

A DEFENSE OF MORAL DEFERENCE

Author(s): DAVID ENOCH

Source: *The Journal of Philosophy*, Vol. 111, No. 5 (MAY 2014), pp. 229-258

Published by: Journal of Philosophy, Inc.

Stable URL: <https://www.jstor.org/stable/43820838>

Accessed: 05-05-2020 16:51 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Journal of Philosophy, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*

THE JOURNAL OF PHILOSOPHY

VOLUME CXI, NO. 5, MAY 2014

A DEFENSE OF MORAL DEFERENCE*

There seems to be something fishy about the idea of moral expertise and indeed testimony, the idea of one person forming her moral judgments merely based on the view of another, and then acting on them. But it is not immediately clear how serious the problem with moral expertise is, or how it is best understood. In making progress on these matters, it will prove useful, I think, to start with two examples.

The first comes from Bernard Williams:

There are, notoriously, no ethical experts....Anyone who is tempted to take up the idea of there being a theoretical science of ethics should be discouraged by reflecting on what would be involved in taking seriously the idea that there were experts in it. It would imply, for instance,

*For helpful conversations and comments on previous versions, I thank Sam Asarnow, Kent Bach, Dorit Barhana-Lorand, Hagit Benbaji, Jonathan Berg, William Crouch, Raff Donelson, Dalia Drei, Nir Eyal, Yuval Eylon, Amihud Gilead, John Goldberg, David Hills, Karen Jones, Alon Harel, Liz Harman, Meir Hemmo, Ulrike Heuer, Allison Hills, Frank Jackson, Frances Kamm, Paul Katsafanas, Arnon Keren, Claire Kirwin, Iddo Landau, Gerald Lang, Alex Langlais, Noa Leibowitz, Brian Leiter, Nethanel Lipshitz, Don Loeb, Peter Ludlow, David Lyons, Don Maier, Sarah McGrath, Barak Medina, Ariel Meirav, Aaron Meskin, Martha Nussbaum, Sven Nyholm, Mike Otsuka, David Owens, Philip Pettit, Christian Piller, David Plunkett, Jim Pryor, Joseph Raz, Hili Razinsky, Grant Rozeboom, Ronni Sadovsk, Tamar Schapiro, Re'em Segev, Assaf Sharon, Ken Simmons, Paulina Sliwa, Saul Smilansky, Michael Smith, Levi Spectre, Gopal Sreenivasan, Daniel Star, Daniel Statman, Sharon Street, Ken Taylor, Sergio Tenenbaum, Pekka Väyrynen, Bruno Verbeek, Ruth Weintraub, Robbie Williams, and two readers for this JOURNAL. I wrote the paper while being a fellow at the Institute for Advanced Studies at the Hebrew University of Jerusalem, where I also presented the paper (in our research group's conference) for the first time. I thank the Institute for its support. I presented earlier versions of the paper in colloquia and workshops at Boston University, the University of Chicago, the Hebrew University in Jerusalem, the Haifa University, Stanford University, Princeton, King's College, Harvard, and Leeds. I thank the participants for the discussion that followed. This research was also supported by the Israel Science Foundation (grant no. 136/09).

0022-362X/14/1105/229-258

© 2014 The Journal of Philosophy, Inc.

that a student who had not followed the professor's reasoning but had understood his moral conclusion might have some reason, on the strength of his professorial authority, to accept it....These Platonic implications are presumably not accepted by anyone.¹

This paper will constitute a refutation (pretty much) of this last sentence. But it cannot be denied that Williams is on to something here, indeed, something important about moral deference and expertise, and presumably about moral testimony as well.

Here is the second example: whenever yet another violent interaction erupts in the Middle East, Israel sees (like most other places, I think) a rallying-around-the-flag initial reaction of its public opinion. It takes some time before more critical views are voiced and allowed to be heard more clearly and loudly. In the past, I have noticed that I too seem to go with the flow on these matters. In the first few days—weeks, sometimes—I tend to see the war as in-principle justified. But often, within a couple of weeks, I come to see the war differently, indeed much more critically. And I then come to view with great shame and embarrassment my views from just a couple of weeks earlier, thinking to myself things like, “How could I have missed how wrong this was?” But I have this colleague—Alon, let us call him—who is different. For he—I noticed—voices the moral criticism I come to endorse as the war proceeds much earlier, indeed, from its very outbreak. And now—you guessed it—another armed conflict erupts. I think about it, familiarize myself as much as possible with the relevant facts, and it seems to me that we are by and large in the right. But I talk to Alon, who once again tells me how wrong this whole affair is. Let us further suppose that I do not have to worry just about forming my opinion. Rather, I have to *do* something, something that will have an effect in the real world, on the lives of real people—say, I have to cast a deciding vote on funding for the war, or some such. Now, war issues are hard. Innocent lives are on both sides. Making the wrong decision—*any* wrong decision, either supporting an unjustifiable war or objecting to a justified war—may cause the unjustifiable deaths of innocents. Erring on the side of safety is not a strategy I could implement here (or so, at least, are things in my stipulated example), and so not a way of circumventing the need to make a more case-specific decision. What should I believe about the moral

¹ Bernard Williams, *Making Sense of Humanity* (New York: Cambridge, 1995), p. 205. I take this quote from Alison Hills, “Moral Testimony and Moral Epistemology,” *Ethics*, cxx, 1 (October 2009): 94–127, at p. 96. Williams’s aims—though related—are not precisely the ones for which I (following Hills) use him. I am not invested, of course, in any interpretative claim about Williams.

status of the war? How should I vote? Here, it seems to me, there is considerable pressure to say that given that Alon—as I know, or anyway wholeheartedly believe—was much more often right than I was about such matters in the past, I should believe that the war cannot be morally defended, and act accordingly.

These examples—like all others—are dangerous. The cases are not fully described, and they are easily abused. So caution should be taken in relying on them (and more details will be stipulated later on). For now, though, we can think of these two examples as highlighting the problematic nature of moral deference and expertise (and maybe also testimony). What we want from a philosophical discussion of moral expertise is an accommodation of both examples, to the extent possible.

In this paper I will vindicate moral deference and expertise, arguing that sometimes we should defer—we are morally required to, pretty much—to the moral expert. My discussion will be based on placing the current debate in a context in which as far as I know it has not been placed:² that of the moral norms guiding the appropriate response to moral uncertainty.³ Deferring to an expert is sometimes, I will argue, the only morally acceptable way of responding to moral uncertainty.

In the next section I do some stage setting. In section II I offer some diagnostic observations, checking to see the scope of the intuitions pulling in the two directions here. In section III I present my vindication of moral deference. This vindication is based on a moral requirement not to subject others to a higher-than-needed risk of being wronged (by the relevant agent, when all other things are equal). My vindication requires that we straightforwardly bite the bullet on the Williams example with which I started. But this leaves us with an explanatory task—what is it, we can still ask, that seems so fishy in the Williams case? How is moral expertise different from other, presumably unproblematic, cases of expertise? I do what I can to answer this question in section IV. A brief conclusion follows.

I. SOME PRELIMINARIES

The aim of this section is to clarify my question and focus the rest of the discussion, not to defend any substantive conclusions or to

²And so perhaps we will avoid the stalemate that Hopkins recognizes in discussion of moral testimony. See Robert Hopkins, “What Is Wrong with Moral Testimony?,” *Philosophy and Phenomenological Research*, LXXIV, 3 (May 2007): 611–34, at p. 613.

³But see Paulina Sliwa, “In Defense of Moral Testimony,” *Philosophical Studies*, CLVIII, 2 (March 2012): 175–95, at p. 193, for comments that, though underdeveloped, anticipate precisely this point. For an overview of the moral uncertainty literature, and for references, see Andrew Sepielli, “Moral Uncertainty,” forthcoming in the *Routledge Encyclopedia of Philosophy Online*.

seriously engage the literature on them. Some of my choices regarding the framework in which I place my discussion of moral deference may be neither obvious nor uncontroversial. More can be said defending these choices, but for our purposes here they can be treated as stipulatively characterizing the question that interests me.

My question is about the conditions—if there are any—under which it is justified for Williams's student to accept the professor's conclusion "on the strength of his professorial authority," or those under which it is justified for me to vote against funding the war simply because Alon judges that it cannot be morally justified. The point of view most relevant for my inquiry, then, is that of the deliberating agent who—uncertain about what morality requires of her—considers relying in this way on the moral judgment of another. Following McGrath,⁴ I highlight this fact by speaking primarily of moral *deference*. All I will be assuming regarding the person to be deferred to (the *expert*, in one sense of this term) is that the deferring subject believes the expert is much more likely to be right on the relevant question. I will not be assuming anything about the expert's epistemic or practical virtues—perhaps this makes the use of the word "expert" a bit of a stretch, and partly for this reason I will talk almost entirely of deference.

