but even if it just involves the expression of a condemning attitude, this may cause the target of blame to experience emotional pain. In light of this, the

asymmetry in the justificatory standards for harm and benefit would appear to imply an asymmetry in the justificatory standards for harmful blame and beneficial praise.

It may be that not all praise is well-understood as a benefit, and that not all blame is well-understood as a harm. For example, there may be cases where we praise or blame people solely as a way of getting them to reflect on what they have done, without intending or causing them to have valuable or harmful experiences. I am not sure that we can sensibly call such behavior praise and blame, but for present purposes let us suppose that we can. The arguments of this paper would imply no asymmetry in the justificatory standards for such behavior. But it seems clear that a lot of praise is (and is intended to be) beneficial, and that a lot of blame is (and is intended to be) harmful. So the asymmetry at issue here is relevant for a good bit of the praise and blame that actually goes on.

Beneficial praise and harmful blame are special kinds of benefit and harm, since they can only be legitimate if the people who are their recipients deserve them based on how they have acted. That is, if people ever really deserve praise or blame, then they deserve it not just because they are people, but because of how they have acted, and people can only deserve things based on their actions if they are morally responsible for their actions. My own view is that we need to distinguish between personhood-based desert and action-based desert to explain this difference. (I think that all acts of harm and benefit must conform to personhood-based desert, but praise and blame are special in that they must also conform to another kind of desert, action-based desert.[5]) For purposes of this paper, however,

---

[5] Some philosophers think that all desert is action-based. Examples include Rachels (1978) and Sadurski (1985). Smilansky (1996) holds a related position—that giving up the belief that human beings are morally responsible for their actions implies giving up all our morally significant beliefs about desert. It can seem natural to suppose that actions are the only kind

recognizing this difference is more important than explaining it. Some may wonder whether the claims about moral responsibility involved in justifications of 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 at all on 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 for blame to a higher standard than the claims that play roles in justifications for praise. Since claims about moral responsibility play roles in justifications for 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 benefit/harm asymmetry, then it seems natural to suppose that the standard is a slope that rises as the harm at issue increases. But it is beyond my reach in this paper to

---

of desert base, since the category of action-based desert claims is very broad. Yet there are desert claims that can only be supposed to be action-based with great difficulty. It is natural to think that people deserve respect, access to their rights, equal treatment before the law, not to be used as a mere means to others' ends, and to be given the benefit of the doubt, and that they deserve these things not because of facts about her actions, but simply because they are people. I discuss this further in (Vilhauer 2009b and 2013).

survey this whole range. My argument here will focus on the standards for two kinds of limiting cases at the extremes of this range: seriously harmful blame, and praise that can be given without harming 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, or the sort of shaming behavior that marks its victims with an indelible stigma of monstrosity or absurdity, and such emotional violence can sometimes be even more harmful than bodily violence. Other kinds of seriously harmful blame include retributive punishment by execution, and retributively justified imprisonment under brutal conditions like those that prevail in contemporary prisons. To keep things concise, I will call seriously harmful blame 'serious blame' from now on. Probably only a small minority of instances of blame amount to serious blame, and my argument here only applies to these instances. This restricts the argument's applicability, but given the ethical importance of these instances, I do not think this does much to diminish its significance.
In talking about praise which can be given without harming anyone, I have in mind praise that benefits the recipient and doesn't cause any harm as a by-product. (I will call such praise 'harmless praise'.) Praise can cause harm as a by-product if the people we praise have managed to do good things, but have not tried very hard, and our praise causes them to be content with mediocre efforts.[6] Praise can also harm third parties. For example, 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

[6] Bruce Waller makes this point (Waller 2011: 136). Thanks also to Jennifer Morton for suggesting discussion of it.

painful experiences of the others. But harms like these can often be avoided. We can single out individuals for praise 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. And we can balance our praise with exhortations to keep trying harder.[7]

To be morally responsible, we must satisfy the conditions of moral responsibility. Most think there are multiple conditions of moral responsibility: a control condition, as well as conditions having to do with agents' knowledge and motivation. My focus here is on control, but it may be helpful to draw further distinctions between three kinds of control-related conditions:

(I) the control condition of moral responsibility;

(II) the condition of a justified *belief* that someone satisfies the control condition of moral responsibility;

(III) the condition of being justified in *treating someone* as satisfying the control condition of moral responsibility.

