Moral Deference and Authentic Interaction[1]

Knut Olav Skarsaune

Penultimate draft, for *The Journal of Philosophy* (2016)

For official version see:
https://www.pdcnet.org//pdc/bvdb.nsf/purchase?openform&fp=jphil&id=jphil_2016_0113_0007_0346_0357&onlyautologin=true

Most of our beliefs rest on testimony. We read books, go to school, watch the evening news; and with the right sort of book, teacher and broadcaster, this is a responsible way to form beliefs. And it is rational to *act* on the basis of such beliefs; to accost a reliable-looking native, when lost in a foreign city, and go where she points.

But there is something odd about taking *moral* beliefs from other people, or guiding one's actions by someone else's moral compass. Of course, there is nothing wrong in seeking moral advice, if our advisor *gives reasons* for her recommendations, enabling us to judge for ourselves. It can also be fine to trust someone else's moral judgment, at least provisionally, if they have *empirical* information we lack, and there is no time to explain. But most people report unease when presented with cases where an agent rests a moral belief or a moral decision on the say-so of someone else, simply because she takes the other to be a better moral judge.

In "A Defense of Moral Deference," David Enoch begins by granting that there seems to be "something fishy" about such cases.[2] Nevertheless, the paper is a brief for so-called "optimism" about moral testimony, the view that it *is* legitimate to use it. What Enoch defends is primarily the legitimacy of *acting* on the basis of moral testimony (he is less committal about the parallel view concerning belief formation), and to mark this he uses the term 'moral deference' for the thing he is defending. Moral deference, then, is to act in the way someone else says is right, not because she has empirical information that you lack, but simply because she says the act is right

---

[1] Many thanks to Kåre Letrud, Terje Ødegård, Andreas Carlsson, Caj Strandberg and participants at the Ruth Barkan Markus Philosophical Society at NYU for helpful comments.
Work on this paper was supported by the Norwegian Research Council, grant #240069.

[2] David Enoch, "A Defense of Moral Deference," *The Journal of Philosophy*, vol. CXI, no.5 (May 2014): 229-258.

and you think she is a better judge than yourself on the relevant moral issue.

Enoch offers two main arguments. The first is a direct defence of moral deference, namely that, in the relevant sorts of case, deferring minimizes the risk of subjecting other people to moral wrongdoing. The second is an objection to the view that there are reasons having to do with moral understanding or autonomy against deferring. Enoch's objection is that, even granting that understanding and autonomy are important values, they are not the kind of values that can give reason to subject others to a greater risk of being wronged. The third main element in Enoch's article is a deflecting diagnosis of the negative intuitions many have about cases of moral deference. The source of the fishy smell, he thinks, is not the deference itself, but the prior lack of moral achievement that places the agent in need of it.

In this note, I will first briefly present Enoch's two main arguments, to which I am highly sympathetic. The reason he gives to defer is a good one, and he is right that the values of moral understanding and autonomy, properly understood, do not provide reason to avoid moral deference. But I will argue that we should nevertheless temper our optimism about moral deference far more than Enoch does. There is a different value, which Enoch does not consider, which escapes his negative argument, and which must be balanced against the reason he identifies in favour of deference. This is the value of authentic interaction, of acting in a way that communicates our own understanding of the reasons in play. Ultimately, the point of authentic interaction is that it allows us to get to know and engage one another. This value should sometimes take priority over minimizing the risk of wrongdoing.

I. ENOCH'S DEFENSE OF MORAL DEFERENCE

A common reason for *pessimism*[3] about moral deference is simply the difficulty involved in knowing who to trust. There are no easily identifiable moral experts. There are moral philosophers, of course, and seasoned elders, but they disagree amongst themselves just as much as the rest of us. The only way to know which of them to listen to would seem to be to do moral enquiry yourself, in order to figure out who gives the *correct* advice. But once you have done that, you will

[3] That is, the view that there is something inherently problematic about moral deference. There is really a continuum of possible views here, with perfect optimism (there is nothing at all objectionable about moral deference) at one end and perfect pessimism (moral deference is never legitimate) at the other.

in effect be their epistemic peer. In other words, you will either not know whose testimony is reliable, or you will not need it.[4]

Enoch is non-committal about how good he thinks this point is in general. Instead he argues by example that there will in any case be exceptions, cases where we *do* have reason to think that someone else is more reliable than us on some moral issue (and not because they have more empirical information). His main example is this: after initially disagreeing with his friend Alon about the moral justifiability of military actions taken by their country, he later comes to think that Alon was right from the start, and to regret his own, emotionally driven initial response. This pattern repeats itself several times. Then, another round of hostilities erupts, and again Enoch and Alon make opposite moral assessments. In light of their track record, it seems that Enoch has strong reason to believe that Alon is more likely to be right.

