ARTICLE

Deference and Ideals of Practical Agency

Jonathan Knutzen

Institute of Philosophy, University of Graz, Austria
Email: jonathan.knutzen@gmail.com

Abstract
This paper develops a moderate pessimist account of moral deference. I argue that while some pessimist explanations of the puzzle of moral deference have been misguided in matters of detail, they nevertheless share an important insight, namely that there is a justified moral agency ideal grounded in *pro tanto* reasons against moral deference. This thought is unpacked in terms of a set of values associated with the practice of morality. I conclude by suggesting that the solution to the puzzle of moral deference developed here gives us a plausible recipe for generalizing to certain cases of nonmoral deference as well.

Introduction

The key question in the literature on moral deference is whether there is anything problematic about believing moral claims exclusively on the basis of testimony. Pessimists argue that there is. They point to a variety of features in virtue of which moral deference is plausibly problematic. On the best-known suggestion, moral deference is said to limit understanding, which in turn compromises something else of ethical value, like virtue or the moral worth of actions.[1] In response, optimists attempt to show why moral deference can be a good thing. Some optimists emphasize the fact that relying on moral testimony can improve moral decision-making and minimize one's risk of moral wrongdoing.[2] Others argue that relying on moral testimony can bring about additional moral goods, like fostering trust in intimate relationships and empowering persons who have been subject to unjust credibility deficits.[3]

Yet there is something odd about the debate, at least as it is sometimes framed. Conflict emerges only when one presses each side to an extreme. The best-known pessimist, Alison Hills (2009), claims that normal adults have "strong reasons neither to trust moral testimony nor to defer to moral experts" (98), even though she later adds the caveat that "there probably are situations" in which one could have most reason to defer (124). This makes it sound like she accepts a fairly strong and sweeping conclusion: aside from certain exceptional cases, relying on moral testimony is almost never on balance okay. Meanwhile, the best-known optimist, David Enoch (2014), argues that there are powerful general reasons in favor of moral deference given by the duty to minimize one's risk of wrongdoing and that there are no powerful countervailing reasons on the other side. This suggests an equally strong and sweeping claim to opposite effect: as long as one reasonably believes reliance on testimony will improve the accuracy of one's moral beliefs, relying on testimony is more or less always mandatory.

---

[1]See Callahan (2018), Crisp (2014), Hills (2009), Howell (2014), McGrath (2011), Nickel (2001).
[2]See Enoch (2014), Jones (1999), Sliwa (2012).
[3]See McShane (2018), Wiland (2017).

Each extreme is implausible. Optimists are surely right that reliance on moral testimony can be morally salutary. There can be both remedial and nonremedial value to relying on moral testimony, and sometimes one will have decisive reasons to do so. The pro-deference rationale identified by Enoch is especially compelling: if one reasonably believes someone else is better positioned to reach a justified verdict, refusing to defer can amount to unjustified recklessness.

But pessimists are surely right that there is something problematic about moral deference all the same. Consider two asymmetries:

Asymmetry 1: Accepting moral claims solely on the basis of testimony is prima facie problematic, whereas it is prima facie unproblematic to accept ordinary factual or scientific claims in the same way.

Asymmetry 2: It is generally worse to *sustain* moral beliefs on the basis of testimony than it is to *form* moral beliefs on the basis of testimony. The same is not generally true of ordinary factual and scientific claims.

Both asymmetries are plausible. If I learn from you about arctic foxes in Kamchatka, there is nothing even prima facie troubling about this. When it comes to moral content, however, the situation is different. If you tell me that hunting arctic foxes to extinction is morally wrong and I come to believe this solely on the basis of your testimony, something does seem prima facie troubling about this. To be sure, I might go on to think about the question for myself and come to accept the same claim on the basis of moral reasons. This would change the situation: nothing would seem particularly problematic anymore. But suppose I sustain my moral belief indefinitely on the basis of your testimony. If twenty years hence I still believe that hunting arctic foxes to extinction would be morally wrong purely on the basis of your testimony, there does seem to be something disappointing about this. By contrast, when it comes to ordinary factual information, as long as my trust is appropriately calibrated to the trustworthiness of my sources, I might indefinitely sustain belief without acquiring testimony-independent evidence for the claims I accept. Nothing seems problematic if twenty years from now I still cite you as the reason for believing certain things about arctic foxes—about how far they range, about their hunting and mating habits, about the color-phases of their coat, and so on. The difference is not that in one case justification for belief runs out: in neither case does the original justification come time-stamped with an expiration date. Rather, in the moral case, I seem to be failing in a way that has nothing to do with my justification for belief: I seem to be violating an expectation that I reflect on, and take ownership of, my moral beliefs in a way that is inconsistent with continuing to believe on the basis of testimony alone.

Pessimists can explain these intuitive data. Since pessimists claim that moral deference is *generally* problematic, they have resources for explaining both why moral deference should seem prima facie suspect and why there is residual normative "pressure" to not go on sustaining moral belief indefinitely if one can avoid doing so. Optimists, by contrast, cannot so readily explain these data. If optimists go to the extreme and say there is nothing *generally* problematic about moral deference, they have impoverished resources for explaining these asymmetries.

It appears, then, that there is truth on both sides of the debate. The following intermediate position looks attractive: we should avoid claiming that moral deference is almost always bad or wrong just as we should avoid claiming that intellectual outsourcing in the moral life can be laissez-faire. Optimists have nicely captured the thought that we ought sometimes to defer; pessimists have nicely captured the thought that moral outsourcing must be kept in check. We should look for a solution to the puzzle of moral deference that respects both these insights.[4]

This paper develops such a solution. Specifically, it develops a species of what might be called *moderate pessimism* (*cf.* Lewis 2020a). The next section develops an explanation schema for the

---

[4]The attractiveness of such an intermediate approach is emphasized in several recent contributions, e. g., Lewis (2020a), Lord (2018), Mogensen (2017).

problem of moral deference that shows how to think about this sort of *via media* approach to explaining the puzzle. The following sections then consider how to fill in details. The paper concludes by suggesting how the explanatory approach pursued here might fruitfully generalize.

1. An explanation schema

A solution to the puzzle of moral deference needs to explain the above two asymmetries. Each asymmetry captures something important. The first encapsulates the core phenomenon. Intuitively, there is a difference between accepting moral claims purely on the basis of testimony and accepting ordinary and scientific claims in like manner. The literature is replete with terminology to mark this contrast. Unlike deference about ordinary and scientific claims, which raise no special concerns, moral deference is described as "problematic," "fishy," "off-putting," "strange," and "peculiar."[5] The first asymmetry captures this.