---

[7] Egalitarian praise may not be harmless 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 will not take a position on the legitimacy of this protest here. 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. But there may be a real problem about cheapening the practice of praising in some cases, and instances of praise which would cheapen the practice of praising would not be harmless. 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 to harmless praise.

It seems clear that asymmetry in (I) implies asymmetry in (II), and that asymmetry in (II) implies asymmetry in (III). As mentioned at the outset, Wolf and Nelkin argue for asymmetry in (I). On their view, alternative possibilities are necessary to satisfy the control condition of blameworthiness, but not praiseworthiness. Such an asymmetry in (I) would imply an asymmetry in (II). That is, if alternative possibilities are necessary to satisfy the control condition for blameworthiness but not praiseworthiness, then a justified belief that someone had alternative possibilities is necessary for a justified belief that he satisfies the control condition of blameworthiness but not praiseworthiness. This in turn would imply an asymmetry in (III). If a justified belief that someone had alternative possibilities is required for a justified belief that he satisfies the control condition of blameworthiness but not praiseworthiness, then we need a reason to believe that he had alternative possibilities to treat him as satisfying the control condition of blameworthiness, but not to treat him as satisfying the control condition of praiseworthiness.

I will not take a position on Wolf's and Nelkin's view that there is asymmetry in (I). Instead, I want to argue that there can be asymmetry in (III) without asymmetry in (I).
If one starts with some general notion that rational beliefs must conform to the way the world is, and rational actions must spring from rational beliefs, then it may initially seem natural to think that asymmetry in (II) must derive from asymmetry in (I), and asymmetry in (III) must derive from asymmetry in (II). This would imply that there could be no asymmetry in (III) without asymmetry in (I). It may not be absurd to suppose that we could have asymmetry in (II) without asymmetry in (I)—perhaps we have an obligation to hold our beliefs to varying epistemic standards depending upon their moral import, even when we are not acting on them, in such a way that we ought to hold our beliefs about whether people satisfy the control condition to a higher epistemic standard in the context of blame than in the context of praise, even if the control condition is the same in both cases. This view of belief can seem strained

in more familiar cases, however. It seems plausible to think that I should form and justify the belief *that it is going to rain* in the same way whether I am contemplating morally trivial decisions like whether or not to take an umbrella when I go to the market, or morally weighty decisions like whether I ought to contribute to the construction of a new well for an isolated community which once hid me from assassins and is now endangered by drought. So let us assume that there cannot be asymmetry in (II) without asymmetry in (I). Can there be asymmetry in (III) without asymmetry in (II)? It seems clear that this is possible in familiar cases. Consider again the claim that it is going to rain. The belief that this claim is possibly true is enough to make it reasonable to take my umbrella—it is not necessary for me to believe that this claim is true. But believing that this claim is possibly true is not enough to relieve me of an obligation to contribute to well construction. In this case, I must believe that the claim is true, and further, I must believe that it is true with a high degree of confidence. So the epistemic standards we must meet to incorporate the propositions which are candidates for belief into our practical reasoning can vary considerably depending on the sort of practical reasoning at issue. It would not be surprising to find an asymmetry in (III) for similar reasons even without an asymmetry in (I) or (II). I will explore this idea in the next section in more detail. I will argue that (III) is satisfied for serious blame only if the claim *that the target acted freely* cannot be reasonably doubted, but it is satisfied for harmless praise as long as the claim *that the candidate did not act freely* can be reasonably doubted.

## 2. Blame, Praise, and Reasonable Doubt

First I will argue that (III) is satisfied for serious blame only if the claim *that the target acted freely* cannot be reasonably doubted.[8] Call this the 'serious blame principle'. If

[8] The analogy to the criminal conviction standard I make here is drawn from a longer argument about retribution and reasonable doubt which I present in (Vilhauer 2009a).

