Taking Free Will Skepticism Seriously

Benjamin Vilhauer, William Paterson University of New Jersey

Abstract

An apparently increasing number of philosophers take free will skepticism to pose a serious challenge to some of our practices. This must seem odd to many—why should anyone think that free will skepticism is relevant for our practices, when nobody seems to think that other canonical forms of philosophical skepticism (for example, skepticism about induction or other minds) are relevant for our practices? Part of the explanation may be epistemic, but here I focus on a metaethical explanation. Free will skepticism is special because it is compatible with 'basic moral reasons'—moral reasons acknowledged by all mainstream ethicists—and other minds and induction skepticism are not. One example is our reason not to intentionally harm others. Practical seriousness about other minds and induction skepticism undermines this reason, but practical seriousness about free will skepticism only undermines a potential overrider of this reason, that is, the reason of retribution.

1. Introduction[1]

Free will is the kind of control over our actions necessary for moral responsibility, and we must be morally responsible for our actions to deserve particular kinds of treatment based on them, such as praise, blame, reward, or punishment. Until fairly recently, the most energetic debate in the free will literature was about whether free will was compatible with

---

[1] Thanks to Neil Levy, Richard Burnor, the audience at the 2011 Felician Ethics Conference, and the referees at *Philosophical Quarterly*, for their helpful comments.

determinism. Just about everyone involved agreed that, one way or another, human beings did in fact have free will. What they disagreed about was whether determinism could obtain along with free will. But recent decades have seen influential contributions by free will skeptics such as Derk Pereboom, Saul Smilansky, and Galen Strawson.[2] These philosophers differ in many ways, but they all endorse the view that I will refer to here as 'free will skepticism', that is, the view that we do not know that we have free will. Some (e.g. Strawson and Smilansky) even endorse a stronger view—that we can know we lack free will. There are of course precedents for these views in the history of philosophy, but they are being argued with a renewed energy, and their implications for our practices (such as retributive punishment and distributive justice) are being studied in a new level of detail. A small industry has grown up around the question of whether we ought to live differently in light of free will skepticism. In this paper, I will call the view of philosophers who take this question seriously 'practical seriousness' about free will skepticism. To be practically serious about a kind of philosophical skepticism is to accept that it is a live philosophical possibility that arguments for that kind of skepticism are relevant for practical reasoning. (By 'relevant for practical reasoning', I mean relevant for reasoning about at least some actual behavior in the actual world, not merely relevant for how people act in merely possible worlds.)

---

[2] D. Pereboom, *Living Without Free Will* (Cambridge University Press, 2001); S. Smilansky, *Free Will and Illusion* (Oxford University Press, 2000); G. Strawson, *Freedom and Belief* (Oxford University Press, 1991). Smilansky holds that even though we lack free will, we must in some sense maintain a compatibilist outlook in order to avoid an unacceptable impoverishment of ethics.

Many philosophers, some within the free will subfield, and probably many more without, must find practical seriousness about free will skepticism perplexing. Philosophers have long acknowledged that there are all kinds of fascinating and profound reasons for skepticism about many of our ordinary beliefs, for example, our beliefs that we can know that other minds exist, and that induction yields knowledge. But philosophers have almost universally not thought these things to be worth dwelling on outside the study. Experts on other minds or induction skepticism have never, to my knowledge, seriously suggested that these kinds of skepticism are relevant for practical reasoning. If nobody is practically serious about these other forms of skepticism, then shouldn't we worry that people who are practically serious about free will skepticism are taking it too seriously?

I think they are not taking it too seriously, because free will skepticism has a moral status which induction and other minds skepticism lack. In this paper, I will not argue that free will skepticism is true, or even that we ought to be practically serious about it. I will argue for more limited claims. But let me define some terms before I set these claims out. A moral reason is *basic* if all mainstream ethicists accept that we have it. The central example, for purposes of this paper, is an extremely strong reason not to intentionally harm other people—a reason which I will refer to as 'H', to keep things concise. To be *committed to maintaining* a reason is to have resolved to reject theory changes that 'downgrade' that reason (for example, changes that increase the incidence of its being overridden by other reasons). To be *inclined to advance* a reason is to be disposed to accept (all other things being equal) theory changes that 'upgrade' it (for example, changes that diminish the incidence of its being overridden by other reasons). I will argue that if we are committed to maintaining H, then we can be practically serious about free will skepticism, but not other minds or induction

3

skepticism.[3] I will go on to argue that if we are inclined to advance H, then we have good reason to become practically serious about free will skepticism.

Before I begin, let me address some prefatory issues. I don't have in mind principles that I take to provide necessary and sufficient conditions for determining whether a kind of skepticism is a kind of philosophical skepticism. By 'kinds of philosophical skepticism' I mean only to refer to kinds which have found places in the canon of philosophy. I take the kinds of skepticism mentioned thus far—induction skepticism, other minds skepticism, and free will skepticism—to be examples. Induction skepticism is the view that induction does not yield knowledge, or, phrased differently, that we don't know the products of induction. Other minds skepticism is the view that we don't know that other minds exist. Free will skepticism is the view that we don't know that we have free will.