In order to isolate—as much as possible—the phenomenon that interests us, we should abstract from the relevant effects of *non-moral*, *factual* expertise. We are to assume, then, that there are no non-moral facts that Alon knows of and that I do not, and that I know as much.⁵ For similar reasons, I will also abstract from relevant *political* complexities.⁶ I am just going to assume that there are no relevant abuse-of-power worries, and that the question is entirely a private, first-personal question (*how* shall I vote, given what I hear from Alon) rather than one about the political role, if any, moral experts should be allowed to play.

Some of the relevant literature emphasizes the epistemic problems in identifying the relevant moral experts, or those to be morally deferred to. The phenomenon of moral disagreement—especially among the purported experts—is thought to make things even

⁴ See Sarah McGrath, "The Puzzle of Pure Moral Deference," *Philosophical Perspectives*, xxiii, 1 (December 2009): 321–44.

⁵ If I am right in what I am about to argue, it will turn out that the distinction between factual, non-moral expertise and moral expertise is not that important after all. But this should be *shown*; it cannot be assumed from the start.

⁶ Jones also sets these issues aside. See Karen Jones, "Second-Hand Moral Knowledge," this JOURNAL, xcvi, 2 (February 1999): 55–78, at p. 64.

worse. I have some doubts about such moves,⁷ but here I can afford to bypass these issues: for my vindication of moral deference it will suffice that in the Alon example I *wholeheartedly believe*, or perhaps I *justifiably believe*,⁸ that he is more likely to be right than I am. That such cases are possible should be, I think, uncontroversial, regardless of more delicate issues in moral psychology: clearly, such sincere belief is possible. And we can imagine many cases in which it would be justified or best supported by evidence. Perhaps, for instance, evidence of past track record is relevant, as when, in the Alon case, I view my change of mind about previous wars as a correction of a mistake to which the expert was not vulnerable; perhaps I have testimony-based evidence about someone's being much more reliable than I am about such matters;⁹ perhaps I have evidence that Alon is more reliable than me in other domains, and I have reason to think that morality is not different from them in this respect; and so on.

Now, we should be careful about the objective and the subjective ought (or perhaps, about the debate over which it is that best captures *the* sense of "ought").¹⁰ We are here in subjective territory. The question that interests us is not whether the war is just, and so also not whether I objectively-ought to vote for funding it. The question

⁷Jones ("Second-Hand Moral Knowledge") and Sliwa ("In Defense of Moral Testimony") go through several possible explanations of how people could come to have differential moral expertise.

McGrath ("The Puzzle of Pure Moral Deference," p. 333) talks about problems in establishing a moral track record, or in calibration, or the absence of independent checks. Also, she insists on there being something objectionably circular in deciding who the expert is based on her agreement with us on some moral judgments. But it is not clear what an independent check would be, or why the relevant circularity is either objectionable or unique to this case. (In another paper, McGrath gets very close to acknowledging these difficulties. See McGrath, "Skepticism about Moral Expertise as a Puzzle for Moral Realism," this JOURNAL, CVIII, 3 (March 2011): 111–37.) I think—but will not be able to argue the case here—that the problem (to the extent that there is one) in this way of establishing expertise is not circularity but rather transmission failure, in the sense discussed by Wright. See, for instance, Crispin Wright, "Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof," *Philosophical Issues*, x, 1 (October 2000): 140–63.

⁸Whether this is so will depend on some issues regarding the subjective ought. I get to this topic below.

⁹Notice, of course, that the matter is also comparative: what matter for my purposes are not absolute success rates but comparative ones. What matters for me in deciding whether to defer to Alon, in other words, is not so much how reliable he is but to what extent (if at all) he is more reliable than I am.

¹⁰See, for instance, Peter A. Graham, "In Defense of Objectivism about Moral Obligation," *Ethics*, cxxi, 1 (October 2010): 88–115; Michael J. Zimmerman, *Living with Uncertainty: The Moral Significance of Ignorance* (New York: Cambridge, 2008), and the references there.

that interests us is what it is reasonable for me to do given my ignorance, uncertainty, and partial evidence. Even if you are an objectivist about "ought," you should acknowledge that this is sometimes an intelligible question worth asking, and it is this kind of question I am asking in this paper.

So far I have been talking about responding to Alon's purported expertise both by *forming the judgment* that the war is unjust and by *acting accordingly* by voting against funding the war. But the two may come apart. Perhaps, say, though there is nothing *epistemically* problematic with expertise-based moral judgments, still there is something problematic in acting on them.¹¹ Or perhaps, though expertise-based moral judgments are in some way epistemically inferior to moral judgments supported in other ways (or to expertise-based judgments elsewhere), still acting on them is required. My vindication of moral deference and expertise will be more in line with this second suggestion. I will be insisting, in the first place, that I should vote against funding the war based on Alon's expertise, and indeed that Williams's student should *act* on his professor's moral judgments.¹² What I say will also count rather heavily, I think, in favor of a vindication of the *epistemic* status of expertise-based moral judgments. But the main argument will not depend on this: even if, in other words, deference-based moral judgments are epistemically suspicious, still we should act on them.

Finally, a word about metaethics. Some of the literature on moral deference and expertise assumes (sometimes implicitly, sometimes explicitly) that the problem or its solution is sensitive to the differences between broader metaethical views (perhaps roughly divided into realist and non- or anti-realist ones).¹³ As others have noted, though, I do not think that this is so. To see this, let us drop all talk of moral *truth*, and *fact*, and *belief*, and *knowledge*, and *representation*, and *mind-independence*, and *objectivity*, and certainly of *irreducibility*. Let us just assume—incredibly minimally—that not all seemingly moral judgments are on a par, that some are, as I shall say, in some way more okay than others. This can be because some are true and

¹¹ For a somewhat similar distinction in terms of unavailability and unusability, see Hopkins ("What Is Wrong with Moral Testimony?"). Hopkins is explicit about there being no epistemic problem in relying on moral testimony.

¹² Notice that Williams's formulation is ambiguous on this point.

¹³ This is especially clear in McGrath's relevant writing ("The Puzzle of Pure Moral Deference"; "Skepticism about Moral Expertise"). See also Terrance C. McConnell, "Objectivity and Moral Expertise," *Canadian Journal of Philosophy*, xiv, 2 (June 1984): 193–216, who talks in terms of objectivity, though his understanding of objectivity is rather thin.

others false, or because some are assertible and others are not, or because it is a good idea to inculcate some but not others, or because some are in line with our preferences more than others, or because some would be agreed on by parties in some hypothetical conditions and others would not, or because some capture a truth about the maximization of happiness and the minimization of suffering and others do not, or anything at all. So long as some moral judgments are more okay than others, it is possible that some people have a better chance than others of hitting on the more okay ones. And now we can tell the Alon story all over again, this time in terms of my fully appreciating that on matters such as the moral defensibility of a war in its early stages Alon has a better okayness percentage than me. And so we can ask whether I should defer to him for this reason. It seems to me ~~this~~ way of putting things takes nothing away from the initial question. Similarly, we can without any loss re-describe the Williams student-professor example in terms of okayness. If so, the worries about moral deference and expertise apply to any metaethical view, so long as it distinguishes between more-okay and less-okay moral judgments.¹⁴ This condition, it seems to me, applies to all views in the current literature, and indeed, if there is a view to which this condition does not apply then this itself is reason enough to reject it.¹⁵ And so, metaethics is here, I think, pretty much beside the point. In the main text of the paper I allow myself to speak in realism-friendly terms. If you find this to be a distraction, feel free to replace them with your favorite terms instead.

Thus, the question here is not whether I should act on Alon's advice,¹⁶ or on his authority, but whether I should act on his judgment simply because (as I believe) he is much more likely than I am to get it right; this is the only thing I know about him—I do not know how he came to make this judgment, what his reasons

¹⁴For this point see Elizabeth Fricker, "Testimony and Epistemic Autonomy," in Jennifer Lackey and Ernest Sosa, eds., *The Epistemology of Testimony* (New York: Oxford, 2006), pp. 225–45, at p. 237; and in the aesthetic context, Jon Robson, "Aesthetic Testimony," *Philosophy Compass*, vii, 1 (January 2012): 1–10, at p. 7.

¹⁵For a similar point, see Peter Singer, "Moral Experts," *Analysis*, xxxii, 4 (March 1972): 115–17, at p. 115. And there is a hint at this direction in Jones ("Second-Hand Moral Knowledge," p. 56n3).