The second asymmetry points to a further contrast. Consider Bernard Williams's (1995, 205) example of a student deferring to his teacher on a moral question. In Williams's example, a student who does not follow the force of his professor's reasoning for a particular moral conclusion nevertheless adopts the conclusion on the basis of the professor's authority. Imagine two variants of the case. In the first, the student forms the belief on the basis of testimony while sitting in class, but he goes on to think long and hard about the question later. As he does, the professor's reasoning begins to make more sense and a variety of considerations become clearer to him. The student retains the belief, but now holds it on a new basis. Contrast this with a second version, in which the student does no further work to try to come to appreciate nontestimonial grounds for his belief. Years later, he still holds the same belief on the basis his professor's testimony. The second case seems far more problematic than the first. Both cases involve deference, yet in the second something is more obviously amiss.[6]

In general, how one comes to have moral beliefs in the first place matters far less than how they get sustained. Ordinary and scientific beliefs can be maintained indefinitely on the basis of deference without needing independent grounding. Deference in morality, by contrast, requires further work from the deferring agent. Depending on the case, it might be advisable for the moral student to take notes now and think more about the matter later, but either way, there is something problematic about chronic and unnecessary deference.

A couple caveats are important.[7] First, initial deference is only called for in circumstances of sufficient urgency. If deference isn't urgent, agents can suspend belief and think about the matter further. Second, there must be significant scope restrictions on any principle requiring agents to think for themselves. In particular, agents must have relevant reflective capacities.[8] Third, in deciding how much weight the demand to think for oneself has, the opportunity costs of moral reflection must enter in. Given that moral reflection is demanding, it may be that agents' resources are sometimes better spent devoting themselves to other morally important goals. Finally, there may be cases in which people are predictably worse at deliberating later in their lives than earlier. In such cases, the comparative importance of acquiring versus sustaining moral deference may be flipped.

---

[5]Callahan (2018, 440), Enoch (2014, 229), Howell (2014, 391), McGrath (2011, 111, 115), Mogensen (2017, 261).

[6]The operative notion of sustaining here is the idea of acquiring belief on the basis of testimony that is not subsequently grounded in nontestimonial sources, not the idea of acquiring a nontestimonial belief that is subsequently grounded, in full or in part, through testimonial sources. Thanks to a reviewer for the *Canadian Journal of Philosophy* for prompting me to clarify.

[7]Thanks to a reviewer from the *Canadian Journal of Philosophy* for suggesting several qualifications to my argument.

[8]What if an agent would not discover the truth for herself by making use of her capacities? Proponents of (factive) understanding-centered explanations are committed to treating this type of case as an exception to the rule that agents ought to think for themselves. My view is more complicated. As will become clear, it does not hitch the anti-deference case to success at tracking the reasons that make moral claims true. Nevertheless, since accuracy matters, that case won't consistently favor independent reflection either.

The asymmetry between acquiring and sustaining moral belief has not been emphasized in the literature.[9] Yet it is significant. For one thing, it helps sharpen the contrast between deference about different kinds of content. For another, it makes pessimism more compelling. Focusing on belief acquisition plays into the hands of optimists for, as optimists have amply demonstrated, it is easy to think of cases in which agents should outsource their moral judgment. Such cases, however, are also highly circumscribed: they are not cases in which those who defer are forced to do so indefinitely and they are certainly not cases in which agents are indefinitely off the hook for taking appropriate ownership of their moral beliefs. The most convincing cases given by optimists involve one-off deference. These are cases in which the deferrer can and should go on to later examine her commitments. Moreover, the narrow preoccupation with belief acquisition has made it seem like pessimists must be committed to an incredible and morally dangerous fetishism about moral testimony, as if there were something nearly sacrosanct about how moral beliefs get formed in the first place. But they shouldn't be committed to the idea that merely acquiring moral beliefs via testimony somehow taints them in perpetuity. Suppose I wake in the middle of the night finding that I believe eating meat is wrong. I have no reasons to support this claim; it just occurred to me. If I go on to think about the matter and find that moral reasons best support this claim, then who cares how it arose in the first place? Once we accept that what matters is not so much how moral beliefs first arise but how they get sustained within an individual's psychology, this shifts the debate. If we think chronic and unnecessary moral deference is problematic, as the second asymmetry suggests, we need some explanation for this. We need some explanation of why moral deference cannot be as casually sustained as deference about garden-variety facts, like the mating habits of arctic foxes in Kamchatka. It is this sort of explanation that pessimists are poised to offer.

One more point about the target of explanation. When we are looking to explain a general though not exceptionless pattern, we need to be careful about appealing to individual cases as counterexamples to the pattern. The asymmetries capture general patterns, not claims about the all-things-considered appropriateness of moral deference in one-off cases. This means that the kinds of cases that have been given by optimists need to be handled with care. If the asymmetries are not exceptionless patterns, then they are consistent with pro-deference verdicts in many individual cases.

Having clarified the explanatory target, how is it best explained? There are at least three broad forms of explanation that have been entertained in the literature. The first is epistemic. This form of explanation appeals to epistemic conditions within the domain of morality—that it is difficult to identify experts, that there is rampant disagreement, that everyone is roughly on a par with respect to the facts, and so on.[10] The second is metaphysical. This form of explanation appeals to the nature of moral facts or the constitution of moral commitments—that morality is nonfactual, that moral claims are subjective or relative, that moral claims are partly constituted by sentiment, and the like.[11] The third is broadly value-based. This form of explanation appeals to some ethical value, ideal, or moral principle that speaks against moral deference.

The most common approach to explaining the puzzle of moral deference in the recent literature has been in the last genre. The basic idea is that, because of special features associated with moral content, moral deference has some unique ethical disvalue that deference about ordinary and scientific facts does not have. This seems to me the most promising approach as well. To conserve space, I therefore bracket epistemic and metaphysical explanations of the puzzle (interesting though they are) and focus entirely on value-based explanations. Others have insightfully explored and criticized epistemic and metaphysical explanations elsewhere.[12]

---

[9]It is perhaps implicit in Callahan (2018).

[10]See Cholbi (2007), Davia and Palmira (2015), McGrath (2009, McGrath (2011, 118–20).

[11]See Fletcher (2016, 59–62), McGrath (2011, 121).

[12]See especially Hopkins (2007) and Howell (2014) on epistemic explanations, and McGrath (2011) and Fletcher (2016) on metaphysical explanations.

The best way to think about value-based explanations is in terms of *pro tanto* reasons supporting a default stance for the domain in question. If it is true that agents sometimes ought to defer, then the explanation of what is problematic about moral deference should not be understood in terms of the claim that it is always all-things-considered bad or wrong to defer. Rather, it should be understood in terms of something like an overridable presumption against deference for the moral domain supported by general standing reasons. This presumption sets the warranted default stance—don't defer!—but departures from the default can be both permitted and even required. A number of authors have emphasized that the anti-deference rationale should be understood in terms of *pro tanto* reasons.[13] A plausible way of interpreting what value-based explanations attempt to do is to identify *pro tanto* reasons that could play the role of supporting such a warranted default.

