

Running Risks Morally*

Brian Weatherson

2014

I defend normative externalism from the objection that it cannot account for the wrongfulness of moral recklessness. The defence is fairly simple—there is no wrong of moral recklessness. There is an intuitive argument by analogy that there should be a wrong of moral recklessness, and the bulk of the paper consists of a response to this analogy. A central part of my response is that if people were motivated to avoid moral recklessness, they would have to have an unpleasant sort of motivation, what Michael Smith calls “moral fetishism”.

This paper is part of a project defending **normative externalism**. This is the view that the most important norms concerning the guidance and evaluation of action and belief are external to the agent being guided or evaluated. The agent simply may not know what the salient norms are, and indeed may have seriously false beliefs about them. The agent may not have any evidence that makes it reasonable to have true beliefs about what the salient norms are, and indeed may have misleading evidence about them. But this does not matter. What one should do, or should believe, in a particular situation is independent of what one thinks one should do or believe, and (in some key respects) of what one’s evidence suggests one should do or believe.

There are three important classes of argument relevant to the debate between normative externalists, in the sense of the first paragraph, and normative internalists. One class concerns intuitions about cases. For instance, we might try to defend normative externalism by arguing that according to the internalist, but not the externalist, there is something bad about Huckleberry Finn’s actions in helping Jim escape. Nomy Arpaly (2002) uses this example as part of an argument for a sophisticated form of externalism.

*I’ve discussed this paper with just about everyone I know. Thanks to Elizabeth Anderson, Rachael Briggs, Lara Buchak, Sarah Buss, Justin D’Arms, Tom Dougherty, Dmitri Gallow, Alex Guerrero, Elizabeth Harman, Scott Hershovitz, Ishani Maitra, Julia Markovits, Jill North, Timothy Schroeder, Andrew Sepielli, Ted Sider, Rohan Sud, Sigrún Svavarsdóttir and Julie Tannenbaum for suggestions that particularly improved the paper. This paper was presented to the EDGe group at the University of Michigan and the philosophy department at Ohio State University, and I got valuable feedback at both of those presentations. The paper also served as my inaugural lecture as the Marshall M. Weinberg Professor at the University of Michigan. Marshall has been a wonderful supporter of the University of Michigan for many years, and especially of its philosophy department, and this was a tremendous honour. And the paper was presented at the 2013 Bellingham Summer Philosophy Conference. This is close to the Platonic Ideal of a philosophy conference. I’m incredibly grateful to Ned Markosian, and to all of the people who work with him to make this conference happen every year. And I’m very happy to have been able to present this paper at the 2013 conference. So to Marshall and to Ned, thanks.

Another class concerns views about the nature of norms. Internalists think that externalists have missed the need for a class of subjective norms, that are sensitive to agents' views about the good. Externalists think that the norms internalists put forward are incoherent, or do not meet the internalists's needs. I'll gesture at these arguments below, but they are made in much more detail in recent work by Elizabeth Harman (2015) responding to internalist proposals.

But there's a third class of argument where the internalist may seem to have an edge. Internalists can argue that there is a wrong of moral recklessness, and externalists cannot explain what is wrong about moral recklessness. My response will be fairly blunt; I do not think moral recklessness is wrong. But I'll start by trying to state the case for the wrongness of moral recklessness as strongly as I can, including clarifying just what moral recklessness is, before moving onto a response on behalf of the externalist.

1 Moral Uncertainty

Some of our moral opinions are pretty firmly held. Slavery really is wrong; rescuing drowning children is good; and so on. But others might be more uncertain. To use an example I'll return to a lot, even a lot of carnivores worry that it isn't obvious that killing animals to eat their flesh is morally permissible.

We might wonder whether this uncertainty should have practical consequences. Uncertainty in general does have practical, and even moral, consequences. If you're pretty sure the bridge is safe, but not completely certain, you don't cross the bridge. If you're only sorta kinda confident that an action won't kill any innocent bystanders, and there is no compelling reason to do the action, it would be horribly immoral of you to do it.

There are (at least) two ways to be uncertain about the morally significant consequences of your action. You might know the moral significance of everyone who might be harmed by your action, but not know how many of them will be harmed, or how seriously they will be harmed. Someone who habitually runs red lights is in this position. They know there's an elevated risk that they'll kill another human this way, and they know the human they would kill is morally valuable. Alternatively, you might know who or what is affected by your action, but not be sure of their moral status. The hesitant carnivore is like this. They know that steak dinners require killing cows, but they aren't sure how morally significant the cows are.

Perhaps that's a distinction without a difference though. In both cases, the action results in a higher probability of something morally significant being killed. And, one might think, that's enough to give the actor reason to pause before acting, and enough to give us reason to condemn the action.

As may be clear from the introduction, that's not how I think of the cases. I think the distinction I just flagged is very important both practically and morally. Being uncertain about the physical consequences of your actions should matter both to what you do, and how you are assessed. The red light runner is immoral, even if she never actually harms anyone, because she endangers morally significant humans. But the meat eater cannot be condemned on the same grounds. If she is wrong that meat eating is morally

acceptable, that would be one thing. But a mere probability that meat eating is immoral should not change one's actions, or one's evaluations of meat eaters.

Now I won't pretend this is a particularly intuitive view. In fact, quick reflection on a few cases may make it seem that it is extremely unintuitive. Let's look at three such cases.

Cake

Carla is baking a cake for a fundraiser. She wants to put some sweetening syrup into the cake to improve its taste. She reaches for an unmarked bottle, which she is pretty sure contains the sweetener she wants. But then she remembers that last week she had some arsenic in a similar bottle. She is pretty sure she threw the arsenic out, but not exactly certain. As a matter of fact, the syrup in the bottle is sweetener, not arsenic, but Carla isn't certain of this. What should she do?

Dinner

Martha is deciding whether to have steak or tofu for dinner. She prefers steak, but knows there are ethical questions around meat-eating. She has studied the relevant biological and philosophical literature, and concluded that it is not wrong to eat steak. But she is not completely certain of this; as with any other philosophical conclusion, she has doubts. As a matter of fact, Martha is right in the sense that a fully informed person in her position would know that meat-eating was permissible, but Martha can't be certain of this. What should she do?

Abortion

Agnes is twelve weeks pregnant, and wants to have an abortion. She has studied the relevant medical and philosophical literature, and is pretty sure that foetuses at this stage of development are not so morally significant as to make abortion wrong. But she is not completely certain of this; as with any other philosophical conclusion, she has doubts. As a matter of fact, Agnes is right in the sense that a fully informed person in her position would know that abortion was permissible, but Martha can't be certain of this. What should she do?

The setup of the last two cases is a bit cumbersome in one key respect; I had to refer to what a fully informed person in Martha or Agnes's position would know. I did this so as to not beg any questions against the internalist. I would rather say simply that Martha and Agnes were simply right in their beliefs. But I'm not sure how to make sense of this from an internalist perspective. If what's right to do is a function of your moral evidence and beliefs, perhaps there is a sense in which meat-eating or abortion is objectively permissible, but Martha and Agnes can't truly believe it is permissible, since it isn't permissible in their subjective state, and that's the really important kind of permissibility. So the retreat to talking about what a fully informed person would know is my attempt to find an objective point at which the internalist and externalist

can agree. It doesn't signal that I think there's anything special about fully informed agents; I'm just trying to avoid being question-begging here.