Enoch's positive argument for moral deference is to consider what we should *do*, in a case such as this, where we believe, and with good reason, that someone else is more reliable on some moral issue. A predicament of the relevant kind arises, and we are involved in it as agents. As we understand the reasons in play, one course of action seems right, but our advisor says another is. By our own lights, since we believe she is more reliable than us, acting according to our own moral understanding brings with it a greater risk of acting wrongly. So by our own lights, we would be subjecting the people affected by our action to *greater risk of being wronged*. Which is an eminent reason not to act according to our own moral understanding, but rather according to the advice we have reason to trust.

Enoch's second argument is an objection to a different kind of pessimist view. Views of this kind grant that there are cases where we will have a better chance of acting rightly if we trust someone else's moral judgment over our own. Even so, it is not given that we ought to defer in such cases, because there is *another substantive value* in play, besides acting rightly, which is undermined by deferring. Hence, increasing the odds of acting rightly must be weighed against the sacrifice of this other value. The most common values cast in this role are moral *autonomy*[5] and *understanding*.[6] Though different, these views

---

[4] Paulina Sliwa, "In Defence of Moral Testimony," *Philosophical Studies*, vol. CLVIII, no. 2 (March 2012): 175-195; Sarah McGrath, "The Puzzle of Pure Moral Deference," in *Philosophical Perspectives*, vol. XXIII, no. 1 (December 2009): 321-344; Sarah McGrath, "Skepticism About Moral Expertise as a Puzzle for Moral Realism," *The Journal of Philosophy*, vol. CVIII, no.3 (March 2011): 111-137.

[5] Julia Driver, "Autonomy and the Asymmetry Problem for Moral Expertise," *Philosophical Studies*, vol. CXXVIII, no. 3 (April 2006): 619-644

are variations on a theme; adult human beings have a responsibility to think for themselves, seek moral understanding, and act on it. Belief and action based on moral deference are defective on that score. Perhaps we all-things-considered ought to defer, sometimes, but only if the added likelihood of acting rightly makes up for the loss in autonomy or acting-out-of-understanding.

Enoch's rejection of these views begins with the following analogy. Compassion is good: other things equal, a world with more compassion is better than a world with less of it. Even so, it is never right to cause or allow more suffering in order to generate more compassion. Even if a small amount of suffering could buy a big amount of compassion, it is a deep confusion to think their respective values can be traded off against each other. "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... Someone willing to tolerate greater suffering ... merely in order to work on his (or anyone else's) compassion ... is acting wrongly, and most clearly uncompassionately."[7]

Likewise, Enoch argues, it is confused to think the value of moral understanding or autonomy could provide reason to take a greater risk of acting wrongly. Even if they are important values, they should not be pursued *at other people's expense*. They may give us reason to sacrifice other desirable activities in order to find time for moral reflection, and so on, but they cannot provide reason to subject other people to greater risk of being wronged. Thus, there is *not even a tradeoff* to consider here. Such an attitude (balancing a greater risk of acting wrongly against the pursuit of moral understanding or autonomy) is inherently a mistaken response to these values.

These arguments lead Enoch to a stout optimism about moral deference. Epistemologically, there is no reason in principle why testimony should be less able to confer true belief about moral issues than other issues. Occasions where we have good reason to think someone else more reliable than ourselves may be rarer in the moral domain, but Enoch thinks there is a fairly straightforward way such occasions can come about. And when they do, there is no general or

---

[6] Philip Nickel, "Moral Testimony and Its Authority," *Ethical Theory and Moral Practice*, vol. iv, no. 3 (September 2001): 253-266; Robert Hopkins, "What is Wrong with Moral Testimony?," *Philosophy and Phenomenological Research*, vol. lxxiv, no. 3 (May 2007): 611-634; Alison Hills, "Moral Testimony and Moral Epistemology," *Ethics*, vol cxx, no. 1 (October 2009): 94-137.