The explanation then works as follows. The puzzle of moral deference is explained in terms of the presence of a warranted default stance against deference within the moral domain, grounded in general standing reasons applying to all competent moral agents. There is no such presumption for ordinary factual or scientific content because there are no reasons supporting a *general* anti-deference presumption with respect to such content. But there *is* such a presumption for moral content. Hence, although people will sometimes have most reason *not* to defer about ordinary factual and scientific questions, and most reason *to* defer about moral questions, it is only in the latter case that a general standing presumption is violated. If that's right, we get an explanation of the special taint that applies to moral deference: there are standing reasons in the background, selectively attaching to moral content, which render moral deference suspect—even if, in fact, it is quite fine in many actual cases to defer. This is the sense in which deference turns out to be "weird," "fishy," "problematic," and so on. It isn't that moral deference is always wrong or bad; it is that moral deference violates a presumption that one ought to think for oneself in moral matters.

The residual effect of *pro tanto* reasons in cases in which they are outweighed is familiar from other kinds of ethical explanation. The imposition of justified harms can require apology or compensation, justified choice of option A over option B may come with warranted regret for unrealized values present in B, and so on. The suggestion here is that *pro tanto* reasons against deference play a similarly explanatory role, even when they are outweighed. Since these reasons speak persistently against moral deference, they explain why, in the absence of stronger counter-vailing reasons, one shouldn't defer (hence, also why casual outsourcing in perpetuity is generally not okay), and why even when an agent ought to defer it is reasonable to feel some residual unease about this (since the reasons support a general default against deference that doesn't disappear just because the circumstances mandate deference).

2. A moral agency ideal

Let's fill in the schema. The best anti-deference rationales for the moral domain have appealed to some ideal of moral agency.[14] That seems the right place to look. In an essay on the aims of moral education, William Frankena (1973, 153) argues that since morality is essentially rational, moral education must be centrally concerned with the importance of teaching reasons. The aim of moral education, says Frankena, is to produce "autonomous moral agents," where this means that the developed moral agent must be able to "make up his own mind." Frankena is clearly on to something here. Hovering in the background of our educational practices is an ideal of mature moral agency as involving a kind of intellectual independence. We expect agents to come to have a relationship to moral norms that differs, say, from their relationship to rules of etiquette or the rules of a game or of an arbitrary convention. When Johnny hits Suzy we don't simply say "don't do that" (end of story). We say things like, "that caused Suzy pain" and "how would you feel if Suzy did that

---

[13]Callahan (2018, 4), Crisp (2014, 134, 137), Hazlett (2014, 19), Howell (2014, 392), Lewis (2020a, 6), Mogensen (2017, 263).
[14]See Callahan (2018), Crisp (2014), Hills (2009), Howell (2014), McGrath (2011), and Nickel (2001).

to you?" And when Suzy asks why she should share her toys with Johnny, a pedagogically sensitive answer would invite her to reflect on the kinds of considerations relevant to the issue.

As children mature, we hope they come to appreciate the reasons supporting the norms they endorse more deeply and exhibit deliberative independence so that they can stand on their own feet in moral matters. Parents and educators typically don't just want their children and pupils to get the answers right but to have a certain relationship to moral norms which includes independence of mind. Think of a child who comes home from school having simply soaked up a moral idea from her teachers or peers but without reflective engagement. A parent might feel there is something disappointing here, even if the moral idea seems right. Conversely, suppose the child comes home with a newly acquired moral belief that seems clearly false to the parent, but which has been reflectively adopted. The parent might worry that the child has acquired mistaken moral beliefs, but not about the child's intellectual independence.

To be sure, deference may play a positive role in moral learning.[15] If we think of deference in a broad sense as including trust-based acquisition of content for which one doesn't (yet) grasp direct justifications, it certainly does. Children are highly attuned to absorbing the norms in their environment. We are cultural animals, strongly biased in favor of imitation and conformity, and wired to assimilate the norms in our local environment. But while this may be an inevitable, indeed salutary, starting point, the ideal embedded in our educational practices suggests it is not an optimal endpoint. That ideal suggests it is important that people should be active moral norm evaluators rather than merely passive moral norm conformers. Developmentally, people's learning of moral norms outstrips any ability to ground, explain, or justify them. Yet as individuals mature, they become better able to grasp the relevant bases of the norms and it makes sense to want them to have a correspondingly deeper and more sophisticated appreciation of the norms. Insofar as moral learning makes use of deference, then, the aim should be to produce agents who no longer rely on it. Children should surely strive to kick off their moral training wheels over time, and parents and educators have a duty to assist them in doing so.

It is this basic ideal of autonomy or intellectual independence, reflected in our educational practices, that is plausibly the source of our intuitions about moral deference. Of course, the ideal of autonomy or intellectual independence only names the problem for which a solution is needed. When we worry about whether our intuitive aversion to moral deference can be justified, it doesn't help to realize that our intuition is driven by an ideal of intellectual independence: we need to know whether that ideal can withstand scrutiny.

The ideal of intellectual independence would be straightforwardly implausible if it implied that one could never rely on others (Fricker 2006). But the ideal need not speak against forms of epistemic dependence that are otherwise salutary, like learning from others and consulting with them. It need only speak against certain *kinds* of intellectual dependence. Paradigmatically, what is ruled out is casual outsourcing of large swaths of one's moral beliefs in perpetuity. There is residual pressure to do one's own homework in moral matters even after one has deferentially acquired moral belief. If that pressure is not illusory, something must speak against abiding moral deference.

What is that something? To account for what is generally problematic about moral deference, a number of authors have suggested explanations of roughly the following sort. The deferring individual doesn't understand the reasons that make true or justify the content she accepts. Since her belief is based purely on the testimony of another, she lacks a deeper grasp of reasons for its truth, reasons which could play the grounding or justifying role vis-à-vis the content in question. To be sure, if her trust is warranted, her belief is sensitive to evidence *that* the content of the belief is true. But this is different from appreciating reasons *why* the claim is true. Understanding why involves precisely such deeper appreciation (Hills 2009). Here, then, is a desirable cognitive good plausibly jeopardized by deference: understanding. Since moral and nonmoral deference alike

---

[15]See Hopkins (2007, 613), Howell (2014, 411), and Hills (2009, 120n41).

shortchange understanding, more is needed. The next layer of explanation is to show why understanding has special value within the moral domain. One suggestion appeals to the importance of understanding for morally worthy action. Since deference bypasses appreciation of reasons that make an action right, this feature limits the moral value of actions flowing from deference.[16] Another suggestion appeals to the role of understanding in moral virtue. The basic idea is that appreciating the moral reasons behind moral judgments is a crucial cognitive and affective ingredient of virtue.[17]

This approach represents the best-known and most promising line about why moral deference is problematic. At an abstract level, the approach is compelling: it locates a plausible cognitive good threatened by deference and shows why that good is morally important. The details, however, are less convincing.