¹⁶Of course, there may be good reasons to listen to what others have to say that are not related to moral deference and are not suspicious in the way deference arguably is—perhaps you can benefit, for instance, from learning about another person's perspective. But here I focus on the narrower, and also more suspicious, phenomenon of moral deference, where I (for instance) vote against funding for the war on Alon's say-so.

are,¹⁷ whether he himself would act on it, and so on. We assume that he is not privy to any non-moral information that I am not privy to, and that I know as much; worries about abuse of power have been set aside (by stipulation), as have worries about how to recognize moral experts; the question whether I should act on his moral judgment employs something like a subjective “should” or “ought,” but in a way that objectivists should not find objectionable; and our question is in the first place about whether I should *act* on his judgment, only secondarily about what I should believe.

To a large extent, these clarifications just serve to characterize the problem I will be addressing—all of this is consistent, of course, with other problems in the vicinity being also worth addressing. But I also think that the problem as characterized by the cleaned-up Alon example is of central importance here.

II. SOME DIAGNOSTICS

It will be worth our while to ask how general the phenomena that interest us are. In particular, it will be interesting to see how widely Williams’s discomfort applies. It seems plausible on methodological grounds that in trying to understand—and justify, or anyway explain—Williams’s discomfort we should aim at an account that is (perhaps roughly) as general as the phenomenon to be explained.

II.1. Opaque Evidence. Suppose that I have come to see the following striking correlation: when people perform wrong actions, their heart rate increases significantly. Now, I see you performing an action, say, turning the trolley in *Loop*,¹⁸ and I get an indication that your heart rate increased significantly. I infer (fallibly, but rather confidently) that turning the trolley in *Loop* is morally wrong. And let us further assume that this is so, that the conclusion of my inference is true. True, some details have to be filled in—most notably, perhaps, in order to deal with what in Kantian contexts is sometimes referred to as the action description problem. But such difficulties do not seem to be insurmountable, so let us assume them away.

Now suppose that I believe that turning the trolley in *Loop* is wrong; I assert as much when the topic comes up; I criticize those who

¹⁷ Actually, in a very indirect way, Alon’s reasons may be relevant. In the example, by deferring I am in effect projecting on Alon’s reliability. If you have doubt whether this is indeed projectable in such a way, you may want to know more about the background facts, such as Alon’s reasons. (I thank Jonathan Berg and Saul Smilansky for this important point.) But I think we can bypass it for our purposes by stipulating that in the relevant case Alon’s success rate is projectable.

¹⁸ Judith Jarvis Thomson, “The Trolley Problem,” *The Yale Law Journal*, xciv, 6 (May 1985): 1395–415.

do (or would) turn the trolley in Loop; having turned the trolley in Loop myself, I feel guilty; and so on. How do we feel about this case?

It seems to me very clear that Williams's discomfort is fully present. There are really two claims here. The first is that there is *something* fishy about this way of forming a moral belief, that it feels importantly different—and not for the better—from more standard ways of forming moral beliefs. The second claim is that the fishy scent is familiar to us already, indeed that it is *the same* fishy scent that Williams diagnosed in the case of the student and his professor. I do not know how to argue for these claims: they seem clear to me on phenomenological grounds. So all I can do is ask you to reflect on the cases and see whether you agree. In what follows, I will be assuming that the answer is “yes.”

The point generalizes, I think, to all cases of moral beliefs based on opaque evidence.¹⁹ What I mean by “opaque evidence” here is, roughly, evidence for the truth of a proposition that does not disclose that in virtue of which the proposition is true, or its truth-makers. In this sense, then, if my reason for believing that turning the trolley in Loop is morally impermissible is that doing so will be using the one person merely as a means—as we never should—then I believe that turning the trolley in Loop is wrong on transparent evidence. But if my reason for believing that it is wrong is the said correlation between wrong actions and the agent's heart rate increasing, this is opaque evidence for the same proposition. Intuitively, transparent evidence is such that someone who believes the moral proposition based on it also has insight into why the relevant moral proposition is true. Not so for moral judgments based on opaque evidence.

This applies to expertise-based moral judgments as well. Even if the student can know that the professor's judgment is true—on the strength of his professorial authority—still if this is his evidence the student has not gained insight into *why it is* that that conclusion is true, or that the relevant action is wrong (or right). I think it is not premature, then, to conclude that the problem with expertise-based moral knowledge is an instance of the more general problem of opaquely based moral knowledge.

¹⁹ As far as I know, this point has not been noticed in the literature. The closest I have found is Hopkins's (“What Is Wrong with Moral Testimony?,” p. 630) “The Requirement,” which seems to apply more generally. But he does not explicitly draw this conclusion.

Whether the awkwardness of beliefs based on opaque evidence is present also in non-moral beliefs—or whether some closely related phenomenon is—is a wider question than I cannot discuss here.

If this is so, the following lesson then follows: the problem of moral expertise and deference is really more general and applies to all opaque evidence. The solution—vindicating moral expertise, explaining Williams's discomfort—should apply to the more general phenomenon of moral judgments based on opaque evidence, and so should not depend on features peculiar to the case of deference and expertise.

II.2. Scope: Other Normative Domains. Does Williams's discomfort arise when expertise is appealed to in other normative domains?²⁰ Here are three examples.

I am trying to make up my mind whether to open the second box in a Newcomb Problem setting. It seems to me—having thought about the issue for a while—that two-boxing is the way to go. But I also know that you have a better track record than I do when it comes to prudence, or rationality (of this kind, perhaps), and I see you refusing to open the second box, and indeed insisting that this is also what I should do. I then form the belief that—my initial contrary belief notwithstanding—one-boxing is the rational way to go, and I act accordingly. (I even criticize those poor two-boxers.)

Having thought about black ravens for a while, it seems to me that it is just not true that every particular instance of a generalization confirms the generalization. But I also know that you have a better track record when it comes to epistemological judgments than I do, and I hear you insisting that the paradox can be solved some other way, and that actually every particular instance of a (non-gruesome) generalization confirms it. I defer to you, then, come to believe your epistemological judgment, and conduct my intellectual affairs accordingly. (I even criticize those who do not.)

Having spent a lot of time studying music and listening to both Bach and Mozart, it seems to me that Bach is by far the greater composer. But you (as I know) have a better track record when it comes to appreciating the value of compositions and composers, and I know that you actually think Mozart the greater composer. So, on your expertise, I come to believe that Mozart is the greater composer, and conduct (this part of my) aesthetic life accordingly. (I even criticize those who believe that Bach is better.)

²⁰ There is an initial remark in this direction in McGrath ("The Puzzle of Pure Moral Deference," p. 339). There she also mentions the case of metaphysical expertise, which I discuss in the next subsection.

How do we feel, then, about expertise and deference when it comes to prudence²¹, epistemology²², and aesthetics²³? Is Williams's discomfort present? Well, *some* discomfort is present, I think. At the very least, there is something odd about normative judgments so formed. Whether it is exactly *the same* oddity seems harder to tell. So I am going to return to this question later on. For now, I think we should note that something in the vicinity of Williams's discomfort applies in these other domains. And this means that we should keep an open mind about the possibility that the problem of moral deference and expertise is not really about morality at all—that it arises, rather, because of some features that all normative judgments, moral and otherwise, have in common. Indeed, when I present my vindication of moral expertise, it will emerge that the moral case is special, as it were, in the other direction: it is *easier* to justify moral deference than deference with regard to some other normative domains.

II.3. Scope: Philosophy. I am not much of a metaphysician. I mean, I read and think about the metaphysics of modality like the next guy, but I'm no David Lewis.²⁴ And so I come to believe modal realism—a theory I had initially found surprising, like everyone else—not because I am convinced by Lewis's arguments, but on his professorial authority. Does Williams's discomfort apply?

Again, *some* discomfort is certainly present. This, after all, is not how we do philosophy. And we do not grade philosophy papers or exams by the number of truths in them. But I think that the discomfort here can be explained in a way that is different from—though

²¹ Julia Driver, "Autonomy and the Asymmetry Problem for Moral Expertise," *Philosophical Studies*, cxxviii, 3 (April 2006): 619–44, mentions the prudence case (at p. 620), and (at p. 623) the possibility of generalizing the discussion of moral expertise to other normative domains, but she does not develop this point.

²² Actually, it is hard to get a clean epistemological example. This is so because if we are going for a general epistemological principle—as in the text—then it seems to be a case of purported *philosophical* deference, of the kind I discuss in the next section. And if we go for a more specific judgment—say, whether I am justified in believing that all ravens are black—then the expert that seems relevant is an expert on the subject matter (ravens, that is), not on justification. I thank Yuval Eylon for making me see this.

²³ I should say that I do not know remotely enough about aesthetics to be confident about this kind of case. For an overview, and for many references, see Robson's "Aesthetic Testimony." Still, let me again emphasize how non-normative but normatively relevant judgments may be relevant here. Often the aesthetic experts are just more sensitive than the rest of us to relevant non-aesthetic but aesthetically relevant facts (they have a finer palate, as it were). Driver's discussion of aesthetic expertise in "Autonomy and the Asymmetry Problem for Moral Expertise" ignores this complication.