You might also think that one or other of these cases is very far removed from reality. Perhaps what counts as meat or a foetus would have to be very different for these cases to be possible, perhaps so different that they wouldn't deserve the label 'meat' or 'foetus'. I don't think this should worry us. I don't particularly care if the cases are metaphysically possible or not. There's a world, epistemically if not metaphysically possible, where the medical and biological facts are as they are and meat-eating and abortion are permissible, and that's the world I mean these examples to be set in. By allowing that my thought experiments may well be set in metaphysically impossible worlds, I am going against some recent views on thought experiments as put forward by, e.g. Timothy Williamson (2007) and Anna-Sara Malmgren (2011), but it would take us too far afield to defend this bit of apostasy. Instead, I'll just use the cases as they are.

Finally, note that I've set up the cases where the protagonists are almost, but not entirely, sure of something that is in fact true. And I'm going to argue in the moral case that they should act as if they are right. That's not because I think that a view one is almost sure of should be acted on; one should act on the moral truths, and Agnes and Martha are close to certain of the actual truth. The reason for picking these cases is that they make the issue of recklessness most salient. If any of the three women do anything wrong (and I think Carla does) it is only because they are reckless.

That said, there is something interestingly in common to the three cases. In each case, the agent has a choice that is, if taken freely, clearly morally acceptable. Carla can leave out the syrup, Agnes can continue the pregnancy, and Martha can order the tofu. At least, that's true on the most natural ways to fill out the details of the case.¹ So assume that Carla, Martha and Agnes are correctly completely certain that they have a morally safe option. Also assume, if it isn't clear already, that their only motivation for taking the safe option is to hedge against a possibility that they think is rather unlikely. Hedges can be valuable, so the fact that this is their only motivation is not a reason to not take the safe option.

In contemporary debates, it's not often you see pro-vegetarianism and anti-abortion arguments run side by side. Especially in America, these debates have been caught up in culture war politics, and on the whole vegetarians are on one side of this debate, and anti-abortion activists on the other side. But the debates do have some things in common, and it is their commonality that will interest us primarily here. In particular, we'll be looking at the idea that one should be vegetarian, and refrain from having abortions,

¹Here is one argument against the claims of the last two sentences. Assume that, as is realistic, Agnes wants an abortion because her life will be worse in significant ways if she becomes a parent (again) in the near future. And assume that Agnes has a moral duty to herself; making her own life worse in significant ways for no sufficient reason is immoral. Then it could be immoral for her to continue the pregnancy. I don't find this reason particularly compelling; it seems to me odd to say that people who make heroic sacrifices are immoral in virtue of paying insufficient regard to their own welfare. But the issues here are difficult, and I certainly don't have a strong argument that we should give no credence to the view that there are substantial duties to self that make misguided sacrifices on behalf of others immoral. Still, I'm going to set this whole line of reasoning aside for most of the paper, while just noting that this could be a way even for an internalist to reject the practical arguments I'll discuss below. I'm grateful to conversations with Elizabeth Anderson here (but not only here!).

on the grounds that these are the good safe options to take. (This connection between the debates is not a novel observation. D. Moller (2011, 426) notes it, and makes some pointed observations about how it affects the philosophical landscape.)

I'm going to argue that the idea that all three women should 'play it safe' is entirely the wrong lesson to take from the cases. I think the cases are in important respects disanalogous. It is seriously morally wrong for Carla to include the syrup in the cake, but it is not wrong in the same way for Martha to eat the steak, or for Agnes to have the abortion. A little more precisely, I'm going to be arguing that there is no good way to fill in the missing premise of this argument.

The 'Might' Argument

1. In the circumstances that Agnes/Martha are in, having an abortion /eating a steak might be morally wrong.
2. In the circumstances that Agnes/Martha are in, continuing the pregnancy /eating vegetables is definitely morally permissible.
3. **Missing Premise**
4. So, Agnes should not have the abortion, and Martha should not eat the steak.

When I argue that the 'Might' Argument cannot be filled in, I'm arguing against philosophers who, like Pascal, think they can convince us to act as if they are right as soon as we agree there is a non-zero chance that they are right. I'm as a rule deeply sceptical of any such move, whether it be in ethics, theology, or anywhere else.

But note like someone responding to Pascal's Wager, I'm focussing on a relatively narrow target here. Rejecting Pascal's Wager does not mean rejecting theism; it means rejecting Pascal's argument for being a theist. Similarly, rejecting the 'Might' Argument does not mean rejecting all ethical arguments against meat-eating or abortion. It just means rejecting this one.

I'm also not arguing about public policy here. The 'Might' Argument can be generalised to any case where there is an epistemic asymmetry. The agent faces a choice where one option is morally risky, and the other is not. Public policy debates are rarely, if ever, like that. A legislator who bans meat-eating or abortion takes a serious moral risk. They interfere seriously with the liberties of the people of their state, and perhaps do so for insufficient reason. (This point is well made by Moller (2011, 442).) So there isn't a 'play it safe' reason to support anti-meat or anti-abortion legislation, even if I'm wrong and there is such a reason to think that individuals should not eat meat or have abortions.

There are two ways to try to fill out the 'Might' Argument. We could try to offer a particular principle that implies the conclusion given the rest of the premises. Or we could try to stress the analogy between the three cases that I started with. I'm going to have a brief discussion of the first option, and then spend most of my time on the analogy. As we'll see, there are many possible principles that we could try to use here, but hopefully what I say about some very simple principles, plus what I say about the analogy, will make it clear how I want to respond to most of them.

2 Principles

One way to fill in the **Missing Premise** is to have a general principle that links probabilities about morality with action. The simplest such principle that would do the trick is this.

ProbWrong If an agent has a choice between two options, and one might be wrong, while the other is definitely permissible, then it is wrong to choose the first option.

I think **ProbWrong** does a reasonable job of capturing the intuition that Agnes and Martha would be running an impermissible risk in having an abortion or eating meat. But **ProbWrong** has clearly implausible consequences. Imagine that an agent has the following mental states:

1. She is sure that **ProbWrong** is true.
2. She is almost, but not completely, sure that eating meat is permissible for her now.
3. She is sure that eating vegetables is permissible for her now.
4. She is sure that she has states 1–3.

A little reflection shows that this is an incoherent set of states. Given **ProbWrong**, it is simply wrong for someone with states 2 and 3 to eat meat. And the agent knows that she has states 2 and 3. So she can deduce from her other commitments and mental states that eating meat is, right now, wrong. So she shouldn't be almost sure that eating meat is permissible; she should be sure that it is wrong.

This argument generalises. If 1, 3 and 4 are true of any agent, the only ways to maintain coherence are to be completely certain that meat eating is permissible, or completely certain that it is impermissible. But that is, I think, absurd; these are hard questions, and it is perfectly reasonable to be uncertain about them. At least, there is nothing incoherent about being uncertain about them. But **ProbWrong** implies that this kind of uncertainty is incoherent, at least for believers in the truth of **ProbWrong** itself. Indeed, it implies that in any asymmetric moral risk case, an agent who knows the truth of **ProbWrong** and is aware of her own mental states cannot have any attitude between certainty that both options are permissible, and certainty that the risky action is not, for her, permissible. That is, I think, completely absurd.