Begin with understanding. On most views, understanding is factive. Alison Hills (2009, 99), who has developed the most sophisticated understanding-centered explanation, accepts this standard assumption. Even on nonfactive accounts, understanding typically involves sophisticated inferential and explanatory abilities (Elgin 2009, 327). Whether or not understanding is factive, then, it is a relatively high-grade cognitive achievement.[18] The trouble is that if one appeals to a high-grade cognitive achievement in order to ground the anti-deference rationale, this produces an anti-deference rationale with implausible variability and nearly nothing to be said against deference in many cases. The strength of one's reasons against deference would seem to depend on the expected value of hitting on the truth or achieving the other cognitively sophisticated components of understanding. Since understanding is hard to come by, this will have two unpalatable consequences. First, it will mean one has very little reason not to defer in many cases. Unless the value of the resulting achievement is implausibly magnified to make up for the low probability of getting the relevant good, the reasons against deference will be utterly flimsy in those cases. Second, it is an implication of this explanatory strategy that there is no stable anti-deference rationale; rather it all depends on how likely one is to gain understanding on this or that topic in this or that circumstance.[19]

Suppose, for example, that if you take more time to reflect, it is unlikely you will gain understanding about the ethics of torture but that it is quite likely you will gain understanding about the ethics of meat consumption. This means you would have stronger reasons to defer about the ethics of torture than about the ethics of meat consumption. Or suppose, as seems likely, that you are more likely to achieve understanding of whether you should lie to your colleague than of the correct normative ethical theory (the latter issue being far more intellectually fraught and complex). This suggests you should be more prepared to defer about the correct normative ethical theory than about whether to lie to your colleague. More worrisome still, the chance that one gains understanding will often depend on background factors like education and intelligence. Consequently, it would seem that your friends who have no exposure to philosophy or are less bright than you should

---

[16]See Hills (2009), McGrath (2011), and Nickel (2001).

[17]See Callahan (2018), Crisp (2014), Hills (2009), and Howell (2014).

[18]By "high-grade cognitive achievement," I do not mean to assume an overly intellectualized view of understanding. Hills (2015) maintains that the inferential abilities in question may often be tacit; Callahan (2018) emphasizes the role of emotion in understanding.

[19]One might deny that understanding is hard to come by. For certain elementary moral claims that is surely plausible, though if one needs to defer about elementary claims, a much deeper sort of defect of moral agency is at issue (*cf.* Sliwa 2012, 185; Wiland 2014, 171). For nonelementary moral claims, the suggestion that understanding is easy to come by is less plausible. To be sure, on Sliwa's (2017) account, moral understanding is, in a way, easy to come by secondhand, since it can be acquired via testimony. But Sliwa also denies that moral deference is problematic. I do not deny that there are views of understanding that make it easy (or easier) to come by. However, an understanding-centered explanation of the puzzle of moral deference seems to require a relatively ambitious notion of understanding. Thanks to a reviewer from the *Canadian Journal of Philosophy* for pressing me.

tend to defer more about moral questions than you should. All this seems dubious. Shouldn't the anti-deference rationale be a bit more invariant and democratic?

An additional worry is that facts about the expected value of gaining understanding will often be deliberatively opaque. If there are facts, say, about the likelihood of your gaining understanding about the ethics of torture or meat consumption, these facts will often be nontransparent and even quite inaccessible to you. (Perhaps, for example, the fact that you were spanked as a child clouds your moral judgment and makes it unlikely you will ever achieve understanding of the ethics of torture.) Formulating an anti-deference principle in terms of such facts therefore raises the worry that the resulting principle will fail to be action-guiding, since it cannot meaningfully inform your deliberation.

Turn now to the associated moral goods. Appeals to moral worth and virtue do not take us very far. First, there are obvious complications. One can defer precisely out of respect for the moral importance of getting it right and out of concern to minimize one's risk of wrongdoing, and then act dutifully on the basis of deferential belief. This suggests that actions flowing from deference can have moral worth and be virtuous. Second, and more importantly, virtue and moral worth explanations cannot be fundamental. Take Kant's famous example of the shopkeeper. The fact that the shopkeeper's action (giving the boy the proper change) would be virtuous or morally worthy doesn't explain what is right about refusing to defraud the boy; rather, this is explained by the fact that it is the just and fair thing to, that it appropriately respects the boy, and so on, and the action only has moral worth and displays virtue insofar as, and because, it is sensitive to these features. Likewise, considerations of virtue and moral worth are secondary when it comes to explaining the puzzle of moral deference. What we want is an account of the appropriateness or inappropriateness of moral deference. There are *independent* reasons bearing on the question of whether and when deference is appropriate. For example, Enoch (2014) has argued compellingly that there are powerful reasons in favor of deference given by the duty to minimize one's risk of wrongdoing. I will argue below that there are reasons on the other side having to do with the importance of an independent mind for being a participant in the moral enterprise. And there are likely other moral principles and values speaking for or against deference. But whatever the full story is, it seems that *these* reasons will be explanatorily fundamental. Appeals to virtue and moral worth cannot deliver such reasons. They are parasitic on a background normative account that explains when and why deference is problematic. Since it will sometimes be the case that one ought all-things-considered defer, moral deference will sometimes be both virtuous and have moral worth. What we need is an independent account of the appropriateness of moral deference.

The basic idea that the reasons speaking against moral deference have something to do with moral agency seems promising. As I have suggested, this nicely coheres with our practices of moral education, which assume that mature agents ought to exhibit some autonomy or intellectual independence with respect to moral matters. But appeals to autonomy and intellectual independence don't *explain* the puzzle: they merely restate it. Similarly, I have suggested that appeals to virtue and moral worth don't take us very far. They cannot, in any very fundamental way, explain what is problematic about moral deference. Moreover, I have suggested that specifying the cognitive good threatened by deference in terms of understanding runs into trouble, since it delivers an anti-deference rationale which is implausibly sensitive to one's chances of gaining understanding and is grounded in facts which are deliberatively opaque. If that is right, we should look for an anti-deference rationale connected to an ideal of moral agency, which locates values that are more deeply explanatory and avoids appealing to a demanding cognitive good.

3. The practice of morality

A helpful way to fill out the agency ideal is to think about what is involved in the practice of morality. While there is no single value that speaks against moral deference, the idea of what it takes to engage

well in the practice of morality provides a handy unifying lens. Here I want to highlight four values connected to the practice of morality and show how they are at odds with deference.