[7] Enoch, "A Defense of Moral Deference," *op. cit.,* pp. 248-49.

systematic *normative* reason against deferring. In such cases, deference is normally the way to go.

Only normally, because the requirement to minimize the risk of wrongdoing comes with an other-things-equal clause. Enoch discusses two ways this clause can be triggered. The first is that it is sometimes reasonable to err on the side of safety, by preferring a greater risk of a smaller wrong over a smaller risk of a greater wrong. The second is that there could be *instrumental* reasons against deference, in that acting on one's own judgment might enable moral learning, making the agent more likely to act rightly in the future. Enoch thinks it will be hard to say much in general about how such instrumental considerations play out. But he denies that there is any general, non-instrumental value that can stand up to the requirement to minimize the risk of wrongdoing. After rejecting, as we saw above, the view that autonomy or understanding can play such a role, he continues:

> 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.[8]

I agree with Enoch all the way up to this quote. But in the rest of this paper, I will argue that there is a value that systematically provides reason not to defer, and which even outweighs the requirement to minimize the risk of wrongdoing, in a subclass of cases I will identify.

II. A PATIENT-CENTERED REASON NOT TO DEFER

Enoch's argument usefully shifts our attention from the agent to the *patients* of her action, the people affected by it. Once we take this perspective, it does indeed seem misguided to place a burden on them in order for the agent to develop her moral autonomy or understanding. But it does not follow that there is anything wrong with the *form* of explanation the second type of pessimists offer, that is, pointing to a substantive value that has anti-deference implications. Even if we grant that Enoch's argument is sound, what it shows is that the particular values discussed, moral autonomy and understanding, cannot provide reason not to defer if that will raise the risk of acting wrongly. But there could be other values that can. Before we embrace a general optimism about moral deference, we

[8] *Ibid.*, p. 250.

should pause to see if there are any other values, *that the patients have reason to care about*, that count against the agent deferring. In other words, we should investigate whether there are any values for the sake of which it can ever make sense to welcome a greater risk of being treated wrongly.

That might seem like a strange suggestion. What value could possibly give us reason to welcome a greater risk of being wronged? In this note, I will investigate a surprisingly potent value of this kind. I will suggest that there is a tradeoff, after all, between this value and the requirement to minimize the risk of wrongdoing. The fact that it has seemed to so many writers that there is a problem with moral deference having to do with autonomy/understanding suggests that we should look for a value similar or related to these, but that engages the interests of the patients. I will suggest that *authentic interaction* is just such a value.

I begin with an extreme thought-example. Now, reasonable people can disagree about the proper role of far-fetched thought-experiments in ethical theory. But one thing they surely can do is bring values to our attention that we have underestimated because we take them for granted. The ethos of personal autonomy and responsibility is entrenched in our culture. Moral deference is not a significant ingredient in my social experience, and it may not be in yours either. And familiar societies where it is, say strongly religious or traditionalist communities, have other, distracting features, chiefly perhaps that their moral authorities tend to disseminate values we disagree with. So it may be useful to consider what would be lost if people always deferred their moral decisions, but to a more reliable authority. With that preamble, consider the following scenario:[9]

You take a job in a city far away, and at first it seems like utopia. People are honest and fair; they are friendly and hospitable; the trains run on time. You notice that they fiddle with their smartphones a lot, but you assume they are simply avid users of social media. Eventually, however, you discover what they are really doing: whenever they have to interact with another person, they consult the app *Google Morals*, and do what it advises. Google has produced this app by feeding every article published in peer-reviewed ethics journals into a "big data" algorithm, which generates instructions for how to behave in any social interaction. We need not stipulate that Google Morals tracks the moral truth perfectly; only that it does so

[9] Adapted from Robert J. Howell, "Google Morals, Virtue, and the Asymmetry of Moral Deference," *Nous*, vol. XLVIII, no. 3 (September 2014): 389-415.

with considerable reliability, as witnessed by the utopian character of the society it generates.

