

Deference to moral testimony and (in)authenticity

Shannon Brick

Abstract

The philosophical debate over moral testimony is a debate over whether it is appropriate to trust moral testimony. Most people who have engaged in the debate agree that it seems as though people ought to form moral beliefs by thinking through the issues on their own, rather than by trusting another person's testimony. This chapter reports the results of two studies designed to investigate one hypothesis about why things seem this way. According to this hypothesis, we think it is inappropriate to trust moral testimony because we are committed to the ideal of personal authenticity, and we think that trusting moral testimony would put one in a condition of inauthenticity.

1. Introduction¹

Some philosophers think it is bad to form a moral belief by simply trusting what another person tells you (Nickel 2001; Hopkins 2007; McGrath 2009; Hills 2009; Howell 2014; Fletcher 2016; Lewis 2020). Accordingly, these philosophers think there is a difference between the way we should respond when someone tells us that, say, it is raining outside, and when someone tell us that, say, eating meat is wrong. After all, it seems just as good to believe that it is raining outside on the basis of testimony, as it is to believe it on the basis of your walking to the window to investigate the matter yourself. Other philosophers think there is nothing wrong with deferring to moral testimony (Sliwa 2012; Groll and Decker 2014; Mogensen 2017; Lord 2018). People in this camp often claim that it is merely suboptimal, but not necessarily bad, to form a moral belief by trusting moral testimony (Lord 2018). Sometimes, they reject the claim that moral deference is even suboptimal (cf. Sliwa 2012).

Most parties to the debate agree that there *seems* to be a difference between the way we do (and should) respond to moral testimony and the way we do (and should) respond to testimony about,

¹ For extremely helpful conversations about this project I am grateful to Josh Knobe, Rebecca Keller and Tomasz Zyglewicz.

say, the weather (Nickel 2001; Hopkins 2007; McGrath 2009; Hills 2009; Groll and Decker 2014; Howell 2014; Fletcher 2016; Mogensen 2017; Lewis 2020.) Accordingly, most people agree that it seems better to form one's moral beliefs by thinking through the issues on one's own. This fact is supposed to stand in need of explanation. Arguments about moral testimony thus usually proceed from a claim about how trusting moral testimony seems, to the issue of whether there is, in fact, a normative difference between trusting moral testimony and trusting non-moral testimony.²

Some labels will help clarify the argumentative terrain. *Pessimists* think there is a normative difference between trusting moral testimony and trusting non-moral testimony. *Optimists* reject Pessimism.³ Optimists either accept the appearance of a psychological asymmetry between moral and non-moral testimony but attempt to explain it away, or simply accept that there is an asymmetry but deny that it entails that it is wrong to trust moral testimony.

What explains the fact that most Optimists and Pessimists agree that there seems to be a difference between trusting moral testimony and trusting non-moral testimony? Here, examples do quite a bit of work. Allison Hills (2009: 94) opens her argument for Pessimism with the following example:

Eleanor has always enjoyed eating meat but has recently realized that it raises some moral issues. Rather than thinking further about these, however, she talks to a friend, who tells her that eating meat is wrong. Eleanor knows that her friend is normally trustworthy and reliable, so she believes her and accepts that eating meat is wrong.

Fellow pessimist, Guy Fletcher (2016: 45), begins with the following case:

You are on holiday and someone gives you tickets to a boxing match. En route you bump into a friend who asks your plans for the evening. You tell her that you are headed to the stadium, though you are not sure where it is. She replies: "It's on 21st street." Taking her to be reliable, you defer to her and head for 21st street. Before

² Strictly speaking, the asymmetry is not between moral testimony and non-moral testimony, as some kinds of non-moral testimony (like testimony about aesthetic properties) generate the same worries as does moral testimony. Here, when I speak of an apparent asymmetry between moral and non-moral testimony, it is descriptive non-moral testimony (and not, say, aesthetic testimony) that I have in mind.

³ The "Optimist" and "Pessimist" labels come from Hopkins (2007).

parting she asks what you're going to see at the stadium. You tell her that you are going to watch boxing. She replies: "It's morally wrong to watch boxing." Taking her to be reliable, you defer to her and head back to your hotel.

And here's how pessimist Robert Howell (2014: 389) begins his argument:

Suppose those wizards at Google come out with a new app: Google Morals. No longer will we find ourselves lost in the moral metropolis. When faced with a moral quandry or deep ethical question we can type a query and the answer comes forthwith. Next time I am weighing the value of a tasty steak against the disvalue of animal suffering, I'll know what to do...There seems to be something wrong with using Google Morals. But what is it?