The Economy of Interpersonal Justification. The ideal of interpersonal justification is central to our understanding of morality. According to this ideal, agents ought to be able to justify their decisions and actions to each other. One important problem with moral deference is that it makes use of second-best reasons in contexts of interpersonal justification. Suppose Tammy defers to Sam about whether eating meat is permissible. If we query her about her attitudes and choices, she will cite Sam and facts about him (e.g., his moral sensitivity and thoughtfulness) as the reason for her commitments. But when we confront Tammy about her moral behavior and commitments, these are not the reasons we are primarily interested in. At least in the first instance, we want to know what she sees as the reasons *explaining the truth* of her commitments. If she responds that she is a vegetarian because of Sam's testimony, that's a bit disappointing. She seems to have settled for a kind of justification which is less than ideal.

Of course, if Tammy is a normally functioning adult, she will have the capacity to appreciate various kinds of moral considerations bearing on the issue. She will thus still be able to cite such reasons when asked for justification. However, if she is clear about the justifying "why" of her commitment, she will have to admit that it is not grounded in her own assessment of the moral reasons bearing on the issue but in her reliance on Sam. To refine things a bit, consider Eric Wiland's (2014) distinction between two kinds of deference: deference about which reasons bear on an issue and deference about how to weigh the reasons that bear on an issue. For convenience, call these A- and B-deference, respectively. Tammy might engage in either kind of deference. She might not feel entirely confident settling on which reasons bear on the issue: it's not that she cannot, in principle, recognize various categories of moral reasons that might be relevant; it's that she's not very clear what their fine-grained relevance is for the present issue. Similarly, Tammy might not feel confident in weighing the reasons bearing on the issue: it's not as if she cannot, in principle, weigh different kinds of reasons; it's just that she's not confident she's got a good grip on how they stack up in this case. Either way, because Tammy has basic moral capacities, she can search for normatively relevant facts when prompted. If she A-defers, she can search for reasons bearing on the issue; if she B-defers, she can make some assessment of the relative weight of the reasons bearing on the issue. Either way, she would cite normative considerations that are, by hypothesis, not the justifying grounds of her commitment. Her reasoning will be, in an important sense, post hoc. My claim is not that Tammy *cannot* engage in interpersonal justification by searching for A-relevant or B-relevant normative considerations. Rather, my claim is that in doing so, Tammy departs from an ideal of interpersonal justification that is part of an ideal of moral agency: Tammy ought to be able to appeal to nontestimonial reasons that are *her* (actual) reasons.

Appropriate moral reasoning, it appears, is shot through with epistemic considerations. As Max Lewis (2020b) points out, unproblematic forms of moral reasoning make use of intuition as well as various forms of indirect inference (inference to the best explanation, reductio, moral consistency reasoning, etc.). Such epistemically bona fide methods of arriving at moral beliefs may nevertheless leave the *morally* justifying basis for the accepted claims opaque. Hence, the contrast I am after is not best characterized in terms of a contrast between appealing to "moral" and "epistemic" reasons. Instead, it should be put in terms of relevance for determination of the moral truth in contexts of interpersonal reasoning and justification. Lewis suggests "passing the buck" through moral testimony can be epistemically superior to thinking for oneself. That much seems undoubtedly true. But "passing the buck" through deference is not ideal in contexts of interpersonal justification—or at any rate, not if interpersonal justification is understood in the suggested way. Since moral deference is sometimes both justified and required, it follows that moral deference can serve as suitable justification in one obvious sense: as a defeater for warranted blame. At issue, however, is the kind of interpersonal justification we engage in when we reason together about our commitments—where the purpose of challenging one another is to light upon the truth.

Such justification does recruit epistemic considerations. For example, many moral claims invoke the idea of something's being a "reasonable" thing to do. How does one know what is reasonable? Presumably, at least part of the answer is that one has some sense of what *others* deem reasonable, and one's own sense of what is reasonable is partly justified because it is sensitive to such facts. If you and I are discussing what is reasonable for someone to do, it may be appropriate to appeal to such considerations. For example, you might say, "C'mon! Nobody thinks you would have to be *that* circumspect in such a situation!" But this seems importantly different from saying, "Sam told me (and he would know)." If Sam is a trusted source for both of us, perhaps that piece of information could contribute to ratcheting up our confidence that some line of mutual reasoning is on track, but it cannot really help us very far in our search for the moral truth. Assuming we trust Sam, what he believes is merely an "indicator" of the truth. As such, it is a placeholder for another kind of justification that is more deeply explanatory. So far as the project of interpersonally justifying our moral commitments goes, it isn't ideal to have to appeal to Sam. We may treat his beliefs as evidence that we are on track, but we shouldn't treat them as if they made further deliberation unnecessary. Far from it: they are an encouragement to keep thinking. It is only once we've arrived at a set of considerations that make sense to both of us as explaining the truth of our commitments that we can rest content. How exactly to characterize such considerations I'm not sure. If, as I have suggested, moral reasoning is epistemically saturated, appealing to "pure" moral reasons is unhelpful here. But regardless of how they are best characterized, the considerations in question are *not* testimonial. Testimonial considerations may boost our confidence along the way, but the ideal of interpersonal justification sets us the goal of moving to a state where we have upgraded to reasons we both recognize as more deeply explanatory.[20]

Correcting Moral Norms. Moral norms are often flawed and in need of correction. There is reason to think collective moral inquiry is best served by having independent centers of deliberation and that such independence is threatened by deference. First, the more independent centers of deliberation there are, the more cognitive horsepower is devoted to solving a problem. The person who defers is not in possession of the right kind of reasons to make a deliberative contribution, having outsourced those to another. This means that so far as group deliberation about moral problems goes, deference is cognitively idle. Second, independent centers of deliberation are more likely to preserve valuable cognitive diversity, and cognitive diversity can be an asset in collective deliberation. Since deference takes on board someone else's deliberative conclusion, it will tend to diminish valuable cognitive diversity. Third, independent centers of moral deliberation are a hedge against undesirable cascade effects. Because deference echoes the conclusions of another, pernicious beliefs can spread like contagion in a highly deferential population. Naturally, if the source belief is true, cascade effects would be more benign. However, unless there is antecedent reason to think the source beliefs would more often be true than false and the pattern of deference in the population would trend positively toward truth as a result, banking on the lucky outcome is inadvisable. It seems safer to prevent against such cascade effects by promoting independent deliberation.[21]

This broadly Millian anti-deference rationale may appear to badly overgeneralize.[22] After all, it seems to support an anti-deference stance that is quite independent of content. In ordinary and scientific matters, too, independent deliberation serves the instrumentally valuable functions mentioned above. Hence, this particular anti-deference rationale cannot contribute to explaining a more selective anti-deference stance for the moral domain.