What you discover, then, is that your new neighbours may have no understanding of why they treat you the way they do. In fact, let us stipulate that they do not. The colleague who warmly invited you to dinner was not responding to the fact that you are new in town and might appreciate an invitation; she was just following instructions from her smartphone. So was the neighbour who helped you figure out how to work your fuse box, and so on and so forth.

A natural reaction to this revelation would be one of deep disappointment and loneliness. It is in a relevant sense not possible to *interact* with *people* in this society, only officers of Google Morals. And the problem is independent of the details of the moral code to which they defer. If you distrust the bulk of articles published in peer-reviewed ethics journals, change the app in the example to one based on your favourite moral theory, say *Google Kant*, or *Google Singer*. Perhaps living in such a society would have its advantages, but the point is that it would also have disadvantages. You would lose out on a value that matters *to you* if you stay there, even if you do not partake in the deferential practice, and even if it minimizes the risk that you will be treated wrongly.

Of course, our negative reaction to this example is overdetermined. There are all kinds of things wrong with these people (one of them being that they lack a certain kind of moral achievement, which Enoch describes in his article). I do not take the fishy smell here to prove anything. All I hope to have shown is that it is not crazy to think that moral deference on the part of other people can bring with it some kind of loss to us, even if the risk of wrongdoing is minimized. In the next section, I will discuss whether anything like that applies in everyday moral decisions. But first it will be useful to try to isolate the value that seems to be at stake. I have called it 'authentic interaction', but can we make it more precise?

Authenticity is somewhat different from autonomy. Suppose we understand autonomous behaviour as behaviour guided by the agent's own understanding of the reasons in play (and that these reasons do not include her being coerced, blackmailed, etc., but let us put such things aside). In that case, deceptive behaviour can be autonomous (if the agent's own assessment is that what she has most reason to do is to deceive you). But deceptive behaviour is intuitively not authentic. We should demand of authentic behaviour that it be both *guided by* and *expressive of* the agent's own understanding of the reasons in play. This is hardly a razor-sharp definition, but it will do for our purposes.

Why care whether other people's behaviour has this property? Thinking about what social life would be like in its total absence brings to light two ways it can matter to us. The first is that it contributes to make a certain kind of knowledge available. When other people's behaviour is directed by a source outside of themselves, it does not give us much of a clue about who they are, what they are like – apart from the fact that they are somebody who defers to such-and-such advisor or authority. Conversely, when people's behaviour *is* guided by their own understanding of the reasons in play, they tend to reveal something quite important about themselves – namely what they see as a reason for what, and how they weigh reasons of different kinds. We learn how they weigh their own interests against those of other people, how they weigh happiness against achievement, loyalty against honesty. These are things we want to know about people, especially if we have or contemplate opening some kind of personal relationship with them.

Of course, moral deference is not the only way this kind of knowledge can be subverted. Deceptive behaviour has already been mentioned; there is also behaviour that flows from weakness of will or irrational impulse. Or we might misunderstand each other because we have different empirical beliefs. So the absence of deference is not a sufficient condition for revealing one's normative judgment. But it seems to be a necessary condition, for other people's behaviour to reveal this part of who they are, that they act out of their own understanding of the reasons in play.

The second way we would miss authentic behaviour if it were gone is that we would lose a certain kind of personal engagement with each other. If the person you are interacting with is not acting out of her own normative judgment, but rather following instructions from some advisor, there is a sense in which you are not fully interacting *with her as an individual* at all. The most important features of her behaviour towards you do not flow from her, but from the advisor. Her behaviour is to a large degree independent (counterfactually) of who she is – if she were different, her behaviour would remain the same (so long as we hold fixed her motivation to do the right thing and her belief that the way to achieve this is to defer to the advisor). On the other hand, her behaviour is strongly counterfactually sensitive to what the advisor is like. This is what generates the sense, I think, that you would be interacting with the advisor as much as with her.