Some may think that the explanation of practical seriousness about free will skepticism lies in a special epistemic status, rather than a special moral status. They may point out that some practically serious free will skeptics endorse not only the view I am calling free will skepticism in this paper—the view that we don't know we have free will— but also the stronger view mentioned earlier—the view that we can know we lack free will. Things are different with respect to the other kinds of skepticism at issue. To the best of my knowledge, no philosophers who endorse the weaker claim that we don't know other minds

---

[3] Also see T. Sommers, 'The Objective Attitude', *The Philosophical Quarterly*, 57, 228 (2007), pp. 321-341. Sommers also contrasts free will skepticism and other minds skepticism, but his argument focuses on the psychological possibility of believing that we lack free will versus believing that there are no other minds, while my argument focuses on these beliefs' differing implications for practical reasoning.

exist also endorse the stronger claim that we know they don't exist, and no philosophers who endorse the weaker claim that we don't know inductively-generated claims to be true endorse the stronger claim that we know these claims to be false. It makes sense to worry more about a claim presupposed by our practices if we take ourselves to know that it is false than if we merely take ourselves to lack knowledge of its truth.

This difference no doubt explains part of the phenomenon at issue. If it explains all of it, then there is no point in appealing to the idea that free will skepticism has a special moral status. But I don't think it explains all of it. I don't want to rule out the possibility that it is a complete explanation for some philosophers' actual practical seriousness about free will skepticism. But I don't think it can be the whole story, because I don't think one has to hold that we know we lack free will in order to have good reason to be practically serious about free will skepticism but not these other kinds of skepticism.

Objectors may continue in this vein. They may point out that even those who don't hold that we can know we lack free will can still hold that we can be more confident about the truth of the claim that we lack free will than we can be about the truth of the claims that induction yields only falsehoods or that other minds do not exist. Some might think that the scientific view of the world provides grounds for this view. All science is evidence in favor of induction, in the sense that induction is the fabric of scientific evidence. We also have some scientific evidence in favor of the existence of other minds, at least in the limited (but probably non-trivial) sense that a scientific study of the contents of my own skull would presumably show them to be of the same general kinds as the contents of other skulls. But we have a growing body of scientific evidence against the claim that we have free will, in the sense that our gradually improving scientific understanding of the causes of our actions thus far gives no role to anything resembling the agent causation which many traditional

explanations of free will require. Some might take these points to give us good enough grounds to regard free will skepticism with special seriousness.

I will confess that I am not entirely unmoved by this sort of argument, but it is obviously controversial, and one must already lean toward incompatibilism and a fairly reductive variety of physicalism to find it very attractive. Most free will theorists, including compatibilists, libertartians, and many free will skeptics, reject the notion that we need to find a role for traditional notions of agent causation in scientific explanations of action in order to make free will compatible with the scientific image of the world. In this paper, I wish to present an account of practical seriousness about free will skepticism which will be of interest to everyone committed to maintaining or advancing H, not just those who are already convinced that we have good grounds for doubting that we have free will. So, for purposes of this paper, I will not assume that free will skepticism has an epistemic status which induction and other minds skepticism lack.

2. Skepticism and Commitment to Basic Moral Reasons

As mentioned earlier, a moral reason is basic just in case all mainstream ethicists agree that we have it. I won't be too precise about the boundaries of the mainstream, since I take the claims I'm making about it to be uncontroversial, but the mainstream, in the sense I intend, includes typical varieties of consequentialism, deontology, and virtue ethics, and no doubt lots of other views too. Let me emphasize that I am not claiming that we ought to agree that we have basic moral reasons, or that we ought reflexively to defer to the views of mainstream ethicists. My main interest in using the notion of basic moral reasons is simply to identify a category of reasons that happens to be important to a great many philosophers. I readily concede that there are sensible philosophers who would not accept that we have the moral reasons I am calling basic—for example, philosophers who hold that there can be no

6

such things as moral reasons, because they hold that there are no moral facts which have the right sort of objectivity to ground reasons. But anyone who does think that there are moral reasons should probably acknowledge that there are worse ways to proceed in lines of inquiry which bear on metaethics than by paying special attention to points of agreement among mainstream ethicists.

The central example of a basic moral reason to be used in this paper is H (an extremely strong reason not to intentionally harm other people). There are no doubt other basic moral reasons, for example, reasons to benefit others, as well as reasons to benefit and avoid harming ourselves, but these will not play central roles here. We can accept that H exists without holding that it is never justified to intentionally harm another person—we need only hold that a reason even stronger than H is needed to justify harming someone.

Philosophers who endorse different mainstream ethical theories explain H differently, and have different views about the conditions under which it is overridden (which I will refer to as H's 'overriding conditions'). But they agree that we have this reason. Kantian deontologists explain it as an imperative to which we must conform unless in harming someone (e.g. by punishing him) we do not use him as a mere means. Virtue ethicists might explain it as a principle followed by those who possess the virtue of justice except in cases where justice permits or requires harm. Act-utilitarians explain H as derived from our reason to maximize overall happiness, and as overridden in some cases by that same reason.

We can acknowledge these differences while at the same time recognizing a shared commitment to H. That is, while these explanations of H and accounts of its overriding conditions are significantly different, I see no reason to think that the different frameworks which surround this reason in these different theories would give us reason to reject the claim that these different approaches to ethics do in fact agree on an extremely strong reason to avoid harming others. It is important to note that the different views held by mainstream

ethicists about the conditions under which H can be overridden mean that none of the overriding conditions just mentioned are basic in the way that H is. (It might turn out that there are other overriding conditions which are basic in the way that H is, for example, reasons of self-defense, but I will not investigate that here.)