(III) is not satisfied for serious blame, then serious blame is not justified, and we are obligated to refrain from it. So the serious blame principle implies that we are obligated to refrain from serious blame whenever it can be reasonably doubted that the target of the blame had free will with respect to the action at issue.

The serious blame principle is supported by an intuition which can be drawn out by considering the following scenario. Suppose that bad neuroscientists coercively implant a device into Skip's brain which can randomly cause him to do terrible things. It does not just give extra 'oomph' to intentions to act which Skip forms based on his native practical reasoning abilities (that is, the abilities which developed in him in the normal, species-typical way). It provides what I will call a 'complete system' of capacities for practical reasoning: identification of reasons for action (bad reasons, in this case), formation of desires, intentions, and volitions, and whatever else one might think necessary in a complete process of practical reasoning that terminates in action. The complete system provided by the implanted device is qualitatively different from Skip's native complete system—that is, it does not merely duplicate the sort of reasons-identification, intention-formation, etc. that Skip's native complete system is disposed to provide. If it operates, Skip's native system is totally bypassed—that is, the practical reasoning capacities provided by the device are the only ones which play a role in the explanation of the action. There is no way for anyone to know if it

---

I recapitulate it here in order to connect it to the more general point about asymmetrical justifiability which is the focus of the present paper. Derk Pereboom first proposes applying the 'reasonable doubt' standard in arguments about free will skepticism (Pereboom 2001: 161).

will operate, and if it does, there is no way for anyone to know that it has done so. The practical reasoning that the device can cause would be seamlessly integrated into Skip's conscious experience in such a way that even he would be unable to tell that it had not originated in him in the ordinary way. After the device is implanted, Skip commits a grisly murder. Good neuroscientists discover the device and promptly remove it, and everyone in the scenario can be sure that the device operated either only once, or not at all. All free will theorists should agree that if the device operates, then Skip does not act freely—that is, that he does not satisfy (I).

Would it be justified for Skip to receive serious blame for the murder? (In the terms used earlier, does Skip satisfy (III) in the context of serious blame?) For example, would we be justified in retributively inflicting intense suffering on him? In the scenario as described, presumably not. What if we could know that there was a 90% probability that the device did not operate? I think we would still hold that serious blame was not justified. What if we could know that there was a 95% or 99% probability that the device did not operate? It would be sensible to be very careful around Skip, and perhaps even to have him temporarily detained for observation, but I think we would still hold that serious blame would not be justified. This line of argument raises the bar until we arrive at the reasonable doubt standard. Only if the claim *that the device did not operate* could not be reasonably doubted could serious blame be justified.[9]

---

[9] My claim is that all free will theorists should recognize that this absence of reasonable doubt is necessary for serious blame to be justified, not that is sufficient. For example, incompatibilists would hold that it is also necessary that Skip does not inhabit a deterministic world.

Some may object that the intuitions drawn out by this scenario are not clear enough to offer unequivocal support to the serious blame principle. I am not sure that this is right, but the argument for the serious blame principle does not have to rest on this scenario alone. It gets some extra support from an analogy with another 'reasonable doubt' principle which is widely recognized to be a requirement of justice, that is, the principle observed 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 often part of a larger argument that that person is to be given a punishment which will cause serious harm such as death, or imprisonment under morally and emotionally corrosive conditions. This is why justice demands that we hold arguments that someone has committed a crime to the 'reasonable doubt' standard. By contrast, in civil trials, the most common sort of penalty involves payment of monetary damages, and in that context, the burden of proof is the lower 'preponderance of the evidence' standard, which is typically understood to require a demonstration that there is a greater than 50% probability that the defendant acted as accused. When a claim about free will serves as a premise in a justification for serious blame, it makes sense to hold it to the same standard as arguments in the criminal courtroom. 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. This suggests that the serious blame principle has the same justification as the criminal conviction standard.[10]

I will now turn to praise, and argue that (III) is satisfied for harmless praise so long as the claim *that the candidate for praise did not act freely* can be reasonably doubted. Call this the 'harmless praise principle'. The harmless praise principle does not imply an obligation in the way that the serious blame principle does. The harmless praise principle merely says that (III) is satisfied for harmless praise if it can be reasonably doubted that the candidate did not act freely. If (III) is satisfied for harmless praise, along with whatever other conditions there may be, then harmless praise is justified. But presumably the fact that acting in some way is justified does not imply that acting in that way is obligatory.