²⁴ This example is inspired by a title Bryan Frances used for a paper, but then he decided to go with another. The title was "Who Am I to Disagree with David Lewis?"

related to—what is going on in the normative cases. Because the story I end up telling in section III does not immediately apply to the philosophical case, let me sketch one that does—it will turn out that what I do say about moral expertise will be interestingly related to this other story.

The main problem with believing modal realism on the strength of David Lewis's professorial authority, I believe, is that it is not an impressive intellectual achievement. What we value in doing philosophy, it seems, is originality, insight, depth, rigor—the virtues that are present when someone thinks of a new theory, or a new argument, or appreciates in a new and vivid way the full depth of a known problem, or highlights unnoticed connections between two philosophical theses, or at least—in grading papers, for instance—when someone appreciates the reasons for and against a philosophical thesis. Just hitting on the true thesis is—though in some respects the aim of the whole project—still much less of an achievement. This is why we are not impressed with the person who accepts the true (let us assume) thesis of modal realism on the strength of Lewis's professorial authority. This is also why we find the idea of accepting a philosophical theory by deference somewhat problematic, at least in the context of doing philosophy: it betrays rejection of the virtues relevant to doing philosophy.

In this respect, I think that philosophy is like mathematics. Accepting a mathematical theorem on the strength of some mathematician's professorial authority—without, say, understanding the proof—is not an impressive intellectual achievement, so there is something fishy about mathematical deference. But of course, those of us who are not mathematicians, and who need to rely on mathematical propositions for some other purposes, would do well to defer to expert mathematicians. The deference is justified, even if not as a manifestation of an intellectual achievement. Similarly, those of us who are not philosophers, and who have to rely on a philosophical proposition for some other purposes, would do well to defer to the philosophical expert—if, that is, *there are* other purposes for which, say, the truth of modal realism is relevant,²⁵ and if other problems can be dealt with (for instance, the fact that there is much more disagreement among the purported philosophical experts than among the purported mathematical ones).

I get back to talk of achievements and their relevance to moral expertise in section IV.

²⁵ Perhaps, as Paulina Sliwa suggested to me, a philosopher working in one field may need to rely on experts' testimony from other philosophical sub-disciplines.

III. THE VINDICATION

In subsection III.1 I present my vindication of acting on the judgment of the moral expert. I then discuss objections having to do with the value of moral autonomy and understanding, and with the role of intentions. I then discuss the scope of my vindication.

III.1. The (Sort-of) Requirement to Minimize Risk of Wrongdoing. What would be involved in *refusing* to defer to Alon, in insisting on finding out for oneself, as it were, unassisted, about the moral defensibility of the war, and in voting about its funding accordingly?

Remember that we are holding constant throughout both that the war seems to me defensible and that I wholeheartedly acknowledge that Alon is much more reliable than I am on such matters in such circumstances. So by refusing to defer to Alon I will be in effect accepting a higher risk of forming a false moral judgment. More to the point, I will be accepting a higher risk of *acting* wrongly, of (for instance) voting to fund an unjust war. But if I have a way to minimize the risk of my wronging people, and if there are no other relevant costs (to which I return shortly), why on earth wouldn't I minimize this risk?²⁶

We are in this context asking about the general norms of appropriate response to moral uncertainty. Not much in this context is either clear or uncontroversial. Let me reiterate that I am trying to abstract from as much of these controversies as possible, partly by being noncommittal on the objective and subjective "ought," restricting my discussion to the way—whatever exactly it is—we can evaluate ways of facing uncertainty as more or less reasonable. I think it is pretty safe to say, regardless of these wider controversies, that when the following "should" is understood in this noncommittal way, and when all other things are equal,²⁷ we should minimize the risk of acting wrongly.²⁸ It is this minimal requirement that I am insisting on—a requirement so minimal that it should be accepted, I believe, by the proponents of just about any plausible normative

²⁶ Jones ("Second-Hand Moral Knowledge," pp. 66–67) notes that it would be astonishingly arrogant to suppose that the importance of the moral supports a do-it-yourself attitude.

²⁷ Doesn't this get me involved in the Problem of Intertheoretic Value Comparisons? Perhaps, but not in a problematic way. See below.

²⁸ If "acting wrongly" already incorporates a subjective element, this may complicate things a bit, but not in ways that it is necessary for our purposes to tidy up. Similarly, some problems may arise here around the precise understanding of *risk*. Once again, I do not think more details are needed here for our purposes (but let me note that with regard to risk I am not as confident as I am with regard to wrongness).

moral theory. (Perhaps this requirement would be grounded in different considerations according to different theories. But as a mid-level principle, it should be fairly uncontroversial.) It is this minimal requirement that I would violate if I refused to defer to Alon.²⁹

Notice that the requirement to minimize the risk of wrongdoing incorporates an other-things-being-equal clause. This is important partly because when other things are not equal it is going to be difficult to weigh this requirement against others, and I do not have anything interesting and general to say about that. But this other-things-being-equal proviso is important also because when other things are not equal, other strategies of coping with moral uncertainty may be relevant. Perhaps, for instance, it is sometimes advisable (perhaps even required, in the given-my-evidence sense) to err on the side of safety, to make the decision such that, if it ends up being the wrong one, less harm is caused, or people are less seriously wronged. So by assuming that other things are equal, I am also assuming that no side is (known to be) the side of safety, that we have no indication that an error will be less costly on one side than the other. I think that many war examples are of this kind: there are dire consequences of making *any* mistake, on either side, in deciding whether to engage in armed conflict. But if you think that war cases are different in this regard, all we need is another example. Suppose, then, that I am a judge deciding a private-law case. We can assume that the parties are quite symmetrical, in the sense that if I erroneously find for the plaintiff, the defendant is harmed and wronged to an extent similar to that in which the plaintiff will be harmed and wronged, if I erroneously find for the defendant. In such a case too, it seems to me, if all other things are equal, I should minimize the risk of wrongdoing. And if I believe that Alon—a colleague judge, perhaps—is much more likely to get it right than I am, then it seems to me I should defer, exactly as in the original Alon case.³⁰

²⁹ I think I can remain neutral here between the following two suggestions. The first is that there is a separate second-order *pro tanto* moral requirement to minimize the risk of my wrongdoing. The second is that there is no need here for a separate requirement, as minimizing the risk of acting wrongly is called for by the relevant first-order norm itself. So if I am morally required not to hurt innocents without sufficient justification, *this requirement* already requires that—given uncertainty—I minimize the risk of violating it. On this understanding of the normative architecture, deferring to Alon is just an indirect way of conforming (to an extent, at least) to the norm that requires that I not hurt innocents unnecessarily.

³⁰ I take the example of two judges from Hills, though with somewhat different details and used for a different purpose.

We can dramatize the matter. Suppose that the war is indeed indefensible, I—erroneously thinking that it is justifiable—vote for funding it, innocents die, and a family member criticizes me for my vote. If I did all that I could (within certain constraints) to find out whether the war was morally defensible, and failed, at least I can explain that I was not certain about the moral status of the war and I did what I could to find out. This defense is not without its problems, of course, but nor is it without weight altogether. Now suppose that Alon told me before the vote that the war was indefensible. If so, it no longer seems that I can answer in good faith that I did what I could to find out about the moral status of the war. The family member can insist that I had a more reasonable way available to me for coping with moral uncertainty—I could have asked Alon, whom I acknowledge as much more reliable than me on such matters in such circumstances. And it seems to me that she would be right.

In these respects, the situation is no different, I think, from analogous cases of coping with *non*-moral uncertainty. Whatever your view on objective and subjective “ought,” surely you should agree that there is *a* sense of reasonability (or some such), such that if you have to either cut the blue or the red wire in order to neutralize the bomb and save lives, and you know it is much more likely that cutting the red wire will neutralize the bomb, then you should—in this reasonable-given-my-information sense of “should”—cut the red wire. This is so even if it turns out that cutting the blue wire would have neutralized the bomb, so that in the objective sense of “ought” you ought to have cut the blue wire. This seems to me another instance of the general norm appealed to above: we should minimize the risk of acting wrongly when we can do so without other relevant costs. Whether the uncertainty involved is factual or moral seems to me neither here nor there.³¹ Now, many interesting questions can be asked about this norm, in particular about its interaction with other norms and values. But notice that the norm I am appealing to here is much more minimal—it is about the *pro tanto* requirement to minimize risk of wrongdoing, when no other relevant costs are present.