Now most philosophers who advocate some principle or other as the **Missing Premise** don't quite advocate **ProbWrong**. We can position some of the rival views by abstracting away from **ProbWrong** as follows.

General Principle If an agent has a choice between two options, and one might be X, while the other is definitely not X, then it is Y to choose the first option.

We get **ProbWrong** by substituting 'wrong' for both X and Y. But we saw a decisive objection to that view. And we get a version of that objection for any substitution where

X and Y are the same. So a natural move is to use different substitutions. If you replace X with ‘wrong’ and Y with ‘irrational’, you get something like a principle defended by Ted Lockhart (2000).

What Might be Wrong Is Irrational If an agent has a choice between two options, and one might be wrong, while the other is definitely not wrong, then it is irrational to choose the first option.

Now at this stage we could look at whether this principle is plausible, and if not whether alternative principles offered by Alex Guerrero (2007), Andrew Sepielli (2009) or others are any better. You can probably guess how this would go. We’d spend some time on counterexamples to the principle. And we’d spend some time on whether the conclusion we get in this particular case is really plausible. (Is it true that Martha is not in any way immoral, but is irrational in virtue of moral risk? That doesn’t sound at all like the right conclusion.)

But I’m not going to go down that path. Shamelessly stealing an analogy from Jerry Fodor (2000), I’m not going to get into a game of Whack-a-Mole, where I try to reject a principle that could fill in for the Missing Premise, and if I succeed, another one pops up. I’m not playing that game because you never actually win Whack-a-Mole; by going through possible principles one at a time it isn’t clear how I could ever show that no principle could do the job.

What I need to show is that we shouldn’t look for a principle to fill in as Missing Premise. One reason we shouldn’t is that the intuitions behind principles like Lockhart’s is really an intuition in favour of **ProbWrong**, and as such should be suspect. But a better reason is that the analogy between Carla’s case and Agnes/Martha’s cases that motivated the thought that there should be some principle here is mistaken. Once we see how weak that analogy is, I think we’ll lose motivation for trying to fix **ProbWrong**.

3 Welfare and Rationality

So my primary opponent the rest of the way is someone who wants to defend the ‘Might’ Argument by pressing the analogy between Carla’s case and the two more morally loaded cases.² My reply will be that there are better analogies than this which point in the opposite direction. In particular, I’m going to draw an analogy between Agnes and Martha’s cases with some tricky cases concerning prudential reasoning. To set up the case, I’ll start with an assumption that guides the discussion.

The assumption is that deliberately undermining your own welfare, for no gain of any kind to anyone, is irrational. Indeed, it may be the paradigmatic form of irrationality. This is, I think, a widely if not universally held view. There is a radically Humean view that says that welfare just consists of preference satisfaction, and rationality is just a

²D. Moller (2011) offers an interesting different analogy to motivate something like the ‘Might’ Argument. I think that analogy is a little messier than the one I’m focussing on, and I’ll discuss it separately below.

matter of means-end reasoning. If that's right then this assumption is not only right, it states the only kind of irrationality there is. But you don't have to be that radical a Humean, or really any kind of Humean at all, to think the assumption is true.

The assumption doesn't just mean that doing things that you know will undermine your welfare for no associated gain is irrational. It means that taking serious risks with your welfare for no compensating gain is irrational. Here is a clear example of that.

Eating Cake

Ricky is baking a cake for himself. He wants to put some sweetening syrup into the cake to improve its taste. He reaches for an unmarked bottle, which he is pretty sure contains the sweetener he wants. But then he remembers that last week he had some arsenic in a similar bottle. He is pretty sure that he threw the arsenic out, but not exactly certain. As a matter of fact, the bottle does contain sweetener, not arsenic, but Ricky isn't completely sure of this. What should he do?

I hope it is plausible enough that it would be irrational for Ricky to put the syrup in the cake. The risk he is running to his own welfare – he literally will die if he's wrong about what's in the bottle – isn't worth the gain in taste, given his level of confidence.

With that said, consider two more examples, Bob and Bruce. Bob has thought a bit about philosophical views on welfare. In particular, he has spent a lot of time arguing with a colleague who has the G. E. Moore-inspired view that all that matters to welfare is the appreciation of beauty, and personal love.³ Bob is pretty sure this isn't right, but he isn't certain, since he has a lot of respect for both his colleague and for Moore.

Bob also doesn't care much for visual arts. He thought that art is something he should learn something about, both because of the value other people get from art, and because of what you can learn about the human condition from it. And while he's grateful for what he learned while trying to inculcate an appreciation of art, and he has become a much more reliable judge of what's beautiful and what isn't, the art itself just leaves him cold. I suspect most of us are like Bob about some fields of art; there are genres that we feel have at best a kind of sterile beauty. That's how Bob feels about most visual art. This is perhaps unfortunate; we should feel sorry for Bob that he doesn't get as much pleasure from great art as we do. But it doesn't make Bob irrational, just unlucky.

Finally, we will suppose, Bob is right to reject his colleague's Moorean view on welfare. Appreciation of art isn't a constituent of welfare. In the example we'll suppose welfare is a matter of health, happiness and friendship. So a fairly restricted version of an objective list theory of welfare is correct in Bob's world. And for people who like art, appreciating art can produce a lot of goods. Some of these are direct - art can make you happy. And some are indirect - art can teach you things and that learning can contribute to your welfare down the line. But if the art doesn't make you happy, as it doesn't make

³It would be a bit of a stretch to say this is Moore's own view, but you can see how a philosopher might get from Moore to here. Appreciation of beauty is one of the constituents of welfare in the objective list theory of welfare put forward by John Finnis (2011, 87–88).

Bob happy, and one has learned all one can from a genre, as has Bob, there is no welfare gain from going to see art. It doesn't in itself make you better off, as Bob's Moorean colleague thinks.

Now Bob has to decide whether to spend some time at an art gallery on his way home. He knows the art there will be beautiful, and he knows it will leave him cold. There isn't any cost to going, but there isn't anything else he'll gain by going either. Still, Bob decides it isn't worth the trouble, and stays out. He doesn't have anything else to do, so he simply takes a slightly more direct walk home, which (as he knows) makes at best a trifling gain to his welfare.

I think Bob is perfectly rational to do this. He doesn't stand to gain anything at all from going to the gallery. In fact, it would be a little perverse, in a sense we'll return to, if he did go.

Bruce is also almost, but not completely certain, that health, happiness and friendship are the sole constituents of welfare.⁴ But he worries that this is undervaluing art. He isn't so worried by the Moorean considerations of Bob's colleagues. But he fears there is something to the Millian distinction between higher and lower pleasures, and thinks that perhaps higher pleasures contribute more to welfare than lower pleasures. Now most of Bruce's credence goes to alternative views. He is mostly confident that people think higher pleasures are more valuable than lower pleasures because they are confusing causation and constitution. It's true that experiencing higher pleasures will, typically, be part of experiences with more downstream benefits than experiences of lower pleasures. But that's the only difference between the two that's prudentially relevant. (Bruce also suspects the Millian view goes along with a pernicious conservatism that values the pop culture of the past over the pop culture of the present solely because it is past. But that's not central to his theory of welfare.) And like Bob, we'll assume Bruce is right about the theory of welfare in the world of the example.