The harmless praise principle is supported by an intuition which can be drawn out by considering a different scenario. Suppose that confusedly public-spirited neuroscientists coercively install a device into Pip's brain which can randomly cause him to do heroic things. It is just like Skip's device, except the reasons-identification capacities it provides identify the right sort of reasons for action, rather than bad reasons. It provides a complete system, and if it operates, Pip's native complete system is totally bypassed. There is no way for anyone to know if it will operate, and if it does, there is no way for anyone to know that it has done so. The practical reasoning which the device can cause would be seamlessly integrated into Pip's conscious experience in such a way that even he would be unable to tell that it had not originated in him in the ordinary way. Pip goes on to rescue Doug from a hungry shark despite an awareness that he is risking his own life, motivated by a belief that Doug needs

---

[10] I am not claiming any sort of necessary connection between legality and morality here—I am merely appealing to the intuition (which I think many share) that the reasonable doubt standard in criminal law is in fact morally appropriate.

help and a desire to provide that help. Other neuroscientists promptly discover the device and remove it despite some consequentialist misgivings, and everyone in the scenario can be sure that the device operated no more than once.

All free will theorists should agree that if the device operated, then Pip did not satisfy (I), that is, the control condition for moral responsibility.[11] Since we cannot know whether the device operated, there is good reason to doubt that Pip satisfied the control condition of praise. But it nonetheless seems justifiable to give Pip harmless praise. Suppose that we could know that there was a 10% chance that the device did not operate. Would it be

---

[11] The claim that all free will theorists should agree that the agent did not satisfy (I) may require more argument in the Pip case than in the Skip case, in light of the view held by Wolf and Nelkin that (I) is asymmetrical, and that the requirements agents must meet to satisfy (I) are weaker in the context of praise than in the context of blame. In my view, all defensible theories of free will must endorse what we might call the 'minimal mechanism ownership' requirement. This requirement says that agents cannot satisfy (I) in any context unless their own capacities for practical reasoning play some role in the explanation of the action at issue. If an agent is manipulated through the coercive implantation of a complete system which (a) is qualitatively different from that agent's native complete system, (b) operates no more than once, and (c) can randomly and totally bypass the agent's native system without his consent, then the agent does not own the implanted system. As explained earlier, the total bypass means that the practical reasoning capacities provided by the device are the only ones which play a role in the explanation of the action. This is of course what happens in the Pip case if the device operates. Since the only practical reasoning capacities which play a role in the explanation of the action do not belong to Pip, Pip's own capacities play no role in the explanation, so Pip does not meet the minimal mechanism ownership requirement.

appropriate to give him harmless praise then? I think so. How about a 5% or 1% chance? Pip deserves the benefit of the doubt, and since harmless praise is the only thing at stake, it is not clear that we could have a reason to accept 10% as good enough, but not 5% or 1%. Even though the grounds for thinking that Pip acted freely eventually become quite weak as we proceed with this line of thought, they seem to remain strong enough, given the context.

This line of argument lowers the bar until we arrive at the reasonable doubt standard. No reasonable person could suppose that Pip could be justifiably praised if it could not be reasonably doubted that the device operated. The fact that praise has a control condition means that there is a conceptual connection here which is such that we cannot consistently regard our treatment of someone as praise if we know he did not satisfy the control condition.[12] But it seems reasonable and justifiable to praise him as long as it can be reasonably doubted that the device operated.[13] To put the intuition drawn out by this scenario

---

[12] We could consistently give Pip praise-like beneficial treatment despite knowing that the device operated if we were, for example, pretending to praise Pip, but we could not consistently regard this as praise. There might be consequentialist reasons in favor of pretending to praise Pip—for example, we might cause him to have a valuable emotional experience if we told him he ought to be proud of his courageous willing. But it is worth noting that if Pip too knew that the device operated, it is unlikely that we would succeed in causing him to have a valuable emotional experience by telling him this, because he too would know that the courageous willing was not his. Even if Pip knew that the odds of the device not having operated were very low, but knew there was room for reasonable doubt, he could consistently hope that the courageous willing was his own, and have a valuable emotional experience of praise on the basis of that hope.