---

[20]For somewhat different suggestions about the importance of interpersonal justification, see Hills (2009, 106–8) and Nickel (2001, 256); Thanks to a reviewer and the editor from the *Canadian Journal of Philosophy* for pressing me to articulate the issues with a bit more nuance.

[21]Hazlett (2017, 61) expresses sympathy with this kind of rationale against moral deference, though he goes on to defend a different sort of view of the social value of intellectual diversity.

[22]Thanks to a reviewer and the editor of the *Canadian Journal of Philosophy* for pressing this worry.

The objection is basically right that there isn't a categorical distinction here between morality and other domains. However, this doesn't undermine the more domain-specific anti-deference case. What is needed for a selective anti-deference case capable of explaining the contrast between moral content and ordinary factual and scientific content is not anything so strong as the claim that having an independent mind matters greatly in morality but not at all in other matters. Rather, a weaker comparative claim will do: in ordinary and scientific matters, there are no reasons sufficiently powerful and general to support a warranted default to think for oneself, whereas there are such reasons in the moral domain grounded in special values internal to it.

Consider how it makes sense to divide intellectual labor in the moral and scientific domains. In science, there is a producer/consumer distinction reflecting a highly efficient division of intellectual labor. Ordinary people support the advancement of science chiefly as voters, taxpayers, and consumers—that is, indirectly, rather than as direct contributors to the intellectual enterprise. Most people do not have the ability to contribute productively to the advancement of science and there is, in any case, no need for them to do so. In morality, by contrast, a producer/consumer distinction makes little sense. Laypersons *can* be productive members in the social exchange of moral ideas. This does not imply that the moral truth isn't often hard to discern, that moral facts are equally accessible to all, or that there can be no genuine moral experts. What it does imply is that when it comes to morality, there is not the same yawning gulf between lay and expert competence that is characteristic of science. This is plausible. Morality's basic concepts and modes of reasoning are anchored in ordinary thought and experience in ways scientific theory is not. Thanks to the role of morality in the fabric of human life, ordinary adults are capable of grasping subtle distinctions in the moral landscape and deploying a conceptually sophisticated vocabulary that reflects that competence. There is, therefore, not the same stark contrast as there is in science between folk and expert representations of the domain.

The response to the overgeneralization worry, then, is to admit that while Millian considerations do not support anything like an absolute asymmetry, against the backdrop of considerations about sensible divisions of intellectual labor, they nevertheless support a significant asymmetry. This asymmetry is compatible with some amount of moral outsourcing and certain forms of ethical expertise; it is compatible, too, with certain select individuals having special reasons not to defer about scientific content. What it requires is that division of labor considerations favor different default settings for the population as a whole.

Accountability and Participatory Depth. Consider what it means to participate in a rule-governed practice as an "insider." In the standard case, participants are insiders in at least the minimal sense that they take the norms to be regulative of their conduct and to furnish standards against which their own conduct and the conduct of others is appropriately measured. In some cases—games, conventions—what this requires is thin and undemanding. One might be more or less emotionally invested in the practice, but for the most part no deeper intellectual engagement is called for. Morality is plausibly different in virtue of special features: moral reasons are highly important, they apply universally and are binding on all, and they are associated with practices of censure and sanction that cut deep.

Imagine the following Kafkaesque scenario. Suppose we create a norm-governed institution *de novo* which requires participants to do certain silly things, like hopping on one foot while waiting at the bus stop or tapping one's head when in line. The rules are completely arbitrary and pointless, yet they are associated with hefty accountability practices and are binding on all. Perhaps, for example, participants in the practice subject noncompliers to prolonged social ostracism. Presumably, such a system would be morally unjustified, but it would also have the troubling feature that its participants would have no idea why they should follow these onerous norms.

Given the rational structure of morality, participation in moral practice need not be like this. In principle at least, moral norms can be rationally accessible to participants in the enterprise, not only in the sense that they can appreciate *what* the rules of the game are but *why* they are what they are. Because of the gravity and inescapability of the norms and because compliance with the norms can

be highly consequential for participants, it seems best that participants own the norms in a deeper sense, which includes appreciation of the rational principles behind the norms they accept. In short, moral norms are special, and the features that make them special also make it natural to accept a kind of meta-norm or ideal for the moral domain, namely that those agents to whom the norms apply ought not only to accept and internalize the relevant norms in such a way that they can be held accountable for following them but appreciate the rational grounds for these norms, to the extent this is possible for them. If participants fail to grasp the justifying grounds for the norms they embrace, they are only shallow or halfway-insiders to those norms. They may govern their conduct according to the norms and even be held accountable for following them, but their mode of norm compliance is suboptimal. They fail to be insiders in the deep and full sense which it is desirable that those subject to moral norms should be.

A straightforward anti-deference rationale follows. Deference about moral content involves acceptance of norms to which deferring agents are themselves subject. Because deference bypasses appreciation of the grounding or explanatory reasons for the norms accepted, deferential uptake of moral content will tend to be bad from the point of view of participatory depth. This speaks against moral deference in perpetuity. If the deferring agent is rationally competent, then abiding deference suggests she is falling short of a reasonable expectation that she should seek to appreciate reasons for the normative content she accepts.

Moral Reasoning Competence. In specifying the cognitive good that deference jeopardizes, we need not appeal to anything as ambitious as understanding. A lower-grade cognitive good will do. Consider the corrosive effects of deference on moral reasoning competence. Moral reasoning, like reasoning in general, involves making inferences. To make inferences, moral reasoning uses theory-like representational structures which encode normative content—if you like, maps or models or recipes for the moral domain. Like other forms of inference, moral inference is frequently unconscious. Moreover, people's grasp of the principles which feature in their moral reasoning is often inchoate. Still, whether conscious or unconscious, people's reasoning about moral issues makes use of more or less articulate and systematic background commitments about norms and reasons. These representations appear to be similar in many ways to the folk theories people use to reason in other domains.[23]

Deference plausibly affects moral reasoning competence by affecting the quality of the background representational structure: if one has filled in portions of one's moral map through deference, one will tend to be able to explain less, and explain less well, than if one's moral map is filled in by explanatorily relevant information. Deference involves adding to one's stock of beliefs on the cheap, without the benefit of grasping explanatory relations. It therefore amounts to outsourcing explanatorily relevant information. To be sure, adding to one's stock of beliefs and commitments by way of deference can extend the range of one's background theory. But this gain in scope is shallow, hampering informative generalization to novel cases.[24] Interestingly, the activity of seeking explanations has been shown to facilitate learning. When people are prompted to search for explanations, they are more likely to discover rules, patterns, and principles that support informative generalizations (Lombrozo 2006). So while deference may play some role in moral learning, that role should not be exaggerated. Deference can involve trading a less valuable for a more valuable learning opportunity.

