

Why don't we trust moral testimony?*

James Andow

February 25, 2019

Abstract

Is there a problem with believing something on the basis of someone else's testimony? Is there a problem with deferring to someone else's judgment? While it is typically okay to form beliefs in this way, many have thought there is something suspect about forming moral beliefs in this way. This basic intuition has been a starting point for much research on moral testimony and moral deference. To arbitrate between various attempts to account for our intuitions about moral testimony, it is important to know the exact nature of those intuitions. This is an empirical question and one which has not previously been addressed empirically. The current study attempts to get some empirical traction on how we intuitively think about moral testimony. The findings confirm that the way we ordinarily think about testimony about moral matters differs from the way we think about purely descriptive matters. Mediation analyses were used to explore the extent to which this asymmetry can be explained by appeal to different metaphysical beliefs about the relevant domains, or beliefs about the relative background levels of deception or disagreement. The current study also replicates and extends previous findings relating to a similar asymmetry between aesthetic and descriptive testimony.

1 Testimony and Asymmetry

We have many beliefs. Some of these beliefs are formed firsthand, on the basis of our own experience and without direct testimonial input from others. I believe that my neighbor's cat typically hangs out near the trampoline. I believe that Millias' *Ophelia* is part of the collection at the Tate Britain. I believe that torturing that cat would be morally abominable. I believe that *Ophelia* is captivating in its own way. Of course, these beliefs are not formed in isolation. I had to learn about what torture is and about the typical features of torture, and that information largely came to me via testimony. But the relevant testimony concerns the descriptive facts of the case; my belief that torturing cats is morally abominable is not based simply on

*This is a preprint copy of this paper. Please do not cite without permission. Thanks are due to a number of people whose comments on this and related work have been very helpful: Aimie Hope, Aaron Meskin, Shen-yi Liao, Jon Robson, the anonymous reviewers for this journal, and to various folks at various conferences and seminars with whom I discussed this work.

other people's testimony to that moral fact. I came to my own conclusion about the moral issue based on the descriptive facts rather than accepting the say-so of anyone else.

Some of my beliefs, however, are formed largely on the basis of direct testimony. I believe that one of my colleagues lives near somewhere called Didcot but have only his say-so to go on. I believe that Timișoara has many attractive buildings, but solely on my wife's testimony as I have never been there nor seen any photographs. I believe of many historical figures that they did many morally despicable things. However, in many cases, don't even really have a solid understanding of what it is these figures actually did which people suppose to be so wrong; I have simply accepted the moral judgments of others.

Many have argued that moral testimony (testimony about moral matters) is importantly different from descriptive testimony (testimony about descriptive matters) (see [Hills, 2013](#)). Similar arguments have been made about an asymmetry between aesthetic testimony (testimony about aesthetic matters) and descriptive matters (see [Robson, 2012](#)). One important putative difference concerns the epistemic credentials of beliefs formed solely on the basis of this testimony; in some way, is the thought, beliefs formed solely on the basis of moral (or aesthetic) testimony are in poorer epistemic standing than beliefs formed solely on the basis of descriptive testimony.

Like [Howell \(2014\)](#), who makes the point in a nice vivid way, I focus on the issue of *asymmetry* between moral testimony and descriptive testimony (as well as between aesthetic and descriptive testimony).¹

Although I don't know the streets and subways of New York very well, I get around with little difficulty... I have an application on my phone, Google Maps,