Now Bruce can also go to the art gallery. And, unlike Bob, he will like doing so. But going to it will mean he has to miss a night playing video games that he often goes to. Bruce knows he will enjoy the video games more. And since playing video games with friends helps strengthen friendships, there may be a further reason to skip the gallery and play games. Like Bob, Bruce knows that there can be very good consequences of seeing great art. But also like Bob, Bruce knows that none of that relevant here. Given Bruce's background knowledge, he will have fun at the exhibition, but won't learn anything significant.

Still, Bruce worries that he should take a slightly smaller amount of higher pleasure rather than a slightly larger amount of lower pleasure. And he's worried about this even though he doesn't give a lot of credence to the whole theory of higher and lower pleasures. But he doesn't go to the gallery. He simply decides to act on the basis of his preferred theory of welfare, and since that welfare is correct, he maximises his welfare by doing this.

Now I think both Bob and Bruce are rational in what they do. But there is an argu-

⁴Thanks to Julia Markovits for suggesting the central idea behind the Bruce example, and to Jill North for some comments that showed the need for it.

ment that they are not. I'll focus on Bob, but the points here generalise.

1. Going to the gallery might increase his welfare substantially, since it will lead to more appreciation of beauty, and appreciation of beauty might be a key constituent of welfare.
2. Not going to the gallery definitely won't increase his welfare by more than a trivial amount.
3. It is irrational to do something that might seriously undermine your own welfare for no compensating gain.
4. So it is irrational for Bob to skip the gallery.

I think that argument is wrong. Bob's case is rather unlike Ricky's. There is a sense in which Bob might be undermining his own welfare in skipping the gallery. But it is not the relevant sense. We can distinguish the two senses making the scope of various operators explicit. The first of these claims is plausibly true; the second is false.

- Bob's welfare is such that it is irrational for him to do something that might undermine it for no compensating gain.
- It is irrational for Bob to do something that might undermine his welfare, whatever that turns out to be, for no compensating gain.

If welfare turns out to be health, happiness and learning, then the first claim says that it is irrational to risk undermining your health, happiness and learning for no compensating gain. And that is, I think, right. But the second claim says that for any thing, if that thing might be welfare, and an action might undermine it, it is irrational to perform the action without a compensating gain. That's a much stronger, and a much less plausible, claim.

Importantly, Bob's 'Might' Argument doesn't go through with the first claim. Given that appreciation of beauty is not directly a component of welfare, and that the various channels through which appreciating beauty might lead to an increase in welfare are blocked for Bob, there is no chance that going to the gallery will increase his actual welfare. Going to the gallery will increase something, namely his appreciation of beauty, that is for all Bob knows part of welfare. But that's not the same thing, and it isn't relevant to rationality.

One caveat to all this. On some theories of welfare, it will not be obvious that even the first claim is right. Consider a view (standard among economists) that welfare is preference satisfaction. Now you might think that even the first claim is ambiguous, between a claim that one's preferences are such that it is irrational to undermine them (plausibly true), and a claim that it is irrational to undermine one's preference satisfaction. The latter claim is not true. If someone offers me a pill that will make me have preferences for things that are sure to come out true (I want the USA to be more populous than Monaco; etc.), it is rational to refuse it. And that's true even though taking the pill will ensure that I do well by preference satisfaction. The point is that taking the pill does not, as things stand, satisfy my preferences. If I prefer X to Y, I should aim to

bring about X. But I shouldn't aim to bring about a state of having satisfied preferences; that could lead to rather perverse behaviour, like taking this pill.

4 Duelling Analogies

Here's how I see the six cases we've discussed so far fitting together.

	Factual Uncertainty	Normative Uncertainty
Prudential Risk	Ricky	Bob Bruce
Moral Risk	Carla	Agnes Martha

On the left-hand column, we have agents who are uncertain about a simple factual question; is this syrup sweetener or arsenic? On the right-hand column, we have agents who are uncertain about a question about the nature of value; does the decision I'm facing right now have serious evaluative consequences?

It's even easier to see what is separating the rows. Ricky, Bob and Bruce face questions that, in the first instance, just concern their own welfare. Carla, Agnes and Martha face questions that concern the morality of their actions. I don't mean to say that there's a hard line between these two. Perhaps being moral is an important part of the good life. And perhaps one has a moral duty to live well. I'm a little doubtful on both scores actually. But even if the questions bleed into each other in one or other way, we can separate questions that are in the first instance about the agent's own welfare from questions that bear directly on the morality of the agent. (Recognising, as always, that there will be borderline cases.) And that's how we've split the rows.

One way to motivate the 'Might' Argument is to stress the analogy between Carla and Agnes/Martha. After all, both of them risk killing someone (or something) stasued if they act in a certain way. But once we look at the table more broadly, it is easy to see why we should resist the analogy between Carla and Agnes/Martha. The analogy between Bob/Bruce and Agnes/Martha is much stronger. We can see that by thinking about their motivations.

Why would Bruce go to the gallery? Not for pleasure; he'll get more pleasure out of playing video games with his friends. Not for the educational value; he won't learn more by looking at these kind of paintings again. His only reason for going is that he thinks it might increase his welfare. That is, he can only be motivated to go if he is motivated to care about welfare as such, and not about the things that make up welfare. There is something perverse about this motivation. It is healthy and natural to want the things that make up a good life. It is less healthy, and less natural, to directly desire a good life whatever that may be.

Now think about Martha. Why should she turn down the steak? Not because she values the interests of the cow over her dining. She does not. And not because she

should have that value. By hypothesis, she need not do so. (Remember we're only interested in replying to people who argue from The 'Might' Argument to vegetarianism; if you think there's a direct argument that Martha should value the cow so highly that she doesn't eat meat, that's a different debate.) Rather, she has to care about morality as such. And that seems wrong.

The argument I'm making here owes a lot to a similar argument offered for a somewhat different conclusion by Michael Smith (1994). He compared the person who desires to do what is actually right, as he put it, desires the right *de re*, with the person who desires to do what is right whatever that turns out to be, as he put it, desires the right *de dicto*.

Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read *de dicto* and not *de re*. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue. (Smith 1994, 75)

I think that's all true. A good person will dive into a river to rescue a drowning child. (Assuming that is that it is safe enough to do so; it's wrong to create more rescue work for onlookers.) And she won't do so because it's the right thing to do. She'll do it because there's a child who needs to be rescued, and that child is valuable.

The analogy with the welfare case strengthens this conclusion. The rational person values their health, happiness and friendships (and whatever goes into the actual list of things that constitute welfare.). They don't simply value their welfare, and desire to increase it. That's why it would be perverse for Bruce to go to the gallery. He would only go if he had a strange motivation. And it is why it would be perverse for Martha to turn down the steak. To do so she would have to care about morality, whatever it is, not about the list of things that Smith rightly says a good person will care about.