So, at least when it comes to action, moral deference and expertise are justified. In fact, when someone genuinely believes that another is more likely to be right on a moral question, moral deference is not just permissible but rather is morally required.³² Refusing to

³¹ Though you may think that the Problem of Intertheoretic Value Comparisons means that there is a difference here after all. Again, see below.

³² McConnell (“Objectivity and Moral Expertise,” text near footnote 19) anticipates this point, I think.

defer in such circumstances would amount to unjustifiably accepting a higher risk of compromising others' morally protected interests, or of wronging them. This is inconsistent with giving others' morally protected interests the right place in one's practical reasoning. It would be wrong.

Let me end this section with two worries about the requirement to minimize the risk of wrongdoing, and so also about my vindication of moral deference.

First, a central problem in the literature on moral uncertainty is the Problem of Intertheoretic Value Comparisons.³³ Thus, many of the proposed strategies for responding to moral uncertainty—maximizing expected moral value, maximin (which to an extent is just the strategy of erring on the side of safety), and perhaps others as well—require that we compare moral value under one theory (that is, if one theory is true) with moral value under a competing moral theory (that is, if that other theory is true). And it is not immediately clear how (or even whether) this can be done. I too am vulnerable to this problem, for I have been insisting on the need to minimize the risk of wrongdoing *when all other things are equal*. Presumably, these other things that have to be equal include the seriousness of the purported wrong; surely, we do not want to go for a somewhat smaller risk of committing a much more serious wrong than for a somewhat higher risk of committing a much less serious wrong. So the seriousness of the two purported wrongs (one if Alon is right, one if he is not) has to be roughly equal. And this calls for an intertheoretic value comparison. If such comparisons do not make sense, then I am in trouble.³⁴ True, like all those supporting some way of “moral hedging” (that is, of taking the risk of erring morally into account and trying to reduce it or its effect), I need some intertheoretic value comparisons to make sense. But I do not need *all* of them to make sense. All I need is that sometimes we can know, say, that the wrongs involved in fighting the war (or funding it) if Alon is right are not significantly less serious than the wrongs involved in refusing to fund it if he is wrong; indeed, if the difference in probabilities (between Alon's being right and my being right) is great enough, what is needed is even weaker

³³ See Seppielli, “Moral Uncertainty,” and the references there.

³⁴ Unless the right view of how to cope with moral uncertainty is to always go with the action that is prescribed by the moral theory one is more confident in (regardless of how much more confident one is in that theory, and of how good or bad the action is according to the rival theories). On such a theory, no intertheoretic value comparisons are needed, and my vindication of moral deference still works. But I do not want to rely on this (highly implausible, I would say) solution to the problem of moral uncertainty.

than that. This too, I take it, is controversial, but I do not think it should be. That *some* wrongs according to one theory are less serious (or some things of less value) than others according to another theory should serve as an adequacy constraint on accounts of moral uncertainty. I am happy to take this on board as a premise in my defense of moral deference and expertise.³⁵

Second, we should distinguish between the requirement to minimize the risk of wrongdoing, which I have been emphasizing, and a (purported) requirement to minimize the risk of *harming* others. Perhaps, in other words, given some factual uncertainty about the harmful consequences of alternative possible actions, we should act, when all other things are equal, in a way that minimizes the risk of harming others. However, it does not follow from this plausible idea that even when no greater risk of *harming* is involved, we should minimize the risk of *wrongdoing*. And so perhaps my intuitive case for the requirement to minimize the risk of wrongdoing traded on failing to distinguish between minimizing expected harm in the face of *factual* uncertainty and minimizing risk of wrongdoing in the face of *moral* uncertainty. In response, I agree, of course, that the requirement to minimize the risk of wrongdoing does not follow from a requirement to minimize the risk of harm in the face of factual uncertainty. But I think that the requirement to minimize the risk of wrongdoing is *independently* plausible. To see that, we need to focus on cases (perhaps like the two judges case above) where the harm is both certain and equal on both sides, and where the only uncertainty regards the moral question—who should be made to suffer that harm. In such a case, nothing like a requirement to minimize harm can help. But the requirement to minimize the risk of wrongdoing still applies, it seems to me, and is as intuitively plausible as ever: if in such a case I do not know how to decide, but I know that Alon's chances of getting it right are significantly higher than mine, then I should defer to him. We can run the arguments and thought experiments again—what would be involved in refusing to defer? What would I tell the harmed (and perhaps wronged) party, and her loved ones? In all these respects, this case is just like the case I started with, so the plausibility of a requirement to minimize risk of wrongdoing does not rest on a failure to distinguish it from a requirement to minimize expected harm.³⁶

³⁵ But see also the point in the previous footnote. Sergio Tenenbaum has been extremely helpful in making me see the problem discussed in this paragraph and in discussing my response.

³⁶ For discussion here, I thank Arnon Keren and Mike Otsuka.

III.2. The (Purported) Values of Moral Autonomy and Understanding. You may concede that the line of thought in the previous paragraphs does count in favor of moral deference, but insist that there are also countervailing considerations, that there are also reasons not to defer, and perhaps sometimes the latter outweigh the former. Here are two suggestions for such countervailing reasons.

First, considerations of autonomy have played a central role in resistance to moral expertise and deference.³⁷ Shouldn't we be autonomous, setting rules or ends or principles for ourselves, or at least (if you are as unclear as I am about what these words mean) trying to find out the moral truths ourselves? Acting on the moral expert's judgment seems to conflict with some plausible understandings of the value of autonomy: it is as if one is outsourcing one's agency.³⁸ So there is after all a cost involved in deferring even to an acknowledged expert.

Second, you may think (as do Nickel, Hills, and to an extent Hopkins³⁹) that at least when it comes to morality, there is more to worthy action than right action, and perhaps also more to moral epistemology than knowledge. The crucial thing to note here is the importance of moral *understanding*. Unlike the moral knowledge that *p* (for some moral *p*), and even unlike the moral knowledge why *p*, *understanding* that (and why) *p* involves appreciating the relation between *p*'s being true and the moral reasons that make it true: the ability to articulate the reasons in one's own words, the ability to apply them and closely related reasons in new cases, and so on.⁴⁰ Furthermore, moral understanding is *important* and so worth pursuing.⁴¹ This may be so partly because of instrumental reasons—perhaps, for instance, in many circumstances developing moral understanding improves one's chances of acting rightly over the long run. But more importantly and interestingly, this is so also for more intrinsic reasons. Hills relies primarily on two kinds of reasons. The first is the plausible suggestion that moral understanding is a necessary part of virtue and

³⁷ See, for instance, McConnell ("Objectivity and Moral Expertise," text near footnote 17); Jones's ("Second-Hand Moral Knowledge," p. 57) reading of Wolf; and Driver's ("Autonomy and the Asymmetry Problem for Moral Expertise") extensive—and at the end of the day, critical—discussion of autonomy.

³⁸ I take this term from Andy Egan and Adam Elga, "I Can't Believe I'm Stupid," *Philosophical Perspectives*, XIX, 1 (December 2005): 77–93, at p. 83, who use it in a similar—but entirely epistemological—context.

³⁹ See Philip Nickel, "Moral Testimony and Its Authority," *Ethical Theory and Moral Practice*, IV, 3 (2001): 253–66; Hills, "Moral Testimony and Moral Epistemology"; and Hopkins, "What Is Wrong with Moral Testimony?"

⁴⁰ Hills, "Moral Testimony and Moral Epistemology," p. 102.

⁴¹ *Ibid.*, p. 106 and on.

admirable moral character—for these clearly include more than merely doing (often enough) the right thing. They also include acting rightly for the right reasons, and understanding the relation between the reasons and the rightness or wrongness of the relevant action. The second, more intrinsic way in which understanding can be seen to be morally important is as a constitutive part of morally worthy action.⁴² There is something more morally worthy about an action that is not just motivationally but also cognitively sensitive to the morally relevant features of the situation, and the relevant kind of cognitive sensitivity just is the kind of moral understanding Hills (for instance) is drawing attention to. Now, while the say-so of another may be a legitimate source for moral knowledge, it is not often a source of moral understanding. So there is always something problematic—something less good—about forming moral beliefs by deference. By deferring to another on moral matters, you are always paying a price, a price in the valuable currency of moral understanding. And action on such beliefs inherits this imperfection.

Let me concede, for the sake of argument at least, that moral autonomy and understanding are genuinely of value and so genuinely worth pursuing. This means that we can concede also that my deferring to Alon and voting against funding the war is not the ideal scenario for me. The ideal scenario for me would be to achieve moral understanding and make an autonomous choice to vote against the war. But this is the wrong comparison. When I am deliberating whether to vote for or against funding the war, and indeed whether to defer to Alon, I can perhaps *hope* to have more by way of moral understanding. But this is not an option available to me right now.⁴³ The relevant options available to me are either to defer or to think for myself, going, as it were, for the gold (of acting not only rightly but also autonomously and with moral understanding). The problem is that if I do attempt to think for myself, I am deliberately accepting a higher risk of acting wrongly here (we are holding constant, remember, that Alon is, as I believe, much more likely than I am to get things right). *These* are the two relevant options for me. Though it would be nice if I had a third—acting rightly and also autonomously and from moral understanding—this is not possible for me now except by paying a price in greater chance of wronging people.