I share the sense that something seems "problematic" (Hopkins 2007; McGrath 2009; Howell 2014; Enoch 2014) about each of these cases. Here, I present the results of two studies designed to help clarify this intuition. The studies were designed to test one hypothesis, in particular; namely, that it is our attachment to the ideal of personal authenticity that explains why it seems problematic to trust moral testimony, but not problematic to trust non-moral testimony. Call this the *Authenticity Hypothesis*.

The Authenticity Hypothesis has been defended by Andreas Mogensen (2017), who makes his case for it by drawing on "our intuitive sense of what it means to be authentic, cashed out via familiar platitudes and metaphors whose imprecision [he] leave[s] unresolved" (Mogensen 2017: 276). These are metaphors and platitudes like "be true to yourself" or "you do you" and, as Mogensen notes, they permeate popular Western culture. The view the present studies are designed to test, then, is that the apparent asymmetry in our attitudes towards moral and non-moral testimony is explained by our sense that deference to moral testimony is in tension with personal authenticity.

In investigating the intuitive foundation to the debate over moral testimony, the studies presented here continue a project initiated by James Andow (2019). In a recent study, Andow investigated the extent to which our intuitive responses to moral testimony are a matter of our taking there to be an epistemic difference between beliefs gained via testimony in the moral and descriptive domains. Andow found that moral beliefs formed on the basis of moral testimony are judged to have a worse

epistemic standing than descriptive beliefs based on testimony. More specifically, he found that beliefs about morality are less likely to be considered knowledge when based on testimony than when they are based on first-hand experience. (The same was not true with respect to descriptive beliefs.) Andow also found some sign that higher perceived background levels of disagreement and deception in the moral domain are responsible for the perceived lower ‘Legitimacy’ of forming moral beliefs by trusting testimony (compared to the Legitimacy of forming descriptive beliefs on that basis). One hypothesis, then, is that lower knowledge ratings are due to the lower perceived Legitimacy of trusting moral testimony, which is in turn due to higher perceived levels of disagreement and deception when it comes to moral matters.

For now, I’ll bracket the question of moral testimony’s perceived epistemic status (although I will return to it in the final section). My focus is on whether the Authenticity Hypothesis captures folk intuitions about moral testimony and whether, if it does, that fact provides support for Pessimism. Mogensen himself thinks that if his explanation of our intuitions about moral testimony is right, then it helps vindicate Optimism about moral testimony. This is because he thinks that our attachment to the ideal of authenticity cannot withstand serious philosophical scrutiny (see Mogensen 2021). As I will suggest in the final section, however, the connection between the Authenticity Hypothesis and the debate between Optimists and Pessimists remains unclear.

2. Testing the Authenticity Hypothesis

2.1 Study 1:

The first study attempts to get empirical traction on the claim that deferring to moral testimony seems problematic because it is seen to be in tension with personal authenticity.

The predictions to be tested are as follows:

P1: People judge agents that defer to moral testimony more harshly than those who arrive at their moral beliefs by thinking for themselves.

P2: People tend to think that agents who defer to moral testimony are not being true to themselves.

P3: People judge agents that defer to moral testimony more harshly because deferring agents strike them as not being true to themselves.

The preregistration form can be found on the As Predicted platform: [link anonymized].

Participants. 452 were recruited from Prolific. From there, they were directed to the survey, which was administered by Qualtrics. Each participant received \$.40 for completing the study. Participants were native or fluent English speakers residing in the U.S.

Method and materials. Participants were randomly assigned to one of two conditions and asked to read a vignette and answer four questions about it (the three listed below, as well as a comprehension question). Each vignette described a person realizing that an activity they enjoy raises moral issues, and then asking a friend about it. The details of the vignette closely mirror a well-known example of moral testimony from the literature (see Hills 2009).⁴ The person's response to the testimony was manipulated across conditions, although in both cases the person ultimately ends up coming to the same conclusion.

Condition 1: Eleanor has always enjoyed eating meat but has recently realized that it raises some moral issues. She talks to a friend, who tells her that eating meat is wrong. Eleanor knows that her friend is normally trustworthy and reliable but decides to think through the matter herself. In the end, she concludes that eating meat is wrong, and so becomes a vegetarian.

Condition 2: Eleanor has always enjoyed eating meat but has recently realized that it raises some moral issues. Rather than thinking further about these, however, Eleanor asks a friend, who tells her that eating meat is wrong. Eleanor knows that her friend is normally trustworthy and reliable, so she believes her and becomes a vegetarian.

All participants were asked to respond to the two following questions, presented on separate screens (participants could not go back):

⁴ In Hills' original sketch, Eleanor "talks to a friend," rather than "asks a friend." However, later (see esp. p.123) Hills makes explicit that discussing a moral question with a friend isn't problematic, but that deference is. The vignette used here thus better reflects Hills' own views, and the Pessimist position more generally, insofar as she thinks Eleanor does something problematic. Hills also doesn't conclude the vignette by saying Eleanor becomes a vegetarian. However, her discussion of what's wrong with deference focuses on the status of actions that proceed on the basis of deferentially acquired moral knowledge (as does most of the pessimistic literature (e.g., Hopkins, (20017); Howell (2014)).