5 An Alternative Analogy

Moller offers the following analogy to back up something like the 'Might' Argument.⁵

Suppose Frank is the dean of a large medical school. Because his work often involves ethical complications touching on issues like medical experimentation and intellectual property, Frank has an ethical advisory committee consisting of 10 members that helps him make difficult decisions. One day Frank must decide whether to pursue important research for the company in one of two ways: plan A and plan B would both accomplish

⁵Though note that Moller's own position is more moderate than what the 'Might' Argument suggests; he thinks moral risk should play a role in reasoning, but not necessarily so strong a role as to make the 'Might' Argument go through. I'm advocating what he calls the "extreme view, we never need to take moral risk into account; it is always permissible to take moral risks." (435.)

the necessary research, and seem to differ only to the trivial extent that plan A would involve slightly less paperwork for Frank. But then Frank consults the ethics committee, which tells him that although everyone on the committee is absolutely convinced that plan B is morally permissible, a significant minority - four of the members - feel that plan A is a moral catastrophe. So the majority of the committee thinks that the evidence favors believing that both plans are permissible, but a significant minority is confident that one of the plans would be a moral abomination, and there are practically no costs attached to avoiding that possibility. Let's assume that Frank himself cannot investigate the moral issues involved - doing so would involve neglecting his other responsibilities. Let's also assume that Frank generally trusts the members of the committee and has no special reason to disregard certain members' opinions. Suppose that Frank decides to go ahead with plan A, which creates slightly less paperwork for him, even though, as he acknowledges, there seems to be a pretty significant chance that enacting that plan will result in doing something very deeply wrong and he has a virtually cost-free alternative. (436)

The intuitions are supposed to be that this is a very bad thing for Frank to do, and that this illustrates that there's something very wrong with ignoring moral risk. But once we fill in the details of the case, it is clear that this can't be the right diagnosis.

The first thing to note is that there is something special about decision making as the head of an organization. Frank doesn't just have a duty to do what he thinks is best. He has a duty to reflect his school's policies and viewpoints. A dean is not a dictator, not even an enlightened, benevolent one. Not considering an advisory committee's report is bad practice qua dean of the medical school, whether or not Frank's own decisions should be guided by moral risk.

We aren't told whether A or B are moral catastrophes. If B is a moral catastrophe, and A isn't, there's something good about what Frank does. Of course, he does it for the wrong reasons, and that might undercut our admiration of him. But it does seem relevant to our assessment to know whether A or B are actually permissible.

Assuming that B is actually permissible, the most natural reading of the case is that Frank shouldn't do A. Or, at least, that he shouldn't do A for this reason. But that doesn't mean he should be sensitive to moral risk. Unless the four members who think that A is a moral catastrophe are crazy, there must be some non-moral facts that make A morally risky. If Frank doesn't know what those facts are, then he isn't just making a decision under moral risk, he's making a decision involving physical risk. And that's clearly a bad thing to do.

If Frank does know why the committee members think that the plan is a moral catastrophe, his action is worse. Authorising a particular kind of medical experimentation, when you know what effects it will have on people, and where intelligent people think this is morally impermissible, on the basis of convenience seems to show a striking lack of character and judgment. Even if Frank doesn't have the time to work through all the ins and outs of the case, it doesn't follow that it is permissible to make decisions based

on convenience, rather than based on some (probably incomplete) assessment of the costs and benefits of the program.

But having said all that, there's one variant of this case, perhaps somewhat implausible, where it doesn't seem that Frank should listen to the committee at all. Assume that both Frank and the committee have a fairly thick understanding of what's involved in doing A and B. They know which actions maximise expected utility, they know that which acts are consistent with the categorical imperative, they know which people affected by the acts would be entitled to complain about our performance, or non-performance, of each act, they know which acts are such that everyone could rationally will it to be true that everyone believes those acts to be morally permitted, and so on. What they disagree about is what rightness and wrongness consist in. What's common knowledge between Frank, the majority and the minority is that both A and B pass all these tests, with one exception: A is not consistent with the categorical imperative. And the minority members of the committee are committed Kantians, who think that they have a response to the best recent anti-Kantian arguments.

It seems to me, intuitively, that this shouldn't matter one whit. I think the extreme view I'm defending in this paper is not, in general, intuitive. But it is worth noting how counterintuitive the opposing view is in this extreme case. A moral agent simply won't care what the latest journal articles have been saying about the relative importance of Kant's formulation of the categorical imperative versus either contemporary variants or approaches from very different traditions. It's possible (though personally I doubt it), that learning of an action that it violates the categorical imperative would be relevant to one's motivations. It's not possible that learning that some people you admire think the categorical imperative is central to morality could change one's motivation to perform, or not perform, actions one knew all along violated the categorical imperative. At least that's not possible without falling into the bad kind of moral fetishism that Smith rightly decries.

So here's my general response to analogies of this kind, one that shouldn't be surprising given the previous sections. Assuming the minority committee members are rational, either they know some facts about the impacts of A and B that Frank is unaware of, or they hold some philosophical theory that Frank doesn't. If it's the former, Frank should take their concerns into account; but that's not because he should be sensitive to moral risk, it's because he should be sensitive to non-moral risk. If it's the latter, Frank shouldn't take their concerns into account; that would be moral fetishism.

6 Objections and Replies

I've discussed this paper with many people, and they almost all have objections. I'm going to respond to some of the most pressing, and end with three objections that I don't have a particularly satisfying response to. The most important objection, from my perspective, is the second; it's what most closely links the discussion of this paper to the broader issues about normative externalism that I find most fascinating.

Objection: All you've shown so far is that moral recklessness isn't objectively wrong.

But that's trivial. There's a sense in which ordinary recklessness isn't objectively wrong either. What matters is that both are subjectively wrong, where this tracks what the agent believes.

Reply: Distinguish between two things: doing things that produce bad outcomes, and doing the wrong thing. Unless you are sure that actualist consequentialism is a conceptual truth, this is a conceptually coherent distinction. Among actions that produce bad outcomes, there are easily detectable distinctions we draw that seem to track whether the actions are wrong.

In the paper so far I've usually been focussed on people who are almost certain of the truth. But let's change tack for a minute and look at people who have catastrophically false beliefs. In particular, consider Hannah and Hannibal. (I'm taking the Hannibal example from work by Elizabeth Harman (2011), who uses it for a related purpose.)

Hannah takes her spouse out for what is meant to be a pleasant anniversary dinner. It's a nice restaurant, and there's no reason to think anything will go wrong. But the restaurant gets bad supplies that day, and Hannah's spouse gets very sick as a consequence of going there.

Hannibal is a 1950s father with sexist attitudes that were sadly typical. He has a son and a daughter, and makes sure to put together a good college savings fund for his son, but does not do the same for his daughter. Indeed, if he had tried to do the same for his daughter, he would not have been able to support his son as well as he actually did. As a consequence, his daughter cannot afford to go to college.