One might object that our topic is only *moral* deference, not deference about every aspect of how to behave towards each other. Not every detail of a social interaction is morally significant, so there would remain some space for individual differences in carrying out the same

moral instruction, as it were. For example, the colleague might invite you to dinner by email or in person, she might throw in a joke or a quip about the weather, she might invite you home or to a restaurant. This is true, but I do not think it is enough to enable the kind of personal engagement at issue. For such differences seem to matter much less to how you experience the interaction than if we vary the morally relevant features – if she had briskly ignored you, say, or dumped her committee work on you. In general, the morally relevant features of our behaviour towards each other are highly socially salient, they set the tone of our interactions. If these were all guided by a source outside of the people we interacted with, I suggest we would feel cut off from them in a rather uncomfortable way.

To sum up, if we go to the extreme and consider a social world dominated by moral deference, it seems we would lose two things that we care about: knowledge of how the people around us are disposed to weigh different kinds of reasons, and the sense that our interactions involve a certain kind of personal engagement with them.

III. A TRADEOFF BETWEEN AUTHENTICITY AND PROBABILITY OF RIGHT ACTION

In the previous section, I identified two potential reasons to want others to act authentically towards us. Now I will ask whether, and if so when, these reasons apply in normal social life. Suppose they sometimes do apply, in the kind of case Enoch describes, where the agent can minimize the risk of acting wrongly by deferring to a moral advisor. In that case, there would from our perspective as patients be a tradeoff between our reasons to want the agent to act authentically, and our obvious reasons to want her to minimize the risk of treating us wrongly. I will suggest that there is indeed such a tradeoff, and I will identify two features of social interactions that seem to affect the balance of this tradeoff in a systematic way, namely the degree to which the relation between the parties is a personal one, and the gravity of the practical interests at stake.

Let us begin with an example where I think we will want the agent to defer. We are involved in some kerfuffle, and a police officer arrives on the scene. She is not great at moral reasoning, and she knows it, but precisely for that reason she has memorized her department's Ethical Code of Conduct. If we stipulate that her odds of resolving the dispute in a fair way will be better if she uses the Code rather than her own normative judgment, which should we want her to do? Clearly, we want her to follow the guidelines. In a situation such as this, we simply do not care about learning how she is disposed to

weigh different kinds of reasons. And we are not interested in a personal engagement with her. But we greatly care that she minimizes the risk of acting wrongly.

This generalises to other cases where the other person is acting in some professional or official capacity. When dealing with police, doctors, or landlords, we do not care whether they act authentically or not. What we care about is that they treat us correctly. This fits well with our practice of having institutionalized ethical codes for professions where the practical stakes are high (so that we care greatly about minimizing the risk of wrongdoing), and their interactions with us are of an impersonal nature (so that we do not care whether they act authentically).

But consider next a different sort of case. A close friend finds herself in a moral predicament involving you. Should she be frank about what she thinks about your new romantic attachment, or tell a white lie? Her own understanding of the reasons in play is that she should do one thing, but her moral advisor recommends the opposite. Let us stipulate that her advisor is more likely to be right. Should she act on her own normative judgment, or defer to the advisor?

When the practical stakes are moderate, I think we want our friends to act according to their own normative judgment, even at the cost of a greater risk of acting wrongly. It would be disappointing to be told, if we ask a friend to explain why she treated us the way she did: “well, it didn't seem that way to me, but according to Smith, who you know is an expert in these things, it was the right thing to do.” We would rather have a couple of honest mistakes from our friends, it seems to me, than a string of answers like this. (Contrast, again, with a doctor's reply: “well, it didn't seem that way to me, but according to the hospital's ethical guidelines, it was the right thing to do”.)

These examples suggest that in an impersonal relation and a decision where the practical stakes are high, we prefer deference (when that improves the odds of right action),[10] but in a personal relation and a low stakes decision, we prefer authenticity. What about mixed cases? Consider a high stakes decision in a personal relation. Suppose you break your spine and your spouse has to decide on your behalf whether to do nothing, which will leave you paralysed from the neck, or go through with a risky surgery which may cure you but may also kill you. The doctors say that medically, it is an even decision. My inclination is that your spouse should follow her own normative