- On the basis of what you have just read, how do you feel about Eleanor? ('Feelings')
- Do you agree with the following statement: "Eleanor is being true to herself"? ('Authenticity')

Responses to the 'Feelings' question formed a five-point scale, ranging from "Very negatively" (1), to "Very Positively" (5). Responses to the 'Authenticity' question also formed a five-point scale, ranging from "Completely disagree" (1) to "Completely agree" (5).

Previous studies (e.g., Newman, Freitas, and Knobe 2015) indicate that judgements of authenticity are correlated with a subject's own moral views. In order to be able to control for the impact of participants' own views about eating meat on their responses to the Authenticity question, all participants answered the following question:

- Do you agree that eating meat is wrong? ('Own')

Answers to this question also formed a five-point scale, ranging from "Completely Disagree" (1), to "Completely Agree" (5).

Results. Descriptive results for Authenticity and Feelings questions, by condition, in Figure 1.

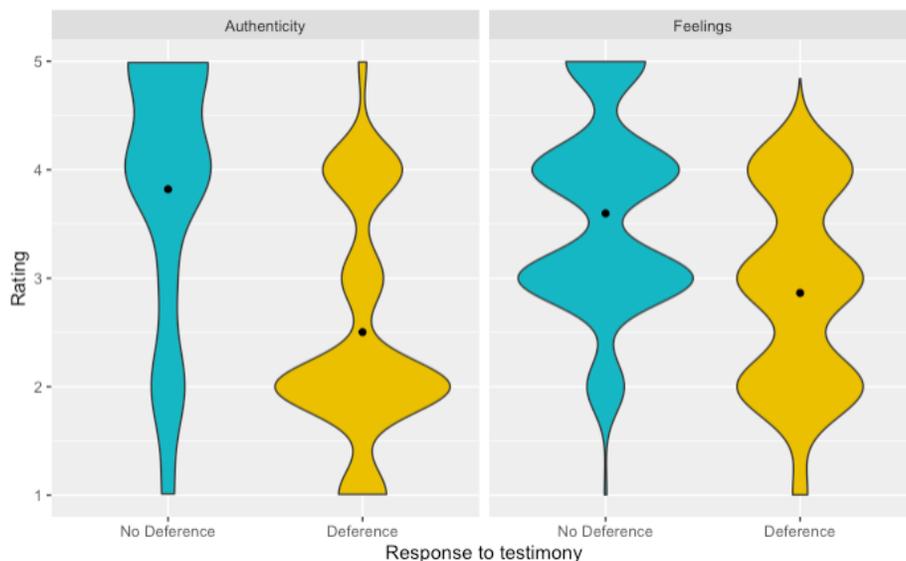


Figure 1: Authenticity ratings and Feelings ratings across the two conditions, with means represented by the black point. The number 3 on the y-axis represents the mid-point, corresponding to the “neither agree nor disagree” (for Authenticity) and “neutral” (for Feelings) answers.

Participants in the deference condition gave lower ratings for Feelings ($M=2.86$ $SD=0.87$) than participants in the no deference condition ($M=3.60$ $SD=0.90$), $t(363.41) = 7.92, p < .001$; Cohen's $d = 0.83$. There was also a large difference in Authenticity ratings between the two conditions, with participants in the deference condition giving lower ratings for Authenticity ($M= 2.50, SD=1.06$) than participants in the no deference condition ($M=3.82, SD= 1.21$), $t(362.16) = 11.11, p < .001$; Cohen's $d = 1.17$.

The results of the present study confirm the finding that judgements of authenticity are correlated with a subject's own moral views. There was a moderate correlation between participants' responses to Own, and their responses to Authenticity ($r=0.29$).

To examine the effect of Authenticity ratings on Feelings ratings, while controlling for the effect of Own, a mediation analysis was conducted using the mediation package in R. ‘Feelings’ was entered as the dependent variable, ‘Authenticity’ was entered as the mediator, ‘Deference’ was entered as the independent variable, and ‘Own’ was entered as the covariant.

As Figure 2 illustrates, the regression coefficient between Deference and Feelings, and the regression coefficient between Authenticity and Feelings was significant. The analysis indicated a significant indirect effect, $-.53$, 95% CI $[-.67, -.40]$. Thus, the effect of Deference on Feelings was partially mediated via ratings of Authenticity.



Figure 2. The relationship between the independent variable (Deference), mediator (Authenticity) and dependent variable (Feelings).