Some act-utilitarians may doubt the claim that there a strong reason not to intentionally harm others.[4] They may suppose that since the only criterion for putting the act-utilitarian seal of approval on any action is that it maximizes overall happiness, our reasons for harming people when it is a means to maximizing overall happiness are just as strong as our reasons for benefiting people when it is a means to maximizing overall happiness. But this objection does not acknowledge the constraints under which we reason practically. Act-utilitarians need to be more confident about their ability to predict the consequences of actions to justify harming people than they do to justify benefiting people. An action which does immediate harm can only improve overall utility through its consequences, while an action which immediately benefits people has an immediately positive impact on overall utility. So when predictability is very limited, as it is under typical conditions for practical reasoning, there is a very strong reason to avoid intentionally harming others. Therefore act-utilitarians must acknowledge H along with other mainstream ethicists.

As mentioned earlier, to be committed to a reason is to have resolved to reject the 'downgrading' of it. One accepts the downgrading of a reason when one accepts a theory change that (e.g.) undermines an existential commitment necessary to make the reason relevant to practical reasoning, weakens it, or increases the incidence of conditions under which it is overridden. The main idea I want to explore in this paper is that practical seriousness about different kinds of skepticism has different implications for H and for

---

[4] Thanks to Neil Levy for pointing out this objection.

overriding conditions for H. Practical seriousness about induction or other minds skepticism downgrades H, but practical seriousness about free will skepticism does not. Instead, it downgrades the reason of retribution, which is a potential overrider of H.

Ethicists who endorse retributivism (retributivists) hold that people can deserve to be harmed because of how they have acted. Most retributivists are deontologists or virtue ethicists (though deontologists and virtue ethicists do not have to be retributivists). For example, Kant holds that we can be harmed without being used as mere means if we deserve to be harmed, and many virtue ethicists hold that the just permit or require people to be harmed when they deserve to be harmed. But no one can deserve to be harmed because of how he acted unless he acted with free will. If we become practically serious about free will skepticism, then we have to accept that free will skepticism might be relevant to retributive reasoning.

Utilitarians are never retributivists. So it may seem that, in advocating practical seriousness about free will skepticism, despite its weakening of retributivism, I am inevitably advocating utilitarianism over deontology or virtue ethics. But that is not my goal here. First, as just mentioned, retributivism is not an essential feature of either deontology or virtue ethics (as I'll explain in more detail later). Second, insofar as I am advocating a positive view of ethics in this paper, it is that when we are in doubt about how to proceed, it makes sense to defer to basic reasons. This view, on its own, does not advantage any variety of mainstream ethics over any other. What it does is to elevate the importance of points of agreement among ethicists over the points of disagreement. Since there is little agreement on overriding conditions for H, this view of ethics sees those overriding conditions as legitimately vulnerable in the process of theory change in a way that H itself is not.

To develop these ideas in more detail, it will be helpful to consider the attitude of practical seriousness in a more fine-grained way. So far I have described practical

seriousness as accepting that arguments for a particular kind of skepticism might be strong enough to be relevant for practical reasoning. But there are various degrees of relevance to practical reasoning which we might accept to be possible. We might become practically serious about a form of skepticism but remain relatively conservative about it, believing that our skepticism could require only the smallest possible changes in arguments about how we should act. I will call this 'best-case practical seriousness'. Alternatively, we might embrace a more radical agenda, and hold that skepticism about some claim C might be so serious that we ought never to assert C, in any way at all, in arguments about how we should act. I will call this view 'worst-case practical seriousness'. I take worst-case and best-case practical seriousness to be the boundaries of a continuum of attitudes, but I will mostly focus on boundary cases because they are clearer-cut.

3. Worst-Case Practical Seriousness

Worst-case practical seriousness means accepting that arguments for skepticism about some claim C might be strong enough to make it the case that we ought never to appeal to C, in any way at all, in arguments about how we should act. Worst-case practical seriousness about other minds or induction skepticism is not compatible with a commitment to maintaining H, but worst-case practical seriousness about free will skepticism is.

Suppose Yelena accepts H, and has always had an unskeptical view of other minds. But suppose she now becomes practically serious about worst-case other minds skepticism. Yelena now has to accept that she might never be entitled to assert the existence of other people, since it seems a safe bet that she cannot assert the existence of other people without also asserting the existence of other minds. This diminishes the confidence with which she can make the existential commitment necessary for H to be relevant to her practical reasoning. Yelena may legitimately remain just as confident about the truth of the

conditional claim that *if* other people exist, *then* H is relevant to her practical reasoning. But since she holds that she may never be entitled to assert the antecedent, she must acknowledge that H may never be relevant to her practical reasoning. So worst-case practical seriousness about other minds skepticism downgrades H.