You may think that this may be a price worth paying. After all, sometimes going for the gold is worth the risk of losing even the

⁴² *Ibid.*, pp. 113–19.

⁴³ See Sliwa, “In Defense of Moral Testimony,” for a similar point.

silver. I agree that this is sometimes the case, but I do not think this is so here. To see why, think of the following analogy.

Compassion is of value. And compassion's value is reason-giving—we often have reason to see to it that we become more compassionate, or that those for whose moral education we are responsible develop appropriate compassion, and so on. Plausibly, developing (or maintaining) compassion requires encounters with human (and perhaps other) suffering. So is it ever permissible (or even virtuous, or morally worthy) to act in a way that is likely to increase human suffering in order to secure opportunity for bettering ourselves and others in terms of compassion? We should be careful here, because of the possible significance of instrumental effects. Perhaps, for instance, unless our children can genuinely encounter others' suffering they cannot develop compassion. And perhaps then they are likely to behave in ways that will cause even more suffering. If this is so, it seems plausible to say that at times we, as parents and moral educators, should take our children's moral training wheels off⁴⁴ and be willing to accept greater unalleviated suffering (or perhaps greater probability of unalleviated suffering) in one case in order to decrease the overall "amount" of unalleviated suffering (by helping to train more compassionate agents). Things are complicated here, both empirically (I do not know how exactly education for compassion works) and morally (we are here in the territory of suspicious interpersonal tradeoffs). But it is at least plausible to say that sometimes this is the way to go. Notice, crucially, that what justifies in this case the willingness to tolerate greater suffering is not the intrinsic value of compassion itself; rather, it is compassion's instrumental value—its likeliness to result in less unalleviated suffering overall. And the point I now want to emphasize is that without this instrumental feature, it cannot be justified to accept greater suffering just in order to develop compassion. The point is not, to repeat, that compassion is not of intrinsic value. Nor is the point that compassion's intrinsic value is normatively inert. The point, rather, is that though compassion is quite plausibly of intrinsic value, and though this value does give us reasons for action, *it never by itself gives us a non-instrumental reason to act in a way that tolerates greater suffering*. The reason for this is that compassion's value constitutively depends on the compassionate's motivation to relieve or prevent suffering. Tolerating greater suffering for the value of compassion is thus self-defeating, indeed perhaps even practically inconsistent. Someone willing to tolerate greater

⁴⁴I take this metaphor in this context from Jones, "Second-Hand Moral Knowledge," p. 56.

suffering—even by a small margin, and even just the greater *probability* of greater suffering—merely in order to work on his (or anyone else's) compassion, that is, independently of the supposed instrumental payoffs of having more compassionate people around, is acting wrongly, and most clearly uncompassionately.

Returning to moral understanding: I am happy to concede for the sake of argument that autonomy and moral understanding may have instrumental value, in allowing those who have them to act more reliably in permissible ways. And the point I have been making using the requirement to defer to Alon was not that autonomy and moral understanding are not of intrinsic value. Nor was it that these values do not give us reasons. The point was, rather, that *the value of moral autonomy and understanding never by itself gives us a non-instrumental reason to act in a way that tolerates a greater risk of injustice or of acting immorally*. The reason for this is that the value of moral understanding (and perhaps also of autonomy) constitutively depends on the autonomous, understanding person's motivation to respond to the morally relevant features of the situation, and thus to act morally (though perhaps *de re*, rather than *de dicto*). Tolerating a greater risk of wronging others merely for the value of moral autonomy and understanding is thus self-defeating, indeed perhaps even practically inconsistent. Someone willing to tolerate a greater risk of acting impermissibly merely in order to work on her (or anyone else's) moral understanding, that is, independently of the supposed instrumental payoffs of having more morally understanding people around, is acting wrongly, and indeed exhibits severe shortage in moral understanding (of the value of moral understanding, among other things).

Notice that my point is not best understood in terms of the *strength* or the *weight* of the relevant reasons.⁴⁵ Rather, it is best understood in terms of the *nature of the relevant value*, and so the nature of the actions for which it (non-instrumentally) provides reason. Compassion is not the kind of value that can recommend tolerating greater suffering (without some fairly strong instrumental rationale, at least). And moral autonomy and understanding are not the kinds of values that can recommend accepting a greater risk of acting immorally (without some fairly strong instrumental rationale, at least). In both cases, the relevant attitude (tolerating greater

⁴⁵ Hills ("Moral Testimony and Moral Epistemology," pp. 123–24) is willing to allow other considerations—having to do with the importance of avoiding wrongdoing in the specific case—to sometimes outweigh the considerations having to do with the value of moral understanding. But this is not good enough, as I point out in the text. Notice, by the way, that by acknowledging this much Hills already bites the bullet on Williams's example (without saying so explicitly), at least for *some* cases.

suffering, or accepting a greater risk of acting impermissibly) is inherently ill-suited for the relevant value (of compassion, or of moral autonomy and understanding), and so is not supported by it.⁴⁶

What about the worry that moral deference amounts to outsourcing agency in a way that is morally unacceptable, perhaps because it offends against some notion of autonomy?⁴⁷ The first thing to say is that the point above applies—even assuming there is something to be said against such outsourcing, avoiding it cannot justify accepting a higher risk of wrongdoing. But more can be said. For *agency* here is not being outsourced: if I go ahead and defer to Alon, the action (refusing to fund the war) is still very much mine, as indeed is the very decision to defer to Alon. What I am outsourcing is the moral deliberation on the merits of the case. But in order to think that this is wrong, you would have to think that there is a duty to deliberate morally for oneself, without the aid of others (or at least without the kind of aid that calls for deference). It seems to me very hard to believe that there is such a duty, and *exceedingly* hard to believe that it can outweigh the reasons we have to minimize the risk of wronging others.

Let us take stock. When it comes to the reasonable response to moral uncertainty, and when no other significant costs are involved, the thing to do, I have argued, is to minimize the risk of wrongdoing. We then embarked on a long detour regarding the (purported) values of moral autonomy and understanding, as possible candidates for other significant and relevant costs. But it has emerged that these values—even assuming that they are indeed of value—cannot count non-instrumentally against deferring to Alon. As far as I can see, the points made regarding moral understanding and autonomy generalize. There is no plausible candidate for a value here that could outweigh the general requirement to minimize the risk of wrongdoing. This means that we are back to the no-other-significant-costs scenario. Deference is the way to go. Moral deference has been vindicated.

Notice that, as promised, nothing in this vindication at all depends on any seriously controversial metaethical assumptions. Furthermore,

⁴⁶ For a structurally similar point (though about moral worth), see Saul Smilansky, "Morality and Moral Worth," in *Ten Moral Paradoxes* (Malden, MA: Blackwell, 2007).

In the text I say that the values of moral autonomy and understanding do not give rise to reasons for tolerating a greater risk of wronging others. Another possibility—which would do just as well for my purposes, I think—is that they *do* give rise to such reasons (perhaps along straightforward instrumental lines), but these reasons are excluded.

⁴⁷ Christian Piller was helpful in pressing me on this point especially forcefully.

nothing in the vindication depends on the precise epistemic status of moral judgments based on deference to experts,⁴⁸ or even on the epistemic status of the expert. The vindication I offer of moral deference is entirely a first-order, moral, and so practical matter, and it relies on the requirement to minimize the risk of wrongdoing. This requirement applies to the Alon case, it seems to me, even if for some reason basing a moral judgment about the war just on Alon's say-so falls short of knowledge, or is for some other reason epistemically suspect. And even if Alon's judgment about the war is epistemically suspicious, still so long as relying on him minimizes (for me) the risk of wrongdoing, this is precisely what I should do.⁴⁹

III.3. Intentions. You may wonder, though, whether I have left something important out of the picture—namely, the moral significance of the relevant *intentions*. Suppose that intentions are morally relevant in some fairly important way. Isn't there an important difference between, say, my intention in voting against funding the war when I do so based on my own moral reasoning and when I do so based on deference to Alon? Can't this difference undermine the vindication of acting on moral expertise?