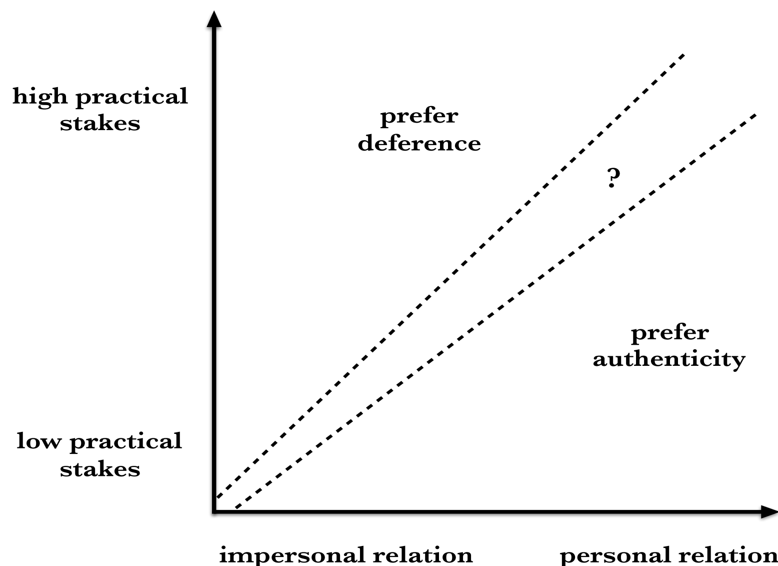
---

[10] Enoch's two main examples are extreme cases of this kind: whether to vote for or against funding a war, as a member of the relevant government body; and how to decide a difficult private-law case, as a judge. Enoch, ”A Defense of Moral Deference,” *op. cit.,* pp. 230, 242-45.

judgment, even if a more reliable moral advisor disagrees, but a significant proportion of the people I ask make the opposite judgment about this case.

Likewise, it is hard to make confident judgments about low stakes, impersonal cases. Suppose for example a colleague at work is an average fellow of average decency. Would you rather he defer to some ethical guidelines, if this would lead to a moderate improvement, on average, in how fairly he treats you? My inclination is again to put a significant price on authenticity, even in relations like co-worker or neighbour, but again some of my informal informants disagree.

The pattern that seems to emerge is that our preferences tilt toward authenticity the more personal the relation and toward probability of right action the less personal the relation. They tilt towards authenticity when the practical stakes are low and towards probability of right action when the stakes are high. When a case is both personal and low stakes, or both impersonal and high stakes, intuitions are pretty clear; in mixed cases they are less firm (see figure).



This is precisely the pattern one would expect if we find ourselves balancing the value of authentic interaction, on the one hand, against increasing the odds of right action, on the other.

Suppose that is indeed our situation. In that case we can resist Enoch's argument, in the kind of cases where we would prefer authentic behaviour from others, when the roles are reversed and we are the agent. I can see no reason not to apply the Golden Rule here. Enoch's point, recall, was that even if autonomy/understanding are important values, it is confused to think they could give us reason to

subject others to a greater risk of being wronged. But since authenticity on our part has value *to others*, just as minimizing the risk of wrongdoing has value to them, it is not confused to balance these values against each other. In situations where we would have most reason, all things considered, to want authenticity from others, I submit that the thing to do is to behave authentically towards them.

IV. CONCLUSION

Prior to reading Enoch's article, I assumed that moral autonomy was the key to the issue of moral deference. I thought our duty to be independent moral thinkers and agents could motivate at least a moderately pessimistic view. But Enoch has set me straight. My autonomy, in this regard, is my business, but how I treat other people is not. If there is a general reason to be pessimistic about moral deference, it must have to do with our relations to each other.

Pursuing the line of thought that Enoch opens up reveals that there is indeed such a reason. We want to know what the people around us are like, and we want to engage with *them*, not their moral advisors. Sometimes, I have argued, we care so much about this that we are willing to accept a greater risk that they will do us wrong. And in that kind of case, when I would welcome a greater risk of you treating me wrongly, I should also be willing to take a greater risk of treating you wrongly.

This is in some ways a surprising result. It seems like a self-important thing to say: "Even though my moral advisor, who is more likely to be right, says that I should do something else, I am going to do what seems right to me, because that will give you knowledge about how I weigh reasons, and will enable personal engagement between us." But if the arguments above are right, then even if that would always be an obnoxious thing to *say*, it is sometimes the right thing to think.