Discussion. All three predictions were borne out. First, people judge agents that defer to moral testimony significantly more harshly than those who arrive at their moral beliefs by thinking for themselves (P1). Second, people are significantly more likely to think that agents who defer to moral testimony are not being true to themselves (P2). Third, the results support the hypothesis that people judge agents that defer to moral testimony more harshly because deferring agents strike them as being inauthentic (P3).

2.2 Study 2

The second study builds on the first one, by examining attitudes towards non-moral testimony, in addition to moral testimony. Accordingly, it seeks to get empirical traction on the Authenticity Hypothesis, which is a hypothesis about the asymmetry between attitudes towards moral and non-moral testimony. More specifically, the study sets out to investigate whether the Authenticity Hypothesis holds because, where deference to moral testimony is seen to be in tension with personal authenticity (where authenticity is a positive value), deference to non-moral testimony is not.

In addition to replicating the results of Study 1, the predictions of the present study are:

P1: People do not judge agents who defer to non-moral testimony more harshly than agents who form their non-moral beliefs by thinking for themselves.

P2: People do not think that agents who defer to non-moral testimony are not being true to themselves.

The preregistration form can be found on the As Predicted platform: [link anonymized].

Participants. 903 were recruited from Prolific. From there, they were directed to the survey, which was administered by Qualtrics. Each participant received \$.25 for completing the study. Participants

were native English speakers residing in the U.S. Of the 903 participants, 893 passed the comprehension question.

Method and materials. Participants were randomly assigned to a condition and asked to carefully read a vignette and answer four questions about it (the three listed below, as well as a comprehension question). Conditions 1 and 2 were the same as the conditions in Study 1. The other conditions differed from the first pair in terms of their content: they described an agent realizing that they do not know they answer to a *non-moral* question, and then talking to a friend about it. As in Study 1, the person's response to the testimony was manipulated across this second pair conditions, although in both cases the person ends up forming the same belief.

Condition 3: Eleanor has always enjoyed eating meat but has recently realized that she doesn't know how to cook it well. She talks to a friend, who tells her that meat cooks best when placed on a searing pan. Eleanor knows that her friend is normally trustworthy and reliable, so forms the belief that meat is best cooked on a searing pan, and starts to cook it like this herself.

Condition 4: Eleanor has always enjoyed eating meat but has recently realized that she doesn't know how to cook it well. She talks to a friend, who tells her that meat cooks best when placed on a searing pan. Eleanor knows that her friend is normally trustworthy and reliable, but decides to look into the matter herself. In the end, she concludes that meat is best cooked on a searing pan, and starts to cook it like this herself.

All participants were asked to respond to the same two questions asked in Study 1 ('Feelings', and 'Authenticity').

Responses to the 'Feelings' question formed a seven-point scale, ranging from "Very negatively" (1), to "Very Positively" (7). Responses to the 'Authenticity' question also formed a seven-point scale, ranging from "Strongly disagree" (1) to "Strongly agree" (7).

To control for the impact of participants' own views about eating and cooking meat on their responses to the Authenticity questions, all participants answered one of the following questions (depending on the condition they were assigned):

- Do you agree that eating meat is wrong. ('Own Moral')
- Do you agree that meat cooks best when placed on a searing pan? ('Own Cooking')

Answers to these questions also formed a seven-point scale, ranging from “Strongly Disagree” (1), to “Strongly Agree” (7).

Results: Feelings about deference were analyzed using a 2 (moral content: yes or no) X 2 (deference: yes or no) ANOVA. The analysis indicated small main effects of content ($F(1, 889)=39.69, p < .001, \eta^2 = 0.04$) and response-to-testimony ($F(1, 889)=20.71, p < .001, \eta^2 = 0.02$), which were qualified by a small moral content X deference interaction, ($F(1, 889)=26.99, p < .001, \eta^2 = 0.02$).

Descriptive results for Authenticity and Feelings questions, by condition, are shown in Figure 3.