Hannah was mistaken about a matter of fact; whether the food at the restaurant was safe. Hannibal was mistaken about a moral matter; whether one should treat one's sons and daughters equally. Now consider what happens when both see the error of their ways. Hannah should feel bad for her spouse, but there is no need for any kind of self-reproach. It's hard to imagine she would feel ashamed for what she did. And there's no obligation for her to feel guilty, though it's easier to imagine she would feel some guilt. Hannibal, on the other hand, should feel both ashamed and guilty. And I think it's natural that a father who realised too late that he had been guilty of this kind of sexism would in fact feel the shame and guilt he should feel. The fact that his earlier sexist attitudes were widely shared, and firmly and sincerely held, simply seems irrelevant here.

The simplest explanation of this emotional difference is that what Hannibal does is, in an important sense, wrong, and what Hannah does is not wrong. But the wrongness at issue is missing from the objective/subjective distinction the objector here makes. Both Hannah and Hannibal do things that make things objectively worse. Both Hannah and Hannibal do things that are good given their beliefs at the time they act. Yet there is a distinction between them. It's this distinction that the normative externalist wants to stress. There's a normative status that is not wholly objective, insofar as it doesn't reproach Hannah, but not wholly subjective, insofar as it does reproach Hannibal.

Objection: But still, we need a standard that can guide the agent, that an agent can live by. Do the right thing, whatever it turns out to be, is not such a standard. And

what motivates internalism is the thought that this kind of agent-centred norm is most important.

Reply: If this is the motivation for internalism, it is vulnerable to a nasty regress. The problem is that internalists disagree amongst themselves, and there is no internalist-friendly way to resolve the disagreement.⁶ (Much of what I say here draws on arguments that Elizabeth Harman (2015) makes about the nature of internalist norms.)

The examples that illustrate this point are a little convoluted, so I'll just state one example schematically to make the point. And I'll put numerical values on options because it is hard to state the internalist views without doing this.

An agent faces a choice between four options: A, B, C and D. Option A is the right option, both in the sense that the externalist will praise people who take it and criticise others, and in the sense that a fully informed internalist would do A. But our agent is, sadly, not fully informed. She thinks A is a completely horrible thing to do. Her credences are split over three moral theories, X, Y and Z, with credence 0.5 in X, 0.1 in Y, and 0.4 in Z. The moral values of each action according to each moral theory are given by this table. (Higher values are better; non-negative values are for actions that are permissible according to the theory.)

Table 2: The moral payout table for four options

	X	Y	Z
B	0	0	-20
C	0	-30	-10
D	-1	-5	0

So the probability, according to the agent, that each action is permissible is 0.6 for B, 0.5 for C and 0.4 for D. The expected moral value of each action is -8 for B, -7 for C, and -2 for D.

Our agent at this stage is a bit confused. And reading some philosophy doesn't help. She reads Ted Lockhart (2000) saying that what she should do is the thing that is most probably permissible. And she reads Andrew Sepielli (2009) saying that what she should do is the thing that maximises expected moral value. But these pieces of advice pull in opposite directions. She could try and come up with a theory of how to resolve the tension, but that is just as hard as resolving the dispute between Lockhart and Sepielli in the first place. She eventually settles on the rule *Don't do what any plausible meta-theory says is the worst thing to do*. Since Lockhart says D is the worst thing to do (having the lowest probability of permissibility), and Sepielli says that B is the worst thing to do (having the lowest expected moral value), she does C.

Here's the lesson of this little parable. There is a worry that externalism is not sufficiently action guiding, and can't be a norm that agents can live by. But any philosophi-

⁶In Weatherson (2013) I make a similar objection to normative internalism in epistemology. It's this point of connection that's made me focus on *normative* internalism and externalism, not *moral* internalism and externalism. The issues in ethics and in epistemology are very closely connected here.

cal theory whatsoever is going to have to say something about how to judge agents who ascribe some credence to a rival theory. That's true whether the theory is the first-order theory that Jeremy Bentham offers, or the second-order theory that Andrew Sepielli offers. Once you're in the business of theorising at all, you're going to impose an external standard on an agent, one that an agent may, in good faith and something like good conscience, sincerely reject. The externalist says that it's better to have that standard be one concerned with what is genuinely valuable in the world, rather than a technical standard about resolving moral uncertainty. But every theorist has to be a little bit externalist; the objector who searches for a thoroughly subjective standard is going to end up like Ponce de Leon.

Objection: You've focussed on the case where Martha is almost sure that meat-eating is permissible. What do we say about the person who is almost sure that meat-eating is impermissible, eats meat anyway, and gets lucky, because they are in a world where it is permissible? The normative externalist says that they are beyond reproach, but something seems wrong here.

Reply: The externalist is only committed to the view that the most important evaluative concepts are independent of the agent's beliefs. There is something rather simple to say about this person; they are a hypocrite.

Objection: Wait a minute! We wanted something reproachful to say about this person. But all you've said is that they are a hypocrite, by which you presumably mean they don't act in accord with their beliefs about what's valuable. And Huckleberry Finn is a hypocrite in that sense, but also beyond reproach.

Reply: Good point, but I think we can still say something. Huckleberry Finn acts against what he believes to be most valuable in order to preserve a great good: Jim's freedom. Our imagined meat-eater acts against what he believes to be most valuable in order to get a tastier lunch. Someone who will do what they believe to be wrong in order to produce a gain which is both trivial, and entirely accrues to them, reveals a bad character. The gain that Huckleberry Finn's actions produce, note, are neither trivial nor selfish, and that's why his actions do not indicate a character defect. But giving up on morality for a trivial, selfish gain is a sign that things will go very badly wrong, very soon.⁷

Objection: How can you even acknowledge such a thing as hypocrisy? Isn't the positing of such a norm vulnerable to the same regress arguments as you've run against the internalist?

Reply: No, because we can be an externalist about what is and is not hypocritical. We can, at least in theory, imagine these two cases. The first case is a person whose beliefs, credences and values indicate that the best thing to do is B, but who thinks the best thing to do given those beliefs, credences and values is C. They do C. They are hypocritical, although they (falsely) do not believe they are. The second case is a person who is exactly like this, except they do B. They are not acting hypocritically. Or,

⁷The Huckleberry Finn case has been discussed extensively by Nomy Arpaly and Timothy Schroeder (Arpaly 2002, 2003; Arpaly and Schroeder 1999, 2014), and I'm relying heavily on their analysis of the case in what I say here and elsewhere about Huckleberry Finn. More generally, the picture I'm assuming of moral motivation owes a lot to those works.

at least, they are not a first-order hypocrite. Perhaps we can recognise a distinct state of second-order hypocrisy, and say that they fall under it. And you can imagine even higher-orders. The externalist can say all of these exist. They aren't the worst offences ever, but it is coherent to posit all of them.

Objection: Once you recognise hypocrisy, there is a way to reinstate the 'Might' Argument. Martha and Agnes are hypocrites. They shouldn't be hypocrites. So they shouldn't eat meat, or have an abortion.

Reply: I simply deny that they are hypocrites. Compare these three statuses.

- Doing that which you disvalue.
- Doing that which you believe to be less valuable.
- Doing that which you have some credence is less valuable.

The first is clearly hypocrisy, and the second seems similar. But there's no reason to say the third is hypocritical. The following example, closely modelled on one offered by Lara Buchak (2014) makes this point.