Next, suppose Ludwig accepts H, but now becomes practically serious about worst-case induction skepticism. He must now accept that it might never be legitimate to make claims involving the products of induction in practical reasoning. This requires him to accept that it might not be legitimate to reason about means to ends in cases where claims about means include inductively-generated claims. This would downgrade H, since many of the ways in which H can be relevant for practical reasoning depend on inductively-generated claims. Before Ludwig's worries about induction skepticism, H required him not to wave pokers at interlocutors in order to emphasize points in his arguments, but now that he must question the induction-based claim that waving pokers at people is likely to cause them harm, he must question the relevance of H for this practice. So worst-case practical seriousness about induction skepticism downgrades H.

Things are different when we turn to worst-case practical seriousness about free will skepticism. It clearly downgrades the retributivist reason for overriding H, since it undermines an existential commitment required by the retributivist reason. That is, the retributivist reason implies the existence of free will, and practical seriousness about free will skepticism means accepting that we might never be able to legitimately assert the existence of free will. But it does not downgrade H.

It may be objected that downgrading the retributivist reason downgrades H indirectly. We might think this if we think H is more relevant for practical reasoning insofar as we believe that we can deserve to be harmed for failing to act in accordance with in. But this confuses the moral reason H with a prudential reason to avoid harm. It is not that H is any

stronger when backed by the threat of harm, but rather that a prudential reason constrains behavior along with H. However we are to understand moral motivation, we must acknowledge that people are sometimes motivated to act morally even in cases where they can be certain to avoid harm if they act immorally.

Some may think that even if refraining from making claims about free will does not indirectly downgrade H, it undermines too many moral theories that depend on claims about free will, and this constitutes an unacceptable impoverishment of the possibilities open to moral theory. But this is a hard case to make. Utilitarians need not appeal to free will, and there are also varieties of deontology and virtue ethics that do not require appeals to free will.

It is probably obvious that utilitarians need not appeal to claims about free will. Utilitarians are concerned with treating people in utility-maximizing ways, not with responding to claims about how people deserve to be treated in light of their actions. (Rule-utilitarians might come up with rules which involved 'treating people as they deserved to be treated based on their actions', but these rules would have to be grounded in utility, not desert, so it would not matter in practice if such utilitarians acknowledged in theory that action-based desert was merely a useful fiction provided by, say, the peculiarities of human psychology.) One common way of explaining the significance of action-based desert is that it allows us to make backward-looking evaluations of how people should be treated. But utilitarians are only interested in forward-looking evaluations.

The complete dispensability of claims about free will for utilitarians also implies that worst-case free will skepticism is compatible with *all* basic moral reasons. A reason cannot be basic unless it is acknowledged by utilitarians, since they must be counted among mainstream moral theorists. So, since utilitarians do not acknowledge any moral reasons which include or entail claims about free will, there can be no basic moral reasons which include or entail claims about free will.

In light of this point, it may seem that utilitarians *could not* become practically serious about free will skepticism (whether worst-case or best-case), since claims about free will are not relevant to the practical reasoning they endorse. But we can be practically serious about a kind of skepticism so long as we take it to be relevant to practical reasoning that goes on in the actual world, even if it is not our own practical reasoning. Since claims about free will do play a role in the practical reasoning endorsed by many ethical theorists that compete with utilitarians, utilitarians can become practically serious about free will skepticism, and becoming so gives them another way to argue for the dispensability of the backward-looking evaluations favored by those other theorists.

Various writers have defended the view that there are varieties of virtue ethics that do not depend on free will or moral responsibility, including Derk Pereboom, Michael Slote, Gary Watson, and myself.[5] To my mind, the key point is that we can recognize that someone possesses a virtue (a character trait which prompts them to virtuous thought and action) without supposing that they are praiseworthy for possessing that virtue, or for the way it prompts them to think and act. In a parallel way, we can recognize a vice without supposing that the agent who has it is blameworthy for having it.

---

[5] D. Pereboom, *Living Without Free Will*, (pp. 149-150); M. Slote, 'Ethics Without Free Will', *Social Theory and Practice,* 16, 3 (1990), pp. 369-383; G. Watson, 'Two Faces of Responsibility', *Philosophical Topics*, 24 (1996), pp. 227–248; and B. Vilhauer, 'Hard Determinism, Humeanism, and Virtue Ethics', *Southern Journal of Philosophy,* 46, 1 (2008), pp. 121-144.

The view that deontology remains a live option if we cannot appeal to free will has also been advocated by various writers, including myself.[6] It is popular to assume that it is essential to deontology to have a 'backward-looking' element derived from action-based desert. But I think we can have a moral system which is robustly deontological without looking backward. We need only look to the present, and to the people with whom we are interacting, to find a basis for desert-claims that (e.g.) constrain utility maximization. To put the point in different words, personhood can provide a basis for desert-claims which is irreducibly different from action, and which does not depend upon free will in the way action-based desert does. For example, all people deserve to be treated only as they would rationally consent to be treated, just because they are people. Since no one can rationally consent to be treated as a mere means, it violates a legitimate personhood-based desert claim to treat someone as a mere means.