There is something importantly right about this suggestion, as we will see in section IV. But as it stands, this is no challenge to my vindication of deference. True, there *is* going to be a difference in the intentions I act on in the two cases (based on my own reasoning, and based on deference). But the question relevant here is whether this difference makes a moral difference. In order to show that it does, it would have to be shown that the intention *to act in a way that minimizes the risk of wrongdoing* is in some important way less than fully morally acceptable—say, it fails to treat people as ends in themselves, or some such. If this is not the case, if this intention itself is right-making, then acting on it is morally permissible, indeed perhaps morally required, although it involves deference to a moral expert. And I cannot think of any reason to think that this intention

⁴⁸ However, I suspect—as do Hills (“Moral Testimony and Moral Epistemology,” p. 97), McGrath (“Skepticism about Moral Expertise as a Puzzle for Moral Realism”), and Sliwa (“In Defense of Moral Testimony”)—that epistemically, or at least when it comes to knowledge, deference-based moral judgments are not more problematic than other deference-based judgments and beliefs. Notice that in this respect—in being independent of *epistemically* vindicating moral deference—my strategy for vindicating moral deference crucially differs from Sliwa's.

⁴⁹ Levi Spectre suggested the following way of fleshing out this point (though he meant it as an objection): if in fact I know Alon to be an *anti*-expert, so that his moral judgment is much more likely to be *false* than mine is to be true, then upon finding out that he agrees with me that the war is justifiable, I should decide to vote *against* funding it. This result seems exactly right to me.

is not fully legitimate (at least when we are talking about the given-my-uncertainty kind of requirement relevant here). In fact, I suspect that any attempt to show this will beg the question against the argument offered above as a vindication of deference.⁵⁰

III.4. Scope. To what extent—if at all—does any of this apply to cases of relying on normative but non-moral expertise?

Interestingly, not so much. The vindication above relies on the moral requirement to minimize the risk of wrongdoing, and this requirement was based on intuitions about what risks we are allowed to take with other people's morally protected interests. Such intuitions are specific to morality. What can we say about the other normative cases?

Recall the Newcomb Problem case, where it seems to me that two-boxing is the way to go, but I acknowledge that you are much better at such things and consider deferring to your judgment that I should open just one box. Now, there is probably no general moral requirement that I minimize the chances of acting imprudently.⁵¹ But perhaps there is a *prudential* requirement to that effect. I am not sure, but it is this kind of requirement that would be needed to vindicate acting on prudential expertise in a way analogous to the vindication of acting on moral expertise above. Perhaps more generally, a requirement to maximize the chances of acting in accordance with the practical reasons that apply to one would be needed in order to offer a similar vindication of deference to the expert on matters of practical reasons. I do not know whether such a requirement is plausible.

Already, then, the analogous vindication of deference in other normative domains seems a little less promising. And I am not sure what to say of the epistemological case, but when it comes to aesthetics, the analogous vindication looks less promising still. Surely, there is no norm of aesthetic appropriateness that requires that we form our aesthetic judgments in a way that minimizes the risk of aesthetic mistake. Of course, that my suggested vindication does

⁵⁰ For an analogous reason, by the way, I believe that talk of the (purported) universal availability of moral truths is irrelevant to our problem as well, and this despite the historical credentials here. McGrath ("The Puzzle of Pure Moral Deference," p. 325 and on) and Driver ("Autonomy and the Asymmetry Problem for Moral Expertise," p. 621 and on) survey some of these historical sources. The problem is that some story has to be told as to why availability-by-deference is not availability enough. And it is not clear that such a story can be told in non-question-begging terms.

⁵¹ In particular cases there may be such a requirement. Perhaps sometimes, for instance, one has a duty to oneself not to act imprudently. Or perhaps sometimes one has a duty to one's loved ones not to act imprudently. In such cases, there may be a moral requirement to minimize the chances of acting imprudently.

not vindicate aesthetic deference does not mean that it cannot be vindicated in some other way—perhaps sometimes, for instance, it is *morally* important to minimize the risk of aesthetic mistake (as when one is deciding on funding for aesthetic projects)⁵²; perhaps sometimes deferring can be justified instrumentally, for instance as a way of improving one's own relevant aesthetic sensitivities.⁵³ Discussing such other possible vindications of aesthetic deference is not something I can do here. But the point above stands, I think: the vindication that is developed in this paper does not plausibly apply to the aesthetic case (because it is highly unlikely that there is an aesthetic duty to minimize the risk of aesthetic mistake).

So the vindication offered for deferring to the moral expert does not seem to apply to deference to the aesthetic expert, and it is not clear to what extent it applies to deference to the prudential (or more generally, practically rational) expert and the epistemological one. Perhaps this is so because while no norms of aesthetics apply to the appropriate way of proceeding given aesthetic uncertainty, there *are* moral norms that apply to the appropriate way of proceeding given moral uncertainty; there presumably are prudential norms that apply to the appropriate way of proceeding given prudential uncertainty; and it is really not clear whether there are epistemic norms that govern the appropriate response to epistemic uncertainty (that is, not uncertainty regarding some specific *p* but regarding what is sufficient to justify beliefs in general, or some such).

There may be something surprising about this result—that it may be easier to justify (using the kind of argument developed here) moral deference than deference in some other normative domains. Upon reflection, however, I find this result an attractive feature of the vindication above: the practicality of the moral, and even more so its involvement with the important interests of others, makes the requirement to minimize risk of mistake more robust in the case of morality.

IV. THE REMAINING EXPLANATORY CHALLENGE, AND WHAT COUNTS AS A MORAL ACHIEVEMENT

So we should bite the bullet on Williams's example: yes, if the other details are spelled out in the right way, the student should accept the professor's conclusion on his professorial authority—or anyway, on the grounds that the professor is less likely to get things wrong here,

⁵² I thank Sarah McGrath for focusing my attention on this kind of case.

⁵³ A reader suggested the example of deferring to an expert on which recording is better, with the hope of developing, with time and after listening to the better recording, the expert's sensitivities.

and because of the requirement to minimize the risk of (the student's own) wrongdoing, when no other relevant costs are involved.

In a way, this leaves us where we started, for we still need to explain the fishiness of moral deference and expertise.⁵⁴ Why is it, in other words, that moral deference does not seem as legitimate as many other cases of deference?

Recall four results that we already have at our disposal and that can serve as diagnostic hints. First, the problem is not specifically about expertise and deference; rather, it is common to all cases of forming moral judgment on opaque evidence. Second, the problem is not specifically about *moral* deference, and it applies to other normative domains.⁵⁵ Third, in some non-normative cases—like that of deferring to the metaphysical expert—the problem seemed to be related to the thought that there was no *achievement* in forming the judgment by deference. And fourth, we already know that the problem with moral deference is not about action—action on the expert's judgment is sometimes, I have insisted, even *required*—and is probably not epistemic either. What *is* the problem, then?

Return to the Alon case. If I succeed in forming the true moral judgment—that the war is indefensible—on my own, my judgment is an appropriate response to the features that make the war indefensible. Furthermore, given some plausible connections between the making of moral judgments and certain emotional responses,⁵⁶ this may also be an *emotional* achievement: I respond to the wrong-making features of the war in the appropriate way, emotionally and otherwise. But if I have to rely on Alon's expert judgment, I have missed this achievement. Suppose, for instance, that the war is unjust because it involves inflicting harm on innocents that is disproportionate to the war's legitimate aims. Then by coming on my own reasoning

⁵⁴ The discussion of moral understanding and autonomy may supply some explanatory relief: if, as I have conceded, it is better to do the right thing autonomously and with moral understanding than from deference, this already explains some of our unease with moral deference. Also, perhaps we can deal with the examples where moral deference seems problematic piecemeal, offering different explanations for different examples. This is Sliwa's ("In Defense of Moral Testimony") strategy, and though much of what she says seems right to me, I want more—at least if other things are equal, it would be preferable on methodological grounds to have a unified explanation of Williams's discomfort, one that applies wherever the discomfort applies. The one I am about to offer is of this kind.

⁵⁵ In the previous section I argued that my vindication of relying on moral expertise did not generalize neatly to other normative domains. But the *problem* still does.

⁵⁶ Whether these relations are necessary or empirical is another important and controversial metaethical issue bypassed here. For my defense of the claim that empirical connections are sufficient, and that therefore there is no problem here for the robust realist, see David Enoch, *Taking Morality Seriously: A Defense of Robust Realism* (Oxford: University Press, 2011), pp. 247–59.

to view the war as unjust, and by acting on that judgment, I have responded appropriately to the war effort's inflicting disproportionately grievous harm on innocents. But if I rely on Alon's judgment, I have not responded well to *this* feature of the war. I may have responded well to some other relevant features—namely, to the uncertainty, and the fact that acting on my own judgment here would amount to subjecting others to unjustifiably great risks of being wronged. And *this* may be an achievement—as evidenced by the fact that, among those who fail to respond to the war's initial wrong-making features, we should think more highly of the person who responds well to the uncertainty by deferring to the expert. But still, I have missed the appropriate response to the features that make the war unjust, and so the primary features which make voting against funding the war morally obligatory.