If deference meant the agent never went on to think further about the question, the cognitive benefit of adding content to one's background theory via deference would have to be weighed against the opportunity cost of forgone future reflection. As I have stressed, however, it doesn't have to be like that. When the moral student copies another's answer, she is not off the hook for thinking further about the matter. So far as moral reasoning competence goes, temporarily plugging gaps in

---

[23]Gottlieb and Lombrozo (2018), Rhodes and Wellman (2017).
[24]See Hills (2009) and Howell (2014) on the importance of being able to generalize to new cases.

one's moral map is not all that worrisome. Abiding deference, however, is far more worrisome; it will tend to be a less cognitively valuable strategy than reflection.

In sum, moral maps mediate moral reasoning competence. People who defer too much will tend to have shoddy moral maps. As long as reflection is a feasible alternative, one will tend, all else equal, to promote one's competence as a moral reasoner better over the long haul by engaging in reflection than by deferring.

4. Explaining the puzzle

Let's recap. A value-based explanation of the problem of moral deference needs to locate credible *pro tanto* reasons operative in the moral domain. These reasons should speak against moral deference quite generally, thereby explaining why there is something prima facie fishy about moral deference and why moral deference cannot be casually and indefinitely sustained. They should also be consistent with and help illuminate why it is generally worse to sustain moral beliefs deferentially rather than merely acquiring them deferentially. Moreover, the resulting anti-deference rationale should not be too variable or limited, applying only to philosophy professors or to easy moral topics. Instead, it should be quite sweeping, applying to most (competent) agents, most of the time, and covering most, if not all, moral questions.

It seems plausible to look for such reasons in an ideal of moral agency. Yet common ways of spelling out that idea do not work. I have appealed to several alternative values that are closely connected to moral practice: interpersonal justification, moral reform, participatory depth, and moral reasoning competence. The last of these is a general asset in moral matters, so it is an independent contributing good, but it also plays a role in facilitating the other goods while satisfying the desideratum that it not be overly demanding.

Together, these values speak against deference in roughly the way needed to yield an anti-deference rationale with plausible weight and scope. The reasons supported by these values are substantial enough to justify a default anti-deference stance for the moral domain; at the same time, they are not so hefty that they cannot be readily outweighed when circumstances demand. Moreover, the values are quite general and democratic: they are associated with morality as a general practice, not with what agents can achieve on this or that occasion, and they are accessible without high-grade cognitive achievements. Finally, the values are not seriously threatened if deference is one-off and occasional—and certainly not if agents go on to do their moral homework. So while the reasons grounded in these values explain why there is something prima facie troubling about moral deference in general (asymmetry 1), they also explain why it is generally more troubling to sustain moral beliefs deferentially rather than merely forming them deferentially (asymmetry 2).

The fact that the reasons against deference are not too strong and that one-off deference is not all that bad is important. This ensures that the anti-deference case doesn't license moral recklessness. The default norm against deference grounded in the values I have cited does not imply that adults should never engage in moral deference. In Karen Jones's (1999, 59–60) example of the young man who must decide whether to accept testimony from women colleagues about whether some behavior counts as sexist, it seems plausible enough that the young man ought to defer. Or consider David Enoch's (2014) example, which concerns his own moral judgments about the military conflicts of his nation, Israel. Enoch asks us to imagine that when violence erupts in the Middle East, he finds himself too quickly and easily swept along with other Israeli citizens in nationalistic fervor only to find later, when ardor gives way to sobriety, that he regrets his initial reaction. Enoch has noticed that his colleague Alon is far quicker to embrace the view it takes Enoch weeks to adopt, far more likely to maintain the sober and critical attitude in the heated initial phases of conflict and, hence, far more likely to judge the morality of the conflict accurately. Given this track record, when it comes time to form an opinion about a new conflict in which his nation is embroiled, Enoch claims he has most reason

to form his opinion on the basis of his colleague's testimony, especially if he must act on the basis of the opinion, e.g., by casting a ballot.

These examples are important, but their significance should not be overstated. First, as emphasized, the fact that agents ought sometimes to defer is consistent with residual unease about moral deference, an unease which I have argued can be justified. When we think people have most reason to defer, as we might in Jones's and Enoch's examples, this is because there are urgent moral values at stake which justify deference. Our reaction to such cases does not mean our intuitive aversion to deference cannot be underwritten by credible values and must therefore be rejected. Second, although the reasons to avoid harm and wrongdoing are significant, we should not overestimate the relevance of such considerations for the issue of deference. The cases in which it is most compelling to suppose that deference is mandated are those in which refusing to defer would be morally reckless. If you must decide whether to pull the plug on someone in a coma, then if you have reason to think someone else is better positioned to make the right decision, you should defer. If you must decide whether to torture a terror suspect who is credibly believed to have information which could save many lives, and you reasonably believe someone else is in a better position to reach a justified decision, you should defer. In cases like these, the stakes are high and the risks of nondeference are clear and grave. Such cases are genuine and important but, for most people, presumably fairly exceptional. The further we move away from high-stakes urgency, the less plausible it becomes that agents must defer.

If we think *any* risk of harm or wrongdoing, no matter how small, mandates deference, then the pro-deference case massively overgeneralizes. This is a problem for Enoch's view. Since he identifies a powerful class of reasons speaking in favor of moral deference but cannot find significant moral reasons speaking against moral deference, it appears agents should defer just whenever they reasonably believe someone else is more likely to get it right. If the values identified above are plausible, however, then there *is* at least some counterweight to the kinds of considerations cited by Enoch. Moreover, while risk of harm or wrongdoing is serious, such considerations must reach a threshold of gravity before they begin to swamp the deliberative field. You are not required to avoid all moral risk any more than you are required in general to avoid all risk. We don't think agents must try to live the morally safest lives. What you are required to do is to take reasonable measures to minimize your risk of harming or wronging others. Given that the practice of morality works best when agents are deliberatively engaged and not deferring too much, it seems plausible to think that it is acceptable for agents to exercise their moral agency capacities within reasonable limits.

Consider an analogy. If you are a parent, you have reasons to exercise your parenting capacities to the best of your ability, even when you are an imperfect parent. The fact that someone else could do a better job raising your children is not a decisive reason to let them do so. You have reason to try to be a good parent, to aspire to live out the ideals of parenthood, even when you know you will fall short, when this imposes some cost on your child, and when a less risky alternative is available. Morality, I think, is a lot like this. Being a moral agent is *morally* risky—and morality makes some room for this. Many commendable exercises of moral agency run some moral risk. That we encourage children in the independent exercise of their moral agency even when, as adults, we have far greater moral insight, suggests that we think it is an acceptable risk of moral agency that one sometimes gets it wrong.[25] The crucial question is what level of risk is acceptable. The situations in which deference is clearly mandated are those in which agents must act, the stakes are substantial, and agents reasonably believe someone else is better positioned to render a justified verdict. Clear and grave risks aside, however, it seems plausible that thinking for oneself in moral matters falls well within the acceptable range most of the time.