[13] Incompatibilists will object that justified praise requires not just reasonable doubt that the

in different words, we could say that it seems justifiable to praise him as long as the claim *that he did not act freely* can be reasonably doubted. If this is right, then the harmless praise principle is correct.

The harmless praise principle may seem to set a peculiarly low standard. But remember that harmless praise is a kind of harmless benefit. When all parties agree that an action under consideration is a harmless benefit, we usually expect no further justification. Harmless praise cannot be supposed to be entirely typical in this regard, of course. Harmless benefits as such are justified merely by the claim that they are harmless benefits. But harmless 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 we ask how high a justificatory standard to apply to the claim that the candidate acted freely, it makes sense to look at the moral context in which the question is asked. If the primary orientation of morality is to give people reasons to benefit and avoid harming other people, then the harmless praise principle aligns with the primary orientation of morality. I think this makes it clear that the justificatory standard should be low. But even if it is clear that the standard should be low, it may not be clear precisely how low it should be. The reasonable doubt standard built into the harmless praise principle is the lowest

---

device operated, but also the falsity of determinism. But if the argument I go on to make in section 3 is correct, then incompatibilists who think it is truth-conducive to debate whether a deterministic Pip acted freely must accept that it is reasonable to doubt that he did not act freely, and if the harmless praise principle is correct, this suffices to justify harmless praise.

possible standard for practical reasoning.[14] Why should we accept the lowest possible standard instead of a standard which is higher, but still low? In the context of harmless praise, it does no harm to use the lowest possible standard, and it may do some harm to use a higher standard. That is, using a higher standard may deprive some candidates for harmless praise of benefits they might otherwise have.

Some will no doubt reply that if the correct standard is higher than reasonable doubt, then there would indeed be some harm in accepting the harmless praise principle, that is, the harm of accepting a false belief. But the most obvious way in which a false belief can be harmful is by prompting people to act in harmful ways, and mistakenly accepting the harmless praise principle could not have this effect. It may be that we can also be harmed by accepting a false belief even if it has no impact on our actions. The sheer fact of not seeing things as they are may be harmful. But surely the harm of this kind which could arise from mistakenly accepting the harmless praise principle is quite slight. Some philosophical errors might involve significant harm of this kind, for example, believing that God exists if this is false. But the point at issue is small and peripheral by comparison. So it seems fair to assume that the harm of depriving some candidates for praise of benefits they might otherwise have is greater that the harm of mistakenly accepting the harmless praise principle.

There are a couple of points which it may be helpful to reiterate at this juncture. First, in claiming that harmless praise of Pip is justified, I am not claiming that anyone is obligated to give him harmless praise. As far as my argument here is concerned, we may have the right

---

[14] It may be objected that a standard involving a lack of certainty that the agent did not act freely would be lower, but I don't think it is clear what it means to apply standards involving certainty, which are at home in the context of arguments about mathematics and logic, in the messy context of practical reasoning.

to adopt a policy of praising only those who have undeniably satisfied the conditions of moral responsibility in doing truly great deeds. We may even have the right to model ourselves on the 'proud' or 'great-souled'[15] man of the *Nicomachean Ethics* 4.3, who seems to offer praise rarely, if at all. All that I am arguing here is that when we properly understand harmless praise as a kind of harmless benefit, we will see that anyone who is inclined to praise Pip can legitimately do so, as long as it can be reasonably doubted that he did not act freely.

In the next section, I will argue that no matter what theory of free will we hold, if we think that a debate about whether someone acted freely is truth-conducive, we must accept that it can be reasonably doubted both that he did act freely, and also that he did not. If this is right, and if the serious blame and harmless praise principles are correct, then it follows that anyone who thinks a debate about whether someone acted freely is truth-conducive must accept that he does not satisfy (III) for serious blame, and that he does satisfy it for harmless praise.