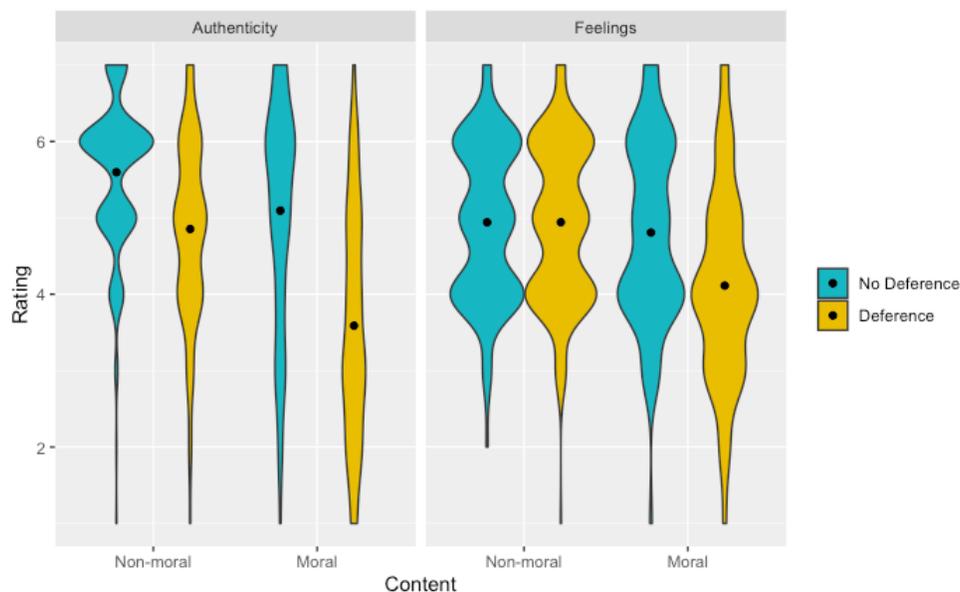


Figure 3: Authenticity ratings and Feelings ratings across the four conditions, with means represented by the black point. The number 4 on the y-axis represents the mid-point, corresponding to the “neither agree nor disagree” (for Authenticity) and “neutral” (for Feelings) answers.

Somewhat surprisingly, post-doc comparisons between the non-moral pair of conditions showed that participants gave lower Authenticity ratings for the agent that defers to non-moral testimony ($M=4.85, SD=1.19$) than they did for the agent that receives non-moral testimony but doesn't defer to it ($M=5.60, SD=0.98$), $t(431.96)=7.22, p < .001$, Cohen's $d = 0.69$. Although the size of this difference was moderate, both means were above the mid-point, meaning that on average participants somewhat agreed that deference to non-moral testimony was consistent with being true to oneself. No difference was found in Feelings ratings of deference ($M=4.94, SD=1.04$) and no deference ($M=4.94, SD=1.02$) to non-moral testimony, $t(445.99)=-5.33e-03, p=0.996$, Cohen's $d=-5.05e-04$. A small correlation between participant's responses to Own Cooking and their responses to Authenticity was also found ($r=0.16$).

Comparisons between the moral pair of conditions confirmed the results of Study 1. That is, participants gave lower Feelings ratings for the condition in which the agent defers to moral testimony ($M=4.11, SD=1.25$) than they gave for the condition in which agents receive moral testimony but don't defer to it ($M=4.81, SD=1.21$), $t(442.06)=5.94, p < .001$; Cohen's $d=0.53$. Participants also gave lower Authenticity ratings for the condition in which the agent defers to moral testimony ($M=3.59, SD=1.50$) than they gave for the condition in which agents who receive moral testimony don't defer to it ($M=5.09, SD=1.47$), $t(442.63)=10.68, p < .001$, Cohen's $d = 1.02$.

As in Study 1, the effect of Authenticity on Feelings across the pair of moral conditions was examined using a mediation analysis. There was a moderate correlation between participants' responses to Own Moral and their responses to Authenticity ($r=0.38$), and so the analysis controlled for the effect of Own Moral. Again, the mediation package in R was used. 'Feelings' was entered as the dependent variable, 'Authenticity' was entered as the mediator, 'Deference' was entered as the independent variable, and 'Own Moral' was entered as the covariant. As Figure 4 shows, the regression coefficient between Deference and Feelings, and the regression coefficient between Authenticity and Feelings, were again significant. The analysis indicated a significant indirect effect, $-.48, 95\% CI [-.61, -.36]$. Thus, the present study confirmed the finding that the effect of Deference on Feelings is partially mediated via ratings of Authenticity.

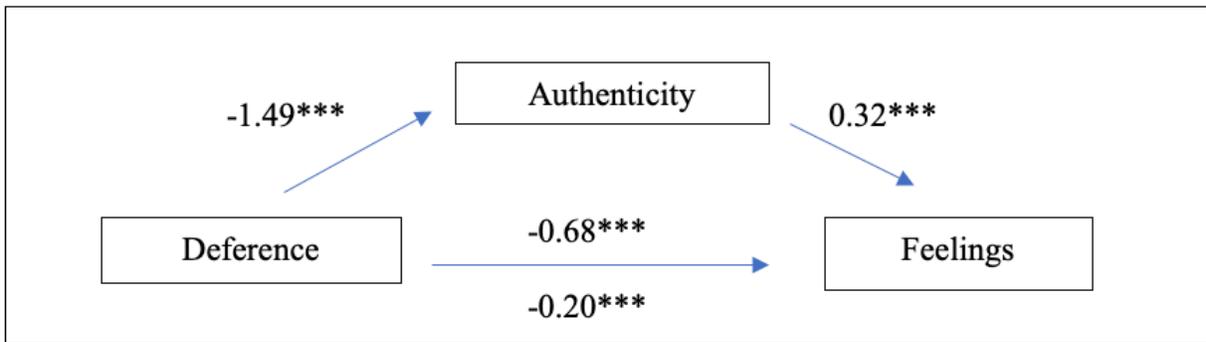


Figure 4: The relationship between the independent variable (Deference), mediator (Authenticity) and dependent variable (Feelings), indicated by results to the moral pair of conditions (Conditions 1 and 2).