---

[25]*Cf.* Nickel (2001, 266).

Conclusion

The problem of deference is broader than sometimes assumed. The types of content for which deference is prima facie suspect include morality, aesthetics, humor, religion, and broader questions of value including meaning and the personal good. It seems potentially problematic to defer not only about the morality of eating meat, but also about the aesthetic merits of Rococo, about whether a joke one has heard is funny, whether there is a god, what the meaning of life is, and sundry practical questions, like what one should do with one's free afternoon, whether one should undergo chemo therapy, whether and whom to marry, and so on.

   Here is a hypothesis. The problem of deference in morality is a special case of a more general phenomenon, and the sort of explanation that works in the moral case works, *mutatis mutandis*, more broadly. What seems problematic across the above cases is that the agent has a defective or subpar relation to certain kinds of content. A plausible conjecture is that our discomfort about forming deferential attitudes of various kinds is informed by ideals of agency. All else equal, one shouldn't defer about the existence of god or the meaning of life or about whether to marry. This is not because one couldn't conceivably improve the accuracy of one's beliefs by taking the word of others, but because making up one's own mind about such questions has special importance. Perhaps I could better judge whether there is a god or whether you should marry your lover. Still, barring very special circumstances, you should settle such questions for yourself.

   Our reluctance about deference thus appears to be inspired by deep agency ideals. What exactly those ideals are, and whether they are good ideals to have, is a further question. The strategy of value-based explanation pursued in this paper suggests what it would take to vindicate such ideals. Ultimately, those ideals need to be underwritten by credible values which supply standing reasons against deference. To decide whether our intuitive reluctance can be rationally supported for other kinds of cases, then, we need to examine the values at stake in our ideals of practical agency.

Jonathan Knutzen is a Lise Meitner Postdoctoral Fellow at the University of Graz. He works primarily in ethics and moral psychology.

References

Callahan, Laura. 2018. "Moral Testimony: A Re-conceived Understanding Explanation." *The Philosophical Quarterly* 68 (272): 437–59.

Cholbi, Michael. 2007. "Moral Expertise and the Credentials Problem." *Ethical Theory and Moral Practice* 10 (4): 323–34.

Crisp, Roger. 2014. "Moral Testimony Pessimism: A Defence." *Aristotelian Society Supplementary Volume* 88 (1): 129–43.

Davia, Cory, and Michele Palmira. 2015. "Moral Deference and Deference to an Epistemic Peer." *The Philosophical Quarterly* 65 (261): 605–25.

Elgin, Catherine. 2009. "Is Understanding Factive?" In *Epistemic Value*, edited by Adrian Haddock, Alan Millar, and Duncan Pritchard. Oxford: Oxford University Press.

Enoch, David. 2014. "A Defense of Moral Deference." *The Journal of Philosophy* 111 (5): 229–58.

Fletcher, Guy. 2016. "Moral Testimony: Once More with Feeling." In *Oxford Studies in Metaethics* Vol. 11, edited by Russ Shafer-Landau, 45–73. New York: Oxford University Press.

Frankena, William. 1973. "Toward a Philosophy of Moral Education." In *Moral Education*, edited by Barry I. Chazan and Jonas F. Soltis. New York: Teachers College Press.

Fricker, Elizabeth. 2006. "Testimony and Epistemic Autonomy." In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 225–50. Oxford: Oxford University Press.

Gottlieb, Sara, and Tania Lombrozo. 2018. "Folk Theories in the Moral Domain." In *Atlas of Moral Psychology*, edited by Kurt Gray and Jesse Graham. New York: Guilford Press.

Hazlett, Allan. 2017. "Towards Social Accounts of Testimonial Asymmetries." *Noûs* 51 (1): 49–73.

Hills, Alison. 2009. "Moral Testimony and Moral Epistemology." *Ethics* 120 (1): 94–127.

Hills, Alison. 2015. "Understanding Why." *Noûs* 50 (4): 661–88.

Hopkins, Robert. 2007. "What Is Wrong with Moral Testimony?" *Philosophy and Phenomenological Research* 74 (3): 611–34.

Howell, Robert. 2014. "Google Morals, Virtue, and the Asymmetry of Deference." *Noûs* 48 (3): 389–415.

Jones, Karen. 1999. "Second-Hand Moral Knowledge." *The Journal of Philosophy* 96 (2): 55–78.

Lewis, Max. 2020a. "Moral Deference, Moral Assertion, and Pragmatics." *Ethical Theory and Moral Practice* 23: 5–22.

Lewis, Max. 2020b. "The New Puzzle of Moral Deference." *Canadian Journal of Philosophy* 50 (4): 460–76.

Lombrozo, Tania. 2006. "The Structure and Function of Explanations." *Trends in Cognitive Sciences* 10 (10): 464–70.

Lord, Errol. 2018. "How to Learn about Aesthetics and Morality through Acquaintance and Deference." In *Oxford Studies in Metaethics* Vol. 13, edited by Russ Shafer-Landau. New York: Oxford University Press.

McGrath, Sarah. 2009. "The Puzzle of Pure Moral Deference." *Philosophical Perspectives* 23 (1): 321–44.

McGrath, Sarah. 2011. "Skepticism About Moral Expertise as a Puzzle for Moral Realism." *The Journal of Philosophy* 108 (3): 111–37.

McShane, Paddy. 2018. "The Non-remedial Value of Dependence on Moral Testimony." *Philosophical Studies* 175 (3): 629–47.

Mogensen, Andreas. 2017. "Moral Testimony Pessimism and the Uncertain Value of Authenticity." *Philosophy and Phenomenological Research* 95 (2): 261–84.

Nickel, Philip. 2001. "Moral Testimony and Its Authority." *Ethical Theory and Moral Practice* 4 (3): 253–66.

Rhodes, Marjorie, and Henry Wellman. 2017. "Moral Learning as Intuitive Theory Revision." *Cognition* 167: 191–200.

Sliwa, Paulina. 2012. "In Defense of Moral Testimony." *Philosophical Studies* 158 (2): 175–95.

Sliwa, Paulina. 2017. "Moral Understanding as Knowing Right from Wrong." *Ethics* 127 (3): 521–52.

Wiland, Eric. 2014. "Rossian Deontology and the Possibility of Moral Expertise." In *Oxford Studies in Normative Ethics* Vol. 4, edited by Russ Shafer-Landau. New York: Oxford University Press.

Wiland, Eric. 2017. "Moral Testimony: Going on the Offensive." In *Oxford Studies in Metaethics* Vol. 12, edited by Russ Shafer-Landau. New York: Oxford University Press.

Williams, Bernard. 1995. *Making Sense of Humanity*. Cambridge: Cambridge University Press.