## 3. Reasonable Doubt About Free Will

In this section, I will argue that no matter what theory of free will one holds, if one thinks that a debate about whether someone acted freely is truth-conducive, one must accept that it can be reasonably doubted both that he did act freely, and also that he did not.[16]

Let me begin with a few terms for describing debates. Some debates can be represented as focused on a central claim. A debate about whether someone acted freely is an example. Its central claim is that this person (i.e. a person of such-and-such a description)

---

[15] Thanks to Nick Pappas for translation help on this point.

[16] This section sets out a more general version of an argument I advance in (Vilhauer 2009a).

acted freely in this situation (i.e. a situation of such-and-such a description). Debates which are focused on a central claim include a pro side and a con side. The pro and the con side each have a basic position. The basic position of the pro side is that the central claim is true. The basic position of the con side is that the central claim is false.

To believe a debate to be truth-conducive is to believe that working on at least some of the objections to one's basic position posed by the opposite side either tends to prompt one to revise the theory one uses to support one's basic position in a way that makes it more likely to be true, or helps one to better understand why one's existing theory is true. One can be on the pro or con side of a debate without believing it to be truth-conducive. Someone might hold that the central claim is true, and someone else might hold that it is false, but they might agree that debate about it will not prompt revisions that make either of their theories more likely to be true, or help either to see better why their existing theories are true.

Now that the terms have been explained, the argument can proceed quickly. Objections can only be truth-conducive if they are reasonable, and reasonable objections are grounds for reasonable doubts about the basic positions to which they are objections. This means that, whether one is on the pro or con side of a debate, if one takes the debate to be truth-conducive, one must accept that it is possible to reasonably doubt one's basic position. So, in the case of a debate about whether someone acted freely, those on the pro side must accept that it is possible to reasonably doubt the claim that he acted freely, and those on the con side must accept that it is possible to reasonably doubt the claim that he did not act freely.

Two points of clarification may be helpful before the implications of this argument for free will are discussed in more detail. First, I am not claiming that we can only get closer to a true theory, or gain a better understanding of why our existing theory is true, when we are working on reasonable objections. One can of course be struck by a good idea at any

time, even when one is working on unreasonable objections. The claim I intend is the weaker claim that, since unreasonable objections do not direct our attention to features of our theory which are reasonably thought of as implausible, there is nothing about unreasonable objections *as such* which would justify us in believing that working on them would be truth-conducive. We might suppose that they could haphazardly cause us to get closer to a true theory, or to a better understanding of why an existing theory is true, but not that they would *lead* us to, or give us *reasons* for, these outcomes.

Second, I am not claiming that everyone who thinks a debate is truth-conducive must actually be able to doubt her own basic position in the debate. It is common for people to become so deeply committed to their basic positions that it becomes psychologically impossible for them to doubt them. But this is no objection to this argument, because the fact that it is psychologically impossible for some people to doubt a claim does not imply that it cannot reasonably be doubted. I do not even claim that *if* the debaters were reasonable, *then* they would doubt their basic positions. Philosophical disagreement is a complicated matter. As far as this argument is concerned, there may be what Richard Feldman calls 'mutually recognized reasonable disagreement' between the pro and con sides (Feldman 2006). That is, it may be that even when both sides recognize that the other side can reasonably doubt their basic position, they can remain reasonable without doubting their own basic position. To make my argument, it is enough to claim that those on both sides would not be unreasonable if they came to doubt their own basic positions, and it seems fair to claim this much.

Philosophers will disagree about when we can have truth-conducive debates about whether someone acted freely. The cases most congenial to free will are ones with normal adults acting in normal conditions—call such cases 'normal cases'. There may be extreme Strawsonians who think the belief that people often act freely in normal cases is so fundamental to the meaning of discussions about free will that those discussions lose all sense

if we put it in question. Such Strawsonians might therefore hold that debating about whether people in normal cases ever act freely is not truth-conducive. But this would presumably be a minority view. Others may be committed to varieties of reductive physicalism which make the idea of free will look too absurd for a debate about it to be truth-conducive even in normal cases. But this too would be a minority view.