A large difference was found in Authenticity ratings of moral and non-moral deference. Participants gave higher Authenticity ratings to non-moral deference ($M = 4.85, SD=1.19$) than to moral deference ($M = 3.59, SD=1.50$), $t(418.83)=9.87, p < .001$; Cohen's $d = .96$. Comparisons between moral and non-moral deference showed a difference in Feelings ratings, with participants giving higher Feelings ratings to non-moral deference ($M = 4.94, SD=1.04$) than to moral deference ($M = 4.11, SD=1.25$), $t(425.76) = 7.60, p < .001$; Cohen's $d=0.74$.

The effect of Authenticity on Feelings across this pair of conditions was examined using a mediation analysis. 'Feelings' was entered as the dependent variable, 'Authenticity' was entered as the mediator, and 'Moral' was entered as the independent variable. Own Cooking and Own Moral were converted into one variable, 'Own', which was entered as the covariant. As Figure 5 shows, the regression coefficients between Moral and Feelings, and between Authenticity and Feelings, were again significant. The analysis indicated a significant indirect effect, $-.33, 95\% CI [-.46, -.22]$. Thus, the present study indicates that the effect of the content of testimony (moral or non-moral) on Feelings is partially mediated via ratings of Authenticity.



Figure 5: The relationship between the independent variable (Moral), mediator (Authenticity) and dependent variable (Feelings), indicated by results to the moral deference and non-moral deference conditions (Conditions 1 and 3).

Finally, participants gave higher Authenticity ratings for non-moral no deference ($M=5.60$, $SD=.98$) than they did for moral no deference ($M=5.09$, $SD=1.47$), $t(388.25)=4.24$, $p < .001$; Cohen's $d = 0.43$. However, no difference was found in Feelings ratings between these two conditions.

Discussion: The first prediction (P1) was supported, as participants did not judge agents more harshly when they deferred to non-moral testimony than when they received that testimony but didn't defer to it. Of course, P1 is a negative prediction. Accordingly, there is a chance that the conclusion that there is no effect of Deference on Feelings, in the non-moral case, is a Type 2 error (that is, a false negative). However, given the sample size (445 participants for the non-moral pair of conditions), Study 2 had an 88% power to observe an effect size of $d = 0.3$. That is, there was an 88% chance of detecting a small effect of Deference on Feelings in the non-moral case. Although it is possible that the true effect size is even smaller, it seems reasonable to conclude that P1 was borne out.

Does the present study support P2? On the one hand, on average people tend to agree, more than disagree, that agents who defer to non-moral testimony are being true to themselves. This provides some support for P2. On the other hand, agents who defer to non-moral testimony are seen as less authentic than agents who think through the non-moral matter for themselves. Accordingly, the study did not show that there is no difference in the perceived authenticity of agents who defer to non-moral testimony and those who think the non-moral matter through for themselves. However, as the pairwise comparison between moral and non-moral deference indicates, people tend to feel more negatively about moral deference than they do about non-moral deference, in part because moral deference is seen as being in tension with authenticity.

3. General Discussion

There are important limitations of the present study. First, the range of scenarios featured in the present studies is limited. As such, the specific findings might not be representative of attitudes towards testimony in general. However, the moral conditions used in the study were adapted from one of the more famous examples from that literature. Accordingly, even if the hypotheses would not be borne out by examining a larger range of cases, the present results may indicate something about the intuitions that drive the academic debate over moral testimony.

The present studies support the Authenticity Hypothesis, according to which our attachment to the ideal of personal authenticity explains why it seems problematic to trust moral testimony, but not problematic to trust non-moral testimony. However, they do not support the idea that agents are equally authentic whether they defer to non-moral testimony or not. Rather, the studies provide some support for the idea that, while deferring to non-moral testimony makes one less authentic, only in the moral domain does deference to testimony put one in a position of inauthenticity, where this inauthenticity partly explains our disapproval of moral deference.