Some may object that worst-case free will skepticism downgrades H in a very direct way that I have not yet mentioned, because of the 'ought implies can' principle. That is, it may be objected that if we accept the 'ought implies can' principle, and we interpret this 'can' as entailing free will, then practical seriousness about worst-case free will skepticism downgrades all reasons for action which depend on this notion of 'ought' ('should', 'must', etc.), and H is clearly an example. There are several ways in which we might handle the

'ought implies can' principle. One way is simply to reject it. This approach is endorsed by John Martin Fischer. Fischer is a 'semicompatibilist'—he holds that moral responsibility is compatible with determinism even if free will is not—and in the service of this view, he argues that 'ought' does not imply 'can'.[7] If it is reasonable to see this strategy as available to a compatibilist about determinism and moral responsibility, then it seems reasonable to see it as equally available to skeptics about free will and moral responsibility. But this strategy may be unnecessarily radical.

A more conservative approach is to interpret the 'can' in such a way that it does not entail free will. One way to do this is with an epistemic interpretation of 'can': that is, to say that I can do x is to say that, to the best of my knowledge, it is possible for me to do x. On this interpretation of 'can', it remains true that we can do a variety of things at any given point in the future, no matter what metaphysics of agency we accept, so long as we make reasonable assumptions about what we can know about the future. Suppose I hold that determinism is true, and that it implies that it is possible for me to act in only one particular way at any given point in the future. So long as I cannot predict what that one particular way is, it remains true, on this epistemic interpretation of 'can', that I can act in more than one way. Even if determinism is in fact true, it seems extraordinarily unlikely that we will ever be able to predict how we will act in any detail, given the vast number of variables that would

---

[7] J.M. Fischer, 'Recent Work on Moral Responsibility', *Ethics*, 110 (1999), pp. 93-139, and ' 'Ought-implies-can', Causal Determinism and Moral Responsibility,' *Analysis*, 63, 279 (2003), pp. 244-50.

presumably have to be measured. Absent this knowledge, there will always be multiple things that I can do, in this sense.[8]

My claim here is that there are at least two senses of 'can' that our conceptual toolkit must contain if we are to work in this region of ethics. The epistemic sense of 'can' is about the kinds of possibilities of action agents need to have for 'oughts' to figure in their practical reasoning. The free-will entailing sense of 'can' is about the kinds of possibilities of action agents need to have for it to be appropriate to hold agents morally responsible for acting (or failing to act) in accordance with 'oughts'. Some may object to this multiplying of senses of 'can', but it seems clear there are many different 'cans' involved in ordinary language as well as philosophical theorizing, so this should not be an objection. (For example, if I claim that my adjustable Crescent wrench can fit fasteners of an infinite variety of widths, I seem to be using a 'can' that need not be connected to claims about knowledge of the actual future or moral responsibility.) If we accept the epistemic sense of 'can', then it follows that we need not interpret the 'ought implies can' principle in such a way that H is downgraded by practical seriousness about worst-case free will skepticism.

Philosophers who hold what we might call a realist view of 'ought implies can' may insist that the truth about what I ought to do is conditioned only by the possibilities that are actually open to me in the future, not by what I happen to know about those possibilities. (Libertarians are typically realists in this sense, but so are compatibilists who advocate

---

[8] This epistemic analysis originated as a compatibilist strategy for defusing the threat of determinism. See e.g. T. Kapitan, 'Deliberation and the Presumption of Open Alternatives', *The Philosophical Quarterly*, 36, 14 (1986), pp. 230-51. See Pereboom p. 137 for discussion of this strategy in the context of free will skepticism.

descendants of the Humean 'conditional analysis' of 'could have done otherwise'.[9]) On this realist view, 'oughts' and 'cans' are ontologically prior, as it were, to our reasoning about them. If the reason that I can consistently believe that I can act in some way in the future is merely that I lack knowledge of facts which imply that I cannot act in that way, then my belief about this 'can', though consistent, is false. Since my belief about 'can' is false, my corresponding belief about 'ought', though consistent, is false. It is not really the case that I ought to act in that way.

Even if we accept such realism about 'oughts', we can hold that what matters for practical reasoning is what agents can consistently believe about 'ought' claims, not truths about 'ought'-claims that they cannot actually know. But it is far from obvious that such realism is necessary for 'oughts' and 'cans' to have the sort of objectivity required for practical reasoning. In the Kantian tradition, for example, 'oughts' are understood as commands which reason gives to itself—so 'oughts' cannot be ontologically prior to reasoning about them. If we understand 'oughts' in this way, then it may become more plausible to suppose that the truth about the 'can' in the 'ought implies can' principle is conditioned by the knowledge about the future to which reason happens to have access. This was not Kant's own view, but it is a view which seems open to contemporary metaethicists working in the Kantian tradition.

---

[9] For descendants of the 'conditional analysis', see J.K. Campbell, 'A Compatibilist Theory of Alternative Possibilities', *Philosophical Studies*, 88, 3 (1997), pp. 319-330, and 'Compatibilist Alternatives', *Canadian Journal of Philosophy,* 35, 3 (2005), pp. 387-406, and also K. Vihvelin, 'Free Will Demystified: A Dispositional Account', *Philosophical Topics*, 32 (2004), pp. 427-450.

It may strike some readers as odd, or even disingenuous, for a paper which advocates skepticism about the 'can' of free will to be so unskeptical about the 'can' in the 'ought implies can' principle. But the motivation for splitting the difference in this way flows directly from the metaethical outlook I am advocating here. That is, all mainstream moral theories would appear to include 'oughts', so 'oughts' are basic moral reasons. But not all mainstream moral theories include the action-based desert-claims that derive from free will and moral responsibility. So we can split the difference in this way even if we are committed to maintaining basic moral reasons. And given that some of the most damaging harm in our current practices is justified by making claims about free will, philosophers inclined to advance H have a very good reason to take this approach.