Anyone who thinks it is truth-conducive to debate whether people in normal cases act freely must accept that it can be reasonably doubted that they do. If normal cases are the situations in which free will is most likely to be found, then it seems to follow that if it can be reasonably doubted that people act freely in normal cases, then it can be reasonably doubted that anyone ever acts freely. If the serious blame principle is correct, this implies that anyone who believes a debate about normal cases to be truth-conducive must accept that (III) is never satisfied for serious blame, and therefore that serious blame is never justified.

Anyone who thinks it is truth-conducive to debate whether people act freely in normal cases must also accept that it can be reasonably doubted that they do *not* act freely. If the harmless praise principle is correct, this implies that anyone who thinks this debate is truth-conducive must accept that people in normal cases satisfy (III) in the context of harmless praise.

Non-normal cases do not provide ideal conditions for finding free will, but even conditions that are not ideal may provide room for a truth-conducive debate. There are lots of cases where there may not be room for a truth-conducive debate about free will, for example, cases where we know the agent did something by accident or because of an irresistible compulsion, and cases involving newborn babies or people with very profound cognitive or behavioral disabilities. But there is probably room for truth-conducive debate in cases involving people who do things because of seemingly resistible compulsions, and cases involving children or people with less profound cognitive or behavioral disabilities. If we

think debates are truth-conducive in these cases, then we must accept that (III) is satisfied for harmless praise in these cases as well.

## 4. Objections and Replies

In this section I will consider two potential objections to the arguments presented above. The first is as follows. The argument of the previous section implies that anytime we think it is truth-conducive to debate a proposition, we must acknowledge that it can be reasonably doubted, and some might object to this. For example, it implies that anyone who thinks the debate about other minds skepticism is truth-conducive must accept that it is possible to reasonably doubt that other minds exist. Does it follow that anyone who thinks the debate about other minds skepticism is truth-conducive should act as if the existence of other minds has been put in doubt? This would only follow if we accepted a principle which said that the possibility of reasonable doubt about the existence of other minds requires us to act as if other minds do not exist. I cannot see any reason to accept such a principle. This highlights an important difference between doubts about free will and doubts about other minds. If the argument of this paper is right, morality requires us to treat reasonable doubt about free will as practically significant. But it is hard to see how morality could require us to treat reasonable doubt about the existence of other minds as practically significant. If relations to other persons are essential to morality, then morality would seem to imply a commitment to the existence of other minds.[17]

The second objection I want to discuss is based on a concern about whether (I) and (III) are really as independent as I have argued. The sort of independence I have claimed to

[17] I discuss this issue in more detail in (Vilhauer 2012).

find may seem to presuppose a metaphysical view of moral responsibility according to which the properties agents must possess to satisfy the control condition must be characterizable independent of our practices of holding agents morally responsible. Presupposing this amounts to rejecting, without argument, the influential views of P.F. Strawson and R. Jay Wallace, who (roughly speaking) define moral responsibility as appropriate accessibility to the reactive attitudes and the practices which express them. (Let us call views like this 'deflationary theories'.) But I do not take myself to be presupposing this. I think that any defensible deflationary theory must allow for cases where agents are in a deep sense appropriately accessible to reactive attitudes and practices, but nobody is justified in treating them as appropriately accessible, because nobody has strong enough evidence. It must also allow for cases where agents are in a deep sense *not* appropriately accessible, but everyone is justified in treating them as appropriately accessible, because everybody has strong enough evidence. Not to allow for such cases would be to slide into what I take to be an unacceptable subjectivism about moral responsibility. Consider a modified version of the Skip scenario. Suppose that Skip has all the abilities that deflationary theorists require for appropriate accessibility to the reactive attitudes and practices, and that he commits a murder without his implanted device operating. Suppose everybody in the scenario knows that that there was a 50% probability that it operated, but suppose that as a matter of fact it does not operate, though nobody in the scenario knows this. I think deflationary theorists ought to say that Skip satisfies the control condition for being appropriately accessible to the reactive attitudes and practices, but that nobody in the scenario is justified in treating him as appropriately accessible. In other words, deflationary theorists ought to acknowledge that in this scenario, (I) is satisfied, but (III) is not. We can modify the Pip scenario in a similar way. Suppose that everybody in the scenario knows that Pip's device has a 50% chance of operating, and that as a matter of fact it *does* operate, though nobody in the scenario knows

this. In this case, I think deflationary theorists ought to hold that Pip does not satisfy the control condition for being appropriately accessible, but that everybody in the scenario is justified in treating him as appropriately accessible. In other words, (I) is not satisfied, but (III) is.