This, then, is what is fishy about moral deference. It is not that it is unjustified as grounds for action—it *is* so justified, indeed, sometimes even required. It is not that it is epistemically unsound—here as elsewhere, expertise can ground knowledge, or help achieve any other epistemic status. It is that forming a moral judgment by deference and then acting on it is much less of a *moral achievement* than forming the true judgment without deference, because it does not constitute the appropriate response to the morally relevant features of the case.⁵⁷ Moral deference indicates failure to achieve something we are expected to achieve. This, perhaps, is the difference compared to other cases of deference—there too is the absence of an achievement, but not always the expectation to so achieve, and so not always also the failure.⁵⁸

The emotional achievement is important here, I think. After all, you may wonder why deferring to Alon's judgment about the war does not amount to an appropriate—if indirect—response to the war effort's inflicting disproportionately great harm on innocents.

⁵⁷ Notice that the point in the text here can be put in a more scalar way, relating not so much to what is or is not an achievement but to greater and smaller achievements, so it can accommodate intuitions not just about moral deference being unjustified but also about it taking more to justify moral deference in some cases than in others. I thank Claire Kirwin for this point.

There is another complication here. We can distinguish between cases in which one had an opportunity to respond correctly to the morally relevant features of the situation but failed so to respond, and cases where one did not even have such an opportunity. Moral deference seems suspicious—though perhaps less suspicious—in the latter as well as in the former. But the explanation in the text does not seem to work as well here, because of the absence of an opportunity to respond appropriately to the morally relevant features of the case. Still, related stories do seem to apply, as the one that follows in the text about the typical emotional responses being *de re* rather than *de dicto*.

⁵⁸ I thank Grant Rozeboom and John Goldberg for help with this point.

In a way, I think it does. But it is not appropriate *enough*, partly because there is something so cold about this response. The response is, as it were, too intellectual, having to do with the attempt to minimize risk of wrongdoing. One is tempted to say, “Don’t you just *see* that it’s wrong? Aren’t you enraged by the harm it inflicts on innocents?” And the complaint seems to be in place even if, by deferring to Alon, I end up voting against funding the war. Now, we may have the emotional responses typical to morality also to those who fail to defer when they should. For instance, the victim’s family member may be (justifiably) enraged with my refusal to defer to Alon about the moral status of the war. But such emotional responses are atypical. Typically, the emotional responses that are closely connected with moral judgments are responses to the morally wrong (say) *de re*, and not merely *de dicto*, that is, under the description “morally wrong.” There need be nothing wrong with such emotional responses to the wrong *de dicto*, or to the second-order wrong of subjecting others to unacceptably high risk of being wronged. But this is certainly not the typical case.⁵⁹ This is why there seems to be something objectionably cold about a moral judgment that is based solely on deference to the expert.⁶⁰

Recall our brief discussion of intentions in section III.3. There I insisted that if the role of intentions is to rule out moral deference, it must be shown that the intention to minimize the risk of (my) doing wrong is not a legitimate intention. It still seems to me that this is a legitimate intention, one that the virtuous will sometimes have. But note that it is far from ideal. The right intention to have in the war case is the intention (perhaps roughly) not to inflict disproportionately great harm on innocents; *this* is the intention with

⁵⁹ Thus, I do not think that caring about the moral *de dicto* is at all objectionably fetishistic. See Michael Smith’s claims to the contrary and my response, in Smith, “In Defense of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord,” *Ethics*, cviii, 1 (October 1997): 84–119, at pp. 112–17; following on Smith, *The Moral Problem* (Oxford: Blackwell, 1994); Enoch, *Taking Morality Seriously*, pp. 255–56, and the references there. See also Ron Aboudi, “Why Do the Virtuous Reevaluate Their Moral Beliefs?,” unpublished manuscript. But I certainly agree that *only* or even *mainly* caring about the moral *de dicto* is both atypical and objectionable. McGrath (“Skepticism about Moral Expertise as a Puzzle for Moral Realism,” p. 135n46) refers to Smith on fetishism in a closely related context.

In seeing the relation between the response to moral uncertainty and the moral fetishism accusation I have been helped by Arnon Keren, “The Risk of Wrongdoing: On the Moral Significance of Moral Uncertainty,” unpublished manuscript.

⁶⁰ There is something similar to this line of thought—though not in its details—in Hills, “Cognitivism about Moral Judgment,” unpublished manuscript. Also related here is McGrath’s (“Skepticism about Moral Expertise as a Puzzle for Moral Realism”) suggestion that what is problematic about moral deference is that it falls short of an ideal of doing the right thing for the right reasons.

which the virtuous would vote against funding the war, and—when I rely on Alon—I lack this intention. True, I have another intention, and a good one too—the intention not to subject others to greater-than-necessary risk of being wronged by me. And given that I lack the previous intention, it is better that I at least have *this* one. Still, I lack the intention it would have been best for me to have. This partly explains the fishiness.

We can reinforce this explanation of the fishiness of moral deference and expertise by doing some amateurish pragmatics. Suppose that I base my moral judgment about the war's status on Alon's judgment, and I proceed to criticize, and indeed blame, those supporting the war. At the very least, there is something odd about such criticism and blame. The oddity may be explained by the fact that such judgments are typically accompanied by emotional responses to the features in virtue of which the war is wrong. This means that my response is atypical, and perhaps also insincere. Even if it is sincere, though—because I respond emotionally to the war's wrongness *de dicto*—this response is sufficiently atypical to be odd.⁶¹

Note how nicely this story, in terms of the absence of the moral achievement, fits with the generalization of the problem to all moral judgments based on opaque evidence. When I form the judgment that turning the trolley in Loop is impermissible based on your heart rate increasing, I am not responding as I should to the features of Loop in virtue of which it is wrong to turn the trolley (if indeed it is). So judging that it is wrong, and indeed refusing to turn the trolley, is less of a moral achievement, is not an appropriate response to the wrong-making features of turning the trolley, and is (typically) less of an emotional achievement than in cases where the relevant moral judgment is based on transparent evidence.

Notice also that this story nicely explains the similar phenomenon in the case of other normative judgments. If one-boxing is the way to go, then it is the response called for by the first-order features of the Newcomb case. If I have to rely on the prudential expert in order to form the judgment that I should not open the second box, and then indeed to avoid opening it, then this is less of a prudential achievement on my part. Similarly, I think, for the epistemological case. And the aesthetic case seems to me especially powerful here: if Mozart is indeed better than Bach, then the relevant aesthetic achievement is constituted by responding appropriately to the features—whatever

⁶¹ For a discussion of such implicatures in the aesthetic context, see Robson ("Aesthetic Testimony," p. 4) and the references there.

exactly they are—of Mozart's music that make it so great. This achievement is absent when I form the judgment based on the expert. Also, assuming that the relation to some typical emotional responses is a feature of the normative in general, not just of the moral, these points too generalize in the right way. Certainly, if having formed the comparative aesthetic judgment by deferring to an expert, I then proceed to emphatically assert how much greater Mozart is than Bach, to criticize those who believe otherwise, and so on, this would be odd at the very least.⁶² And the oddity may be explained by the fact that the relevant emotional responses—typically, to the beautiful *de re*, not merely *de dicto*—are absent here.

V. CONCLUSION

We have a rather full picture now. Often, moral deference is problematic for reasons that in this paper I have set aside by stipulation (for instance, worries about the abuse of power, worries about disagreement among the purported experts, and worries about how to recognize the experts). Still, at times—as in the Alon example—deference is not only permissible but in fact required (in the given-my-evidence sense), and this regardless of further metaethical details, simply in virtue of the moral requirement not to subject others to unnecessarily great risk of being wronged (by the relevant agent). In those cases too, however, there is something suspicious about moral deference and expertise, and an explanation is needed. The story in terms of the relevant moral achievement—and indeed, also the emotional achievement—involved in responding to the moral *de re* and not merely *de dicto* supplies such an explanation, and it does so in a way that nicely accommodates related phenomena, like the scope of the suspicion (applying, as it does, also to other normative contexts and to other opaque evidence).

Often, then, deferring to the moral expert is an indication of some moral imperfection. And acting morally (or forming true moral judgments) is less of an achievement when done by deference to the expert. This is what explains the suspiciousness of moral deference and expertise. Still, as with all imperfections, once they are there, it is better to acknowledge them and act in a way that takes their existence into account. Deferring to a moral expert is sometimes a way of doing just that, and when it is, this is the way to go.

DAVID ENOCH

The Hebrew University of Jerusalem

⁶² All of this is true, it seems to me, even if for some other reason I should defer in the aesthetic case. See the brief examples of such other reasons in section III.4 above.