This approach may still seem wanting to some readers. I think it is fair to say that this is a coherent account of how we can endorse 'oughts' even if we become practically serious about worst-case free will skepticism. But no matter how sympathetic some readers may be to the goal of advancing H, they may find epistemic analyses of the 'can' in 'ought implies can' to be too weak. However, as mentioned earlier, there are degrees of practical seriousness about free will skepticism which we might accept. Readers who would like to be skeptical enough about free will to advance H, but not so skeptical that they have to worry about the 'can' in the 'ought implies can' principle, might do well to consider best-case practical seriousness about free will skepticism as an alternative.

4. Best-Case Practical Seriousness

Suppose we accept that the arguments for skepticism about some claim C might be strong enough to make it the case that we sometimes ought to refrain from asserting C when deliberating about how to act, but only in cases where it is appropriate to apply the highest practical justificatory standard. This view is what I mean by 'best-case practical seriousness'.

Best-case practical seriousness about free will skepticism might appeal to philosophers who think that arguments for free will skepticism might be strong enough to undermine retributive arguments for overriding H, but could not be strong enough to undermine justifications of praise, and could not be strong enough to undermine the 'can' in the 'ought implies can' principle, even if this 'can' entails free will. This is because it seems natural to suppose that we must meet a higher justificatory standard to appeal to claims about free will when we seek to override H than when we seek to justify praise, or to defend the existence of moral reasons involving 'oughts'. It is widely agreed that arguments for overriding H must be held to the highest justificatory standard, at least in cases where serious harm is at issue, and I think this view is correct. To speak impressionistically, I think there is a strong intuition that there is a sort of 'sliding scale' built into H, according to which the extremely strong reason not to intentionally harm others gets stronger as the harm at issue becomes more severe, in such a way that arguments for doing serious intentional harm to others are subject to the highest standard of practical justification. I take there to be the same sort of basic agreement about this point that there is about H itself.[10] I will not try to define precisely how high the highest practical justificatory standard is, since I think it is clear enough that whatever its precise height may be, it is the right standard to apply to arguments

---

[10] Act-utilitarians may object to the claim that arguments for seriously harming people have to be held to an especially high standard. But as explained earlier, under the conditions of limited predictability in which we reason practically, act-utilitarians have a strong reason to avoid intentionally inflicting harm as a means to the end of maximizing happiness, because the immediate effect of this means is to reduce overall happiness. As the harm at issue becomes more severe, act-utilitarians must raise the justificatory standard they apply to arguments for inflicting it.

for doing serious intentional harm to others. (It is probably not as high as certainty, which is arguably an appropriate standard for arguments about mathematics and logic, but which is probably not appropriate in the messy empirical contexts in which we must reason about how to act. I suspect that the 'beyond reasonable doubt' standard applied in the criminal courtroom is about as precise an articulation of this standard as philosophers can hope to find.[11])

If this is right, then best-case practical seriousness includes accepting that arguments for skepticism about some claim C might be strong enough that we ought not assert C in arguments for overriding H, at least where serious harm is at issue. Based on this conclusion, I will now argue that even best-case practical seriousness about other minds or induction skepticism downgrades H.

Suppose that Al, Bernie, Connie, and Dev are the sole members of an island shipwreck society. Al murders Bernie in cold blood with some means that Al could be confident (skepticism about induction aside) would more than suffice to kill Bernie (say stabbing Bernie 50 times). Suppose that Connie and Dev have Al in custody afterwards, and they are debating about how to treat him. Suppose Connie argues for overriding H: 'Al deserves to suffer serious harm for injuring a person in a way that Al should have expected to cause that person to suffer and die. We have a reason to inflict serious harm on those who deserve it, and this reason is stronger than H. So we ought to inflict serious harm on Al.'

---

[11] Pereboom suggests applying the 'reasonable doubt' standard in *Living Without Free Will* (p. 161). I discuss this issue in more detail in B. Vilhauer, 'Free Will and Reasonable Doubt', *American Philosophical Quarterly,* 46, 2 (2009), pp. 131-140.

Now consider three ways in which Dev might reply in defense of Al, by appealing to skepticism about other minds, induction, or free will. First, suppose Dev is practically serious about other minds skepticism. He might challenge Connie's argument as follows: 'We don't know that other minds exist, so it might not be legitimate to appeal to the existence of other minds in arguments that have to meet the standard appropriate for overriding H. Al could only deserve serious harm if Bernie was a person, but Bernie could not have been a person unless he had a mind, so if it might be illegitimate to appeal to claim that Bernie had a mind, it might be illegitimate to appeal to the claim that Bernie was a person.' But upon inspection, Dev's argument begins to unravel in a way that makes its upshot unclear. There are at least two problems with Dev's argument. First, it works by claiming that H is so strong that we cannot assume that the existential conditions necessary for H to be practically relevant are fulfilled when making arguments for overriding H—it claims that the relevance of H in the context of Al's relation to Bernie is not clear enough for Connie and Dev to appeal to it in an attempt to override H. So whatever work it does to support H in one context, it accomplishes only by downgrading H elsewhere. The second problem is that if Connie and Dev reflect on the first problem, then it would be natural for them to ask why, if H now looks oddly weak in the context of Al's relation to Bernie, they are obligated to hold arguments for overriding H to such a high standard in the context of their relation to Al. To ask this would be to downgrade H in the latter context too.