How do the results of the present study relate to the broader results of Andow's study? On the one hand, our intuitive understanding of moral testimony might be shaped by many factors, with the perceived inauthenticity of deferrers just being one of many reasons for which we are attracted to Pessimism about moral testimony. On the other, there may be a connection between judgments of authenticity/inauthenticity, and judgments about the epistemic status of beliefs gained from testimony. For instance, a suspicion that an agent acts inauthentically when she defers may be responsible for the sense that the resulting belief is illegitimate. It may be, too, that background perceptions of disagreement and deception in the moral domain partially mediate the increased perception that people are inauthentic when they trust moral testimony. On this point, it's worth noting that the non-moral scenario described in Conditions 3 and 4 of the Study 2 is plausibly subject to some disagreement—at least, when compared to some other kinds of non-moral testimony (for instance, the location of the city's stadium, or whether it's raining outside) it is easy to imagine reasonable people disagreeing about whether meat is best cooked on a searing pan. Perhaps this explains why participants were more likely to give lower Authenticity ratings for the non-moral deference condition, than for the non-moral no deference condition.

However, Andow's study suggested that higher perceived background levels of disagreement and deception in the moral domain are responsible for lower Legitimacy of beliefs formed via deference to moral testimony. It would be somewhat surprising if the same thing held of the non-moral issue that was recruited in Study 2. After all, participants in the present study were not more likely to disapprove of agents who defer to the non-moral testimony. Assuming, then, that the best way to cook meat is not entirely uncontroversial, the present findings provide some tentative reason to suspect that the perceived level of background disagreement, even if it partly explains Authenticity ratings, may not translate into lower Legitimacy ratings for all non-moral beliefs. It's not obvious, then, that perceived background levels of disagreement and deception are driving the asymmetry in our attitudes towards moral and non-moral testimony. However, future research is needed to confirm this one way or another.

Determining whether the present studies also support Pessimism about moral testimony would require getting clear on what sort of reason, if any, people have to be authentic/avoid being inauthentic, when it comes to moral matters. This would, in turn, require gaining clarity on what, exactly, authenticity entails. It is worth briefly sketching some alternative conceptions of authenticity that might underwrite the Authenticity Hypothesis and asking how each of these bear on the Optimist/Pessimist debate.

The first alternative is one Mogensen recruits and is suggested by empirical work on the "Essential Moral Self" (ESM) hypothesis (Strohming and Nichols 2014). According to the ESM hypothesis, people view a person's identity as defined by her moral sensibility. If the ESM hypothesis is correct, then being authentic would plausibly require acting in a way that accords with one's moral sensibility and, consequently, thinking moral matters through for oneself. Of course, if the ESM hypothesis is true, it is curious that agents who defer to non-moral testimony are also seen as being less authentic than agents who investigated the matter themselves. However, this finding does not necessarily pose a real problem for ESM views; perhaps one's identity is defined by more than one's moral sensibility, with the moral part being more central or somehow more relevant to practical matters. In any case, if the ESM hypothesis is correct, and if it informs the conception of authenticity that underwrites the present results, then whether we accept Pessimism will depend on whether we think the ESM hypothesis can withstand serious philosophical scrutiny—whether, in other words, it provides a good account of personal identity—as well as whether we think have good reason to act in ways that

are in accordance with our identity. On the former point, Mogensen (2021) himself is dubious. If Mogensen is right that authenticity cannot withstand philosophical scrutiny, then the results of the present studies would support Optimism about moral testimony.

A second conception of authenticity is described by Charles Guignon (2004: 76), who notes that at a certain stage in history, being authentic was thought to require having a distinct perspective on things, and thus differentiating oneself from “the herd.” If it is true that a commitment to authenticity is a commitment to differentiating oneself from others, then it is not difficult to see why moral deference is problematic. After all, the agent who defers to moral testimony will necessarily fail to have a unique perspective on the moral question at hand, and so is less likely to be differentiated from ‘the herd.’

It is hard to see what reason we could have for wanting people to have different moral beliefs from each other. Accordingly, the ethical value of authenticity *qua* differentiation appears unlikely to withstand serious normative scrutiny. If a commitment to this conception of authenticity is responsible for the results of the present study, then Mogensen’s suggestion that the Authenticity Hypothesis supports Optimism seems reasonable. Authenticity would be a spurious ideal, and so we should not accord our intuitions about moral testimony too much weight.

Note, however, that if a commitment to the differentiation is responsible for attitudes towards moral testimony, then we should expect deference to moral testimony to be associated with higher, rather than lower, feelings of approval.⁵ After all, given Andow’s finding that people perceive higher background levels of disagreement and deception in the moral domain, then deference about moral matters would be more likely to result in differentiation than would deference about non-moral matters. The results of Study 2, however, show just the opposite. This, then, is one reason to suspect that authenticity *qua* differentiation is not the folk conception of authenticity that participants have in mind.