Annie values her close relationship with her brother Jack. One day, she receives some evidence that marginally raises her credence that Jack did something horrible. She is pretty sure Jack is innocent, but her credence in his guilt does rise a notch. Still, Annie values her relationship with Jack just as much as she did before. If Jack did the horrible thing, she would not value the relationship. But getting some (almost surely misleading) evidence that Jack did something horrible does not change her values at all.

The lesson here is that credences about what is valuable can quite coherently float free from valuing. There is a tricky question about what happens to beliefs about what is valuable in these cases. Buchak thinks they should go with valuing, and this is a problem for theories that reduce credence to belief. I don't agree with this extension of her argument, but I certainly agree that small changes in credence about what is valuable need not, and often should not, change what one values.

Objection: The externalist can't explain why moral ignorance exculpates.

Reply: The short reply is that, following for example Elizabeth Harman (2011), I don't think moral ignorance does exculpate. But the longer reply is that the internalist can't explain why moral ignorance is at best an excuse, not a defence, and why it only works in special circumstances.

We already saw one distinctive aspect of moral ignorance above, in the Hannibal example. Hannibal should feel ashamed, and guilty, about what he did. That's because even if he had an excuse, he did the wrong thing. And this doesn't just mean he made the world worse. This notion of wrongness is an externalist one, even if we allow an internalist friendly excuse for the wrong action.

But when we turn to classic defenders of the idea that moral ignorance can be exculpatory, such as Susan Wolf (1980) and Cheshire Calhoun (1989), we see that it is meant to be an excuse with a very limited scope. And whether the circumstances are such as to furnish this excuse will not always be clear to the wrong-doer. (Indeed, it might be that they are not, and could not, be clear.) So even if moral ignorance was exculpatory,

this wouldn't be much help to the internalist. Since on everyone's view some moral ignorance is blameworthy, and the factors that may make moral ignorance an excuse are external to the agent, only the externalist can offer a plausible theory on which moral ignorance is exculpatory.

Objection: Even if it is fetishistic to be motivated by the good as such, this doesn't extend to thick moral properties. Indeed, the quote from Smith you use explicitly contrasts the thinnest of moral properties with ever so slightly thicker ones. So your objections to arguments from moral uncertainty don't extend to arguments from what we might call virtue uncertainty.

Reply: I agree with this. Here are some things that seem like be non-fetishistic motivations to avoid doing action A.

- It would be cowardly to do A.
- Doing A would be free-riding.
- I would not appreciate if others did A-like actions that could disadvantage me.

The objector draws attention to the distinction between thick and thin moral properties, and I think that's the right way to highlight what's at issue here. But note how thin these are getting. I'm conceding that the fact that something violates the Golden Rule could be a motivation, as could the fact that it violates the categorical imperative.⁸ What I deny is that the wrongness of the action could be an extra motivation over and above these. This was the point of the discussion of Moller's executive in the previous section.

For each of these motivations, there are cases where the risk of violating the relevant standard can be motivating. So one might not do something because there is a risk that it would be cowardly, or free-riding, or violate the Golden Rule or categorical imperative. I don't mean to object to any argument along these lines.

Objection: Now you've conceded that a version of the 'Might' Argument can work. After all, there are vices that might be manifest by eating meat or having an abortion.

Reply: True, but the fact that some action might manifest a vice can hardly be a decisive consideration against doing it. If the vice in question is relatively small, or the chance of manifesting it is relatively small, it is easy to see how this kind of consideration could be overridden.

For instance, imagine an argument for vegetarianism as follows. Eating meat you haven't killed yourself might be cowardly. It certainly isn't obvious that letting someone else do the dirty work isn't a manifestation of cowardice. So that's a reason to not eat meat. I can grant it is a reason while thinking that (a) this kind of cowardice isn't a particularly heinous vice, and (b) it isn't that likely that meat eating is really cowardly in this way, so the reason is a relatively weak one, that can easily be overridden.

⁸To be clear, I'm conceding that these motivations are consistent with the argument of the paper. My own view is that while realising that something violates the Golden Rule could be a motivation, as is evident from how we teach morality to children, realising that it violates the categorical imperative should not be motivating. But the argument of the paper doesn't turn on my quirky views here. What matters is that we distinguish wrongness itself from properties like harming another person, not what other properties we group in with wrongness.

But the concession I want to make is that there could be an argument along these lines that works. In earlier presentations of this paper, I'd tried to extend my argument to respond to the arguments Alex Guerrero (2007) makes for vegetarianism. But I'm no longer sure that was a good idea. But I think Guerrero's arguments can be understood in such a way that they rely only on the idea that we shouldn't risk instantiating certain particular vices. And I don't have a systematic objection to every argument of this form. After all, I do think we have a reason to avoid running a risk of being free-riders, or cowards, even if the action under consideration would not be cowardly, or an act of free-riding.

Objection: Even without getting into debates about *moral* uncertainty, there are other uncertainty arguments against meat eating or abortion. There is some probability that cows or foetuses have souls, and it is a very serious harm to kill something that has a soul.

Reply: Nothing I say here helps respond to this argument. If one thinks that what's wrong with killing is that it kills a soul, thinks that there's a non-trivial chance that cows or foetuses have souls, and eats meat or has an abortion anyway, then one really is being immoral. Whether this should be called recklessness is tricky, since one could understand 'recklessness' as being concerned only with risks that are in a certain sense objective. But it certainly seems that such a person would be morally on a par with the people I've said are immoral in virtue of the risks they pose to others. It's an empirical question, and one I don't have any good evidence about, whether arguments from uncertainty about abortion and meat eating primarily concern uncertainty about facts, as this objection suggests, uncertainty about virtues (broadly construed) as the previous objection suggests, or uncertainty about right and wrong.

Objection: It may be wrong to be only concerned with right and wrong, but it isn't wrong to have this be one of your considerations.

Reply: I don't think you get the 'Might' Argument to work unless concern with right and wrong, whatever they turn out to be, are the only considerations. Assume that they are only one consideration among many. Then even if they point in one direction, they may be overridden by the other considerations. And if the 'Might' Argument doesn't work, then normative internalism, in its strongest forms, is false. So I really only need to appeal to the plausible view that right and wrong as such shouldn't be our only motivations to get the conclusions I want.

But actually I think the stronger, *prima facie* implausible, view is true: rightness and wrongness as such shouldn't even be part of our motivation. My reasons for thinking this are related to my responses to the next three objections. Unfortunately, these are the least developed, and least satisfying, of the responses I'll offer. But I'll conclude with them to leave you with a sense of where I think the debate is at, and what I think future research could assist with.

Objection: Here's one occasion where we do seem motivated by the good as such, or by welfare as such – when we're doing moral or prudential reflection. Sometimes we stop and think, What would be the best thing to do in a certain kind of case? In philosophy departments, people might do that solely because they're interested in the answer.

But most people will think that these projects have some practical consequences. And the strong form of Smith's fetishism objection that you're relying on can't explain why this is a good practice.