## Conclusion

To conclude, I would like to mention two points which are implicit in what I have argued already, but which may be worth making explicit. First, while I have not argued that there is no ontological asymmetry in the control condition, the points discussed here may provide material for such an argument. If the intuition that there is an asymmetry in praise and blame can be explained by general ethical considerations about asymmetry in the justification of benefit and harm, then there is no need to posit an ontological asymmetry to explain the intuition. Second, it seems that both free will believers and free will deniers can accept the argument presented here without giving up their basic positions. By 'free will deniers', I mean those who hold that the control condition of moral responsibility is never satisfied, and by 'free will believers', I mean those who hold that it is sometimes satisfied. It seems consistent to hold both that some claim is true and also that it can be reasonably doubted. If this is right, then free will deniers could accept that it can be reasonably doubted that candidates for harmless praise do not act freely. If they accept the harmless praise principle, they should conclude that harmless praise is justifiable (at least in cases where there is no reason to doubt that any conditions for moral responsibility which do not involve control are satisfied). Similarly, free will believers could accept that it can be reasonably doubted that targets of serious blame act freely, and if they accept the serious blame principle, then they should conclude that serious blame is never justified. This would allow deniers and believers to preserve their basic positions while meeting each other halfway on some important issues.

*City College of New York, USA*

## References

Feldman, R. (2006) 'Epistemological Puzzles about Disagreement', in S. Hetherington (ed.) *Epistemology Futures,* 216-36. Oxford: OUP.

Fischer, J.M. and M. Ravizza (1998) *Responsibility and Control: a Theory of Moral Responsibility.* Cambridge: Cambridge University Press.

Frankfurt, H. (1971) 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 68/1: 5-20.

Nelkin, D. (2009) 'Responsibility, Rational Abilities, and Two Kinds of Fairness Arguments', *Philosophical Explorations* 12/2: 151-65.

----- (2008) 'Responsibility and Rational Abilities: Defending an Asymmetrical View', *Pacific Philosophical Quarterly* 89/4: 497-515.

Pereboom, D. (2001) *Living Without Free Will.* Cambridge: Cambridge University Press.

Rachels, J. (1978) 'What People Deserve', in J. Arthur and W.H. Shaw (eds.) *Justice and Economic Distribution,*150-63. Englewood Cliffs: Prentice-Hall.

Sadurski, W. (1985) *Giving Desert Its Due: Social Justice and Legal Theory*. Dordrecht: D. Reidel.

Smilansky, S. (1996) 'Responsibility and Desert: Defending the Connection', *Mind* 105/417: 157-63.

Strawson, P.F. (1962) 'Freedom and Resentment', *Proceedings of the British Academy,* 48: 1–25.

Vilhauer, B. (2009a) 'Free Will and Reasonable Doubt', *American Philosophical Quarterly* 46/2: 131-140.

----- (2009b) 'Free Will Skepticism and Personhood as a Desert Base', *Canadian Journal of Philosophy* 39/3: 489-511.

----- (2012) 'Taking Free Will Skepticism Seriously', *Philosophical Quarterly* 62/249: 833–852.

----- (2013) "Persons, Punishment, and Free Will Skepticism" (*Philosophical Studies,* January 2013, Vol. 162, No. 2, pp. 143-163).

Waller, B. (2011) *Against Moral Responsibility.* Cambridge: MIT Press.

Wallace, R. J. (1994) *Responsibility and the Moral Sentiments.* Cambridge: Harvard University Press.

Watson, G. (1996) 'Two Faces of Responsibility', *Philosophical Topics* 24/2: 227–248.

Wolf, S. (1980) 'Asymmetrical Freedom', *Journal of Philosophy* 77/3: 151-166.