If Dev instead appeals to induction skepticism, then he ends up in a similar situation. Suppose Dev argues as follows: 'We don't have knowledge about the products of induction, so it might not be legitimate to appeal to them in arguments that have to meet the standard appropriate for overriding H. So it might not be legitimate to claim that Al should have expected the stabbing to cause Bernie to suffer and die, because expectations about the effects of stabbings are products of induction.' Once again, this move seeks to block the

argument for overriding H in the context of Connie's and Dev's relation to Al by downgrading H in the context of Al's relation to Bernie. And once again, if Connie and Dev reflect on this, it would be natural for them to ask why they should be so sure about the strength of H in the context of their relation to Al, which would downgrade H in this context as well.

Things are different if Dev appeals to free will skepticism. Suppose Dev argues that we don't know we have free will, so it might not be legitimate to appeal to the claim that free will exists in arguments for overriding H—and since we can only claim that Al deserves to suffer for his action if we can claim that he acted with free will, it therefore might not be legitimate to claim that he deserves to suffer for his action. The upshot is that Connie's argument for overriding H is challenged without presenting H as downgraded.

5. Advancing the Basic Reason Not to Intentionally Harm

Thus far I have argued that a commitment to maintaining H is compatible with practical seriousness about free will skepticism, but not other minds skepticism or induction skepticism. However, as is probably obvious, a commitment to maintaining H is also compatible with rejecting practical seriousness about free will skepticism. So, on its own, a commitment to maintaining H does not give us a reason to become practically serious about free will skepticism. At best it gives those committed to maintaining H a reason not to regard it as problem if they suddenly find themselves practically serious about it.

However, there is an attitude toward H which I think does provide a reason to become practically serious about free will skepticism, which (as mentioned earlier) I will call the inclination to *advance* H. While commitment to maintaining a reason is a resolution to reject claims that downgrade that reason, to be inclined to advance a reason is to be inclined to accept, all other things being equal, claims that upgrade it (e.g. claims that strengthen it,

22

buttress the existential commitments presupposed by it, or diminish the incidence of its being overridden by other reasons).

People inclined to advance H have a reason to become practically serious about free will skepticism because of the way it downgrades the retributive reason for overriding H. That is, downgrading the retributive reason for overriding H upgrades H, because it diminishes the incidence of H's being overridden.

People inclined to advance H probably have better reason to be practically serious about best-case free will skepticism than about worst-case free will skepticism. This is because practical seriousness about best-case free will skepticism targets arguments for harming others in a way that practical seriousness about worst-case free will skepticism does not. As explained in the last section, practical seriousness about best-case free will skepticism only requires us to recognize the possibility that it might be illegitimate to appeal to free will in arguments that must meet the highest justificatory standard, and this effectively targets arguments for doing serious intentional harm. But practical seriousness about worst-case free will skepticism requires us to recognize the possibility that it might not be legitimate to appeal to free will at all, even in the context of praise that benefits some people and harms no one, and in the context of defending the existence of 'ought'-based moral reasons against objections drawn from the 'ought implies can' principle. Worst-case free will skepticism is compatible with an inclination to advance H, for the same reasons that it is compatible with a commitment to maintaining H, but worst-case practical seriousness is not as well-tailored to the needs of philosophers interested in advancing H as best-case practical seriousness is. Philosophers inclined to advance H might find the best fit of all in an attitude somewhere between worst-case and best-case free will skepticism, since there may be arguments for intentionally harming others in ways that are not serious, which need not meet the highest justificatory standard, but which must meet a higher justificatory standard than

arguments for praise that benefits some people and harms no one. But I will not try to describe that attitude more precisely here.

I will not claim that anyone *should* be inclined to advance H, only that it strikes me as a sensible and natural attitude for philosophers to take. As probably goes without saying, many philosophers choose a goal to pursue through argumentation, and hope that their own small contributions to theory change will help to achieve that goal. This approach to philosophy is not obligatory—one can certainly hope to be driven to one's conclusions by the sheer force of argument alone, without choosing a goal at which one hopes to arrive. But it seems fair to say that such goal-oriented philosophizing is as common in the discipline as philosophizing that eschews goals. It seems to me that the goal of advancing H is as sensible a goal as any other, and more sensible than many, given H's status as a basic moral reason.

6. Non-Retributive Reasons for Overriding H

We can be committed to maintaining H, or inclined to advance H, even if we do not hold the arguably over-optimistic view that there are never good reasons for overriding H. It may be that, given human nature, we cannot keep society functioning without both an institution of punishment and also a moral psychology that includes painful moral emotions (like guilt) which help regulate behavior. Those inclined to advance H should advocate less harmful ways of achieving these goals. Non-retributive moral theories provide a variety of explanations of how we can justify overriding H enough to keep society functioning. I lack the space to give a complete account of these explanations, but let me briefly sketch a few.