Reply: I agree this is a good practice. But I think it is consistent with what I've said so far. Start with an observation also by Michael Smith, that moral inquiry has "a certain characteristic coherentist form" (Smith 1994, 40–41). I think (not originally) that this is because we're not trying to figure out something about this magical thing, the good, but rather because we're trying to systematise and where necessary reconcile our values. When we're doing moral philosophy, we're often doing work that more at the systematising end, trying to figure out whether seemingly disparate values have a common core. When we're trying to figure out what is right in the context of deciding what to do, we're often trying to reconcile, where possible, conflicting values. But as long as we accept that there are genuinely plural values, both in moral and prudential reasoning, we shouldn't think that a desire to determine what is right is driven by a motivation to do the right thing, or to live a good life, as such.

Objection: Sometimes people act from moral conscience. At least by their own account, they do something that involves no small amount of personal sacrifice because it is the right thing to do. And, at least some of the time, these people are highly praiseworthy. The strong version of the fetishism objection you're using can't account for this.

Objection: So I have to bite some bullets here. I have to offer a slightly unnatural reformulation of these cases. In particular, in cases where someone acts from conscience, I have to say that there is something they value greatly, and they are acting on that value. What the value is will depend on the case. It might be welfare, or freedom, or keeping promises, or justice. It might even, and this is the version of the case that's trickiest for me, be a value they can't clearly articulate. A person can know something is the right thing to do and not be in any position to say why it is the right thing to do. And they may do it, even at great sacrifice. I think I'm required to say here that their motivation is the feature of the act that makes it right, not the rightness of the act. That's not optimal, especially since it isn't how the agent themselves would describe the motivation. But I don't think we should assume that agents have perfect access to their own motivations.

I take myself to be here largely in agreement with a line suggested by Sigrún Svavarsdóttir (1999) when she says, in defence of an externalist theory of moral motivation.

The externalist account I propose does not ascribe to the good person a particular concern with doing the right thing. Rather it ascribes to him a more general concern with doing what is morally valuable or required, when that might include what is just, fair, honest, etc. (Svavarsdóttir 1999, 197–98)

There are two points here that are particularly relevant to the current project. The good person has a plurality of motivations, not just one. And the fetishism argument really has a very narrow application: it really only works against theories which say goodness is a matter of having the thinnest of possible moral motivations. It's odd to be solely

concerned with doing the right thing as such. (It's even odd, I say, to have this as one of your concerns, though that's not central to my argument.) It's not odd to have fairness as one of one's concerns, even an important one. Svavarsdóttir suggests that once the range of the fetishism argument is restricted in this way, it can't do the work that Smith needs it to do in his attack on motivational externalism. I don't need to take a stand on this, since I'm not taking sides in the debate between motivational externalists and internalists. All I need is that Smith's objection to fetishism can work, as long as it is suitably restricted.

Objection: Is there any coherent meta-ethical view that can licence all the moves you've made? On the one hand, normative claims must be distinctive enough that uncertainty about them has a very different effect on deliberation and motivation than everyday factual claims. On the other hand, your externalism is the view that the moral facts matter more than anyone's (reasonable) beliefs about the moral facts. The first consideration suggests a strong kind of moral anti-realism, where moral claims are different in kind to factual claims. But the second suggests a strong kind of moral realism, where there are these wonderful moral facts around to do the work that reasonable moral beliefs cannot do. Is this even consistent? And if it is, is there a meta-ethical view we should want to hold consistent with all of it?

Reply: The inconsistency charge isn't, I think, too hard to meet. As long as the 'facts' that I talk about when I say the moral facts matter are construed in an extremely deflationary way, then I'm not being inconsistent. Any kind of sophisticated expressivist or quasi-realist view that allows you to talk about moral facts, while perhaps not meaning quite the same thing by 'fact' as a realist does, will be consistent with everything I've said.

The second challenge is harder, and I don't know that I have a good response. I would like to make the theory I've presented here consistent with a fairly thoroughgoing moral realism, and I'm not sure that's possible. (I'd like to do that simply because I don't want the fate of the theory tied up with contentious issues in meta-ethics.) I think the way to make the view consistent with this kind of realism is to defend the view that neither the metaphysical status of a truth (as necessary or contingent, analytic or synthetic, and so on) has very little to do with its appropriate role in deliberation or evaluation. But defending that, and showing how it suffices to make moral cognitivism consistent with the view I'm describing, is more than I know how to do now.

References

- Arpaly, Nomy. 2002. "Moral Worth." *Journal of Philosophy* 99 (5): 223–45. doi: 10.2307/3655647.
- . 2003. *Unprincipled Virtue*. Oxford: Oxford University Press.
- Arpaly, Nomy, and Timothy Schroeder. 1999. "Praise, Blame and the Whole Self." *Philosophical Studies* 93 (2): 161–88. doi: 10.1023/A:1004222928272.
- . 2014. *In Praise of Desire*. Oxford: Oxford University Press.
- Buchak, Lara. 2014. "Belief, Credence and Norms." *Philosophical Studies* 169 (2):

- 285–311. doi: 10.1007/s11098-013-0182-y.
- Calhoun, Cheshire. 1989. “Responsibility and Reproach.” *Ethics* 99 (2): 389–406. doi: 10.1086/293071.
- Finnis, John. 2011. *Natural Law and Natural Rights*. Second. Oxford: Oxford University Press.
- Fodor, Jerry. 2000. “It’s All in the Mind: Noam Chomsky and the Arguments for Internalism.” *Times Literary Supplement* 23 June: 3–4.
- Guerrero, Alexander. 2007. “Don’t Know, Don’t Kill: Moral Ignorance, Culpability and Caution.” *Philosophical Studies* 136 (1): 59–97. doi: 10.1007/s11098-007-9143-7.
- Harman, Elizabeth. 2011. “Does Moral Ignorance Exculpate?” *Ratio* 24 (4): 443–68. doi: 10.1111/j.1467-9329.2011.00511.x.
- . 2015. “The Irrelevance of Moral Uncertainty.” *Oxford Studies in Metaethics* 10: 53–79. doi: 10.1093/acprof:oso/9780198738695.003.0003.
- Lockhart, Ted. 2000. *Moral Uncertainty and Its Consequences*. Oxford University Press.
- Malmgren, Anna-Sara. 2011. “Rationalism and the Content of Intuitive Judgements.” *Mind* 120 (478): 263–327. doi: 10.1093/mind/fzr039.
- Moller, D. 2011. “Abortion and Moral Risk.” *Philosophy* 86 (3): 425–43. doi: 10.1017/S0031819111000222.
- Sepielli, Andrew. 2009. “What to Do When You Don’t Know What to Do.” *Oxford Studies in Metaethics* 4: 5–28.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Svavarsdóttir, Sigrún. 1999. “Moral Cognition and Motivation.” *Philosophical Review* 108 (2): 161–219. doi: 10.2307/2998300.
- Weatherson, Brian. 2013. “Disagreements, Philosophical and Otherwise.” In *The Epistemology of Disagreement: New Essays*, edited by David Christensen and Jennifer Lackey, 54–73. Oxford: Oxford University Press.
- Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Blackwell.
- Wolf, Susan. 1980. “Asymmetrical Freedom.” *Journal of Philosophy* 77 (3): 151–66. doi: 10.2307/2025667.

Published in *Philosophical Studies*, 2014, pp. 425–431.