However, a folk conception of inauthenticity may also be sufficient to explain the present results. The figure of the poser is one that immediately comes to mind as a paradigm of inauthenticity. In a

⁵ Thank you very much to an anonymous reviewer for pointing this out.

discussion of aesthetic testimony, Madeleine Ransom (2019: 430) describes the aesthetic poser as “someone who pretends to possess a certain competence or character trait in order to impress others or gain their acceptance or who—while relevantly competent to a certain degree—nevertheless attempts to pass off the testimony of others as the product of their own competence.” It is possible that intuitions about moral deference are driven by the suspicion that the agent who defers is engaging in the same sort of deception as does the aesthetic poser.^{6,7} In this sense, the deferrer to moral testimony is not being ‘true to herself’ insofar as she cares more about the opinions of other people than she does about acting in accordance with her own values. Accordingly, even if agents that defer to non-moral testimony are not equally as authentic as those who think the matter through for themselves, we do not view them negatively insofar as they aren’t seen to be posing.

Defending the idea that we should not be moral posers would not be difficult. Moreover, if an aversion to inauthenticity *qua* posing underwrites the results of the present study, that does not mean that those results support Pessimism. It is one thing to say that we are attracted to Pessimism because we suspect that agents are posing when they defer, and quite another to say that, necessarily, whenever an agent trusts moral testimony, she must be posing. It may be that, when it comes to subjects like morality, or at least the vignettes used to motivate Pessimism, we are particularly wary of posing.

In sum, there are at least two possible folk conceptions that could underwrite the present studies’ results. Insofar as these different conceptions may lead to different conclusions as regards moral testimony, it is far from clear whether the results of the present study support Pessimism or Optimism. Further research is therefore needed to substantiate the folk conception of authenticity/inauthenticity that explains the Authenticity Hypothesis.

References

- Andow, James. 2019. “Why Don’t We Trust Moral Testimony?” *Mind & Language*.
<http://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12255>.
- Enoch, David. 2014. “Authority and Reason-Giving.” *Philosophy and Phenomenological Research* 89: 296–332.

⁶ This is an explanation that Sliwa (2012) takes seriously.

⁷ Note that it is also a suspicion that could be motivating similar worries about aesthetic testimony, as Ransom suggests.

Forthcoming in *Oxford Studies in Experimental Philosophy, Volume 5*. Edited by Joshua Knobe and Shaun Nichols. New York: Oxford University Press (2022).

- Fletcher, Guy. 2016. "Moral Testimony: Once More with Feeling." In *Oxford Studies in Metaethics*, 11:45–73. Oxford, U.K.: Oxford University Press.
- Groll, Daniel, and Jason Decker. 2014. "Moral Testimony: One of These Things Is Just Like the Others." *Analytic Philosophy*.
- Guignon, Charles. 2004. *On Being Authentic*. Thinking in Action. London: Routledge, Taylor & Francis Group.
- Hills, Alison. 2009. "Moral Testimony and Moral Epistemology." *Ethics: An International Journal of Social, Political, and Legal Philosophy* 120: 94–127.
- Hopkins, Robert. 2007. *Philosophy and Phenomenological Research* 74: 611–34.
- Howell, Robert J. 2014. "Google Morals, Virtue, and the Asymmetry of Deference." *Noûs* 48: 389–415.
- Lewis, Max. 2020. "A Defense of the Very Idea of Moral Deference Pessimism." *Philosophical Studies*, 1–18.
- Lord, Errol. 2018. "How to Learn About Aesthetics and Morality Through Acquaintance and Deference." Edited by Russ Shafer-Landau. *Oxford Studies in Metaethics* 13: 71–97.
- McGrath, Sarah. 2009. "The Puzzle of Pure Moral Deference." *Philosophical Perspectives* 23: 321–44.
- Mogensen, Andreas L. 2017. "Moral Testimony Pessimism and the Uncertain Value of Authenticity." *Philosophy and Phenomenological Research* 95: 261–84.
- . 2021. "Is Identity Illusory?" *European Journal of Philosophy* 29: 55–73.
- Newman, George E., Julian De Freitas, and Joshua Knobe. 2015. "Beliefs About the True Self Explain Asymmetries Based on Moral Judgment." *Cognitive Science* 39 (1): 96–125.
- Nickel, Philip. 2001. "Moral Testimony and Its Authority." *Ethical Theory and Moral Practice* 4: 253–66.
- Ransom, Madeleine. 2019. "Frauds, Posers And Sheep: A Virtue Theoretic Solution To The Acquaintance Debate." *Philosophy and Phenomenological Research* 98: 417–34.
- Sliwa, Paulina. 2012. "In Defense of Moral Testimony." *Philosophical Studies* 158: 175–95.
- Strohming, Nina, and Shaun Nichols. 2014. "The Essential Moral Self." *Cognition* 131 (1): 159–71.

Forthcoming in *Oxford Studies in Experimental Philosophy, Volume 5*. Edited by Joshua Knobe and Shaun Nichols. New York: Oxford University Press (2022).