Act-utilitarianism provides one obvious example—it has no trouble justifying the imposition of harm in a way that minimizes overall harm, whether the harm comes in the form of self-regulating emotion or punishment. But philosophers committed to advancing H should be cautious about accepting act-utilitarianism. Since act-utilitarianism can sometimes

recommend punishments of terrible severity as means to overall harm minimization, act-utilitarianism does nothing to reduce the amount of harm we can justifiably inflict on individuals. The suffering of a few vivisected pickpockets may in some circumstances be outweighed in the act-utilitarian calculus by all the suffering prevented through deterring crime.

However, the deontological approach mentioned earlier, which appeals to personhood-based desert instead of action-based desert, can generate a justification of punishment which constrains utility maximization.[12] It holds that simply because we are persons, we deserve to be treated only as we would rationally consent to be treated. It is popular to explain rational consent in terms of Rawlsian original position deliberation. Suppose that we had to choose institutions of punishment behind the veil of ignorance, assuming that we had an equal chance of finding ourselves among the punished or among those protected by the institution of punishment. It would be rational to choose to imprison violent criminals in benign prisons (prisons much less harmful than contemporary ones). For people with normal social attachments, and a normal desire to be free from interference in the pursuit of their ends, prison would inevitably be harmful, even under benign conditions.[13] But it would make sense to risk that harm if it significantly diminished our chances of being murdered or seriously injured by a violent criminal. If we chose penal institutions in this

---

[12] I discuss this in more detail in 'Persons, Punishment, and Free Will Skepticism'.

[13] Smilansky argues that free will skeptics who are concerned about justice cannot justify punishing in a way that deters, but I think that even benign imprisonment is harmful enough to deter. See S. Smilansky, 'Hard Determinism and Punishment: A Practical *Reductio*', *Law and Philosophy,* 30, 1 (2011), pp. 352-367.

way, we would probably not choose to incarcerate non-violent criminals—we would probably be content with a system of fines and community service. But I doubt that we need to incarcerate non-violent criminals to keep society functioning.

To keep society functioning, outward penal institutions may need to be supplemented by inward experiences of painful moral emotions like guilt, which help regulate behavior. We can appeal to a non-retributive virtue ethics to explain the appropriateness of such emotional experiences.

Painful moral emotions are often understood on the model of retributive punishment. That is, they are often understood as ways in which agents who have acted immorally impose deserved suffering upon themselves. (This model goes back at least to Freud.[14]) On this model, encouraging others to feel painful moral emotions is a way of exacting indirect retribution, by prompting the targets of retribution to exact retribution upon themselves. This model is probably accurate for many people's experiences of painful moral emotions in contemporary society, in the sense that many people do believe they deserve to suffer painful emotions for their immoral actions, and they may successfully cause themselves to experience painful emotions based on this belief. But this model is not necessary or helpful for understanding all painful moral emotions. For example, even in contemporary society, there is a kind of remorse which is based on an emotional attachment to the victim which the wrongdoer forms in the wake of the wrongful act, which prompts the wrongdoer to suffer empathically along with the person he has wronged.[15] In other words, in this kind of

---

[14] See e.g. S. Freud, *Civilization and its Discontents* (New York: Norton, 1989), pp 83-96.

[15] I discuss this in more detail in 'Hard Determinism, Remorse, and Virtue Ethics', *Southern Journal of Philosophy*, 42, 4 (2004), pp. 547-564.

remorse, the wrongdoer's suffering derives from the fact that he comes to care about the victim in the wake of the wrongful act. In my view, the explanation of why wrongdoers suffer in cases like this need not be fundamentally different from the explanation of why we suffer when our friends or loved ones suffer. This is part of what it means to be attached to others, and insofar as such attachment is virtuous, we can recognize empathic remorse as virtuous. What makes empathic remorse different from friendship or love is that the wrongdoer can recognize that the victim's suffering was caused by his actions, and that such suffering can be prevented in the future if he acts differently. I see no reason why this sort of remorse may not play just as effective a role in regulating behavior as other kinds of painful moral emotions. If this is right, then it seems plausible to claim that it would be virtuous to encourage agents to feel this sort of remorse when they have reason to do so. This would imply that a non-retributive virtue ethics can provide grounds for overriding H in encouraging others to feel painful moral emotions that help regulate behavior.

7. Conclusion

If the arguments I have presented here are correct, then philosophers committed to maintaining or advancing H can be practically serious about free will skepticism, but not induction or other minds skepticism, and philosophers inclined to advance H have good reason to be practically serious about free will skepticism. Are there other canonical philosophical skepticisms which are similar to free will skepticism in this respect? Skepticism about personal identity over time seems to be a likely candidate. Even worst-case personal identity skepticism seems compatible with a commitment to H. That is, we can question claims about personal identity without questioning the more fundamental claim that persons exist, and it is this more fundamental claim upon which H depends. Best-case personal identity skepticism would target arguments for the retributive justification of serious

27

harm without doing anything to downgrade H. If we cannot appeal to the claim that the current target of retributive harm is the same person as the earlier agent of the action for which punishment is proposed, then we cannot retributively justify harm. But questioning the identity of the target does nothing to downgrade H, since we can question whether the target of retribution is the same person as the earlier agent without suggesting that the force of H on the earlier agent was any weaker than a non-skeptic would have taken it to be. If this is right, then philosophers inclined to advance H have good reason to be practically serious about personal identity skepticism as well.