¹Two notes here: (a) There is a more restrictive use of the word 'asymmetry' in this debate, see, e.g., [Groll and Decker \(2014\)](#), to refer only to the idea that there is an *in principle* difference or a difference *in kind* between moral and descriptive testimony (such as that endorsed by [Hopkins, 2007](#)). I use the term to refer to any idea that moral/aesthetic beliefs formed on the basis of testimony are in poorer standing than descriptive beliefs formed on the basis of testimony (at least in general, or on average). (b) The debate about the 'problem' with moral testimony doesn't always explicitly focus on the issue of asymmetry. [Fletcher \(2016\)](#) deliberately avoids the comparative issue. [Sliwa \(2012\)](#) neatly rehearses an argument that moves from a non-comparative psychological claim to a non-comparative epistemic claim. The non-comparative psychological claim featuring in this rehearsed argument is that 'there is something intuitively problematic about deferring to others for one's moral beliefs' (p.175) which is supported by appeal to cases that only mention the moral domain, e.g, [Hills \(2009\)](#)'s case of Eleanor who, rather than thinking through the moral arguments about meat eating for herself, defers to the judgment of her trustworthy friend and accepts that eating meat is wrong. The relevant non-comparative epistemic principle which is supposed to explain the psychological claim is that 'there is something wrong with relying on testimony for one's moral beliefs even if one knows one's source to be reliable and trustworthy' ([Sliwa, 2012](#), p.176). Nonetheless, although the debate is sometimes framed in ways that don't explicitly make comparisons or talk about asymmetries, I am inclined to think that the philosophically interesting issue is always a comparative one. [Fletcher \(2016\)](#)'s account of the problem with moral (and some other types of) belief is one which appeals to factors which are conspicuously not in play for most everyday descriptive beliefs. And, in the argument rehearsed in [Sliwa \(2012\)](#), the interest of the psychological claim is largely that an analogous claim about the descriptive domain is false, and likewise for the epistemic principle supposed to explain the psychological claim. [Sliwa \(2012\)](#) herself agrees that the philosophically interesting issue concerns an *asymmetry*, concluding that 'moral testimony is no more problematic than nonmoral testimony' (p.194).

which gives me directions and directs me to restaurants at the push of a few buttons. . . it's a perfectly reasonable way to get around.

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 quandary 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. Never again will I be paralyzed by the prospect of pushing that fat man onto the trolley tracks to prevent five innocents from being killed. I'll just Google it. . . in this case it seems to many, myself included, that this is not a good way to go. There seems to be something wrong with using Google Morals. But what is it?

The issue, of course, is whether or not there is something particularly wrong with forming beliefs based on deference in moral matters. Is there an asymmetry of deference between the moral and the non-moral? If there is, what is the nature of the asymmetry?

The basis of the philosophical debate about moral testimony is almost always a psychological claim: a claim about how we tend to treat testimony about moral matters (in comparison with others). One can see a similar pattern in the passage quoted from [Howell \(2014\)](#) above: a move from a claim about something *seeming wrong* with moral deference to the issue of whether there is something wrong with forming beliefs in this way.² This psychological asymmetry is taken to demand an explanation. One might seek to explain the psychological asymmetry away as a mistake. Although we might typically treat moral deference as problematic, the argument might go, it is not and then an account might be given as to where we tend to go wrong. Or one might try to explain away the appearance of a psychological asymmetry. Although we do treat the moral and descriptive cases typically appealed to by philosophers differently, the argument might go, this is due to some systematic bias as to the kinds of cases philosophers typically appeal to.³ More typically, however, the appearance of a psychological asymmetry is accounted for in terms of a genuine epistemic difference between accepting a moral claim on the basis of testimony and accepting a descriptive claim on the basis of testimony: the former tend to be in poorer epistemic standing than the latter.

There are various different kinds of epistemic asymmetry which might hold between moral and descriptive testimony. Perhaps, while beliefs formed solely on the basis of moral or aesthetic testimony are typically (or always) in poor epistemic standing, beliefs formed solely

²There are others in the debate who approach the issue in different ways. For example, [McShane \(2018\)](#) argues that 'dependence on moral testimony is not simply for those who cannot realize the ideals of moral agency; it is among those ideals.'

³[Groll and Decker \(2014\)](#), for example, suggest that philosophers typically appeal to cases involving 'normal' moral propositions—those we expect people to know in virtue of 'having the know-how that's required to navigate the world in the way that we're expected to navigate it qua human beings'—but 'non-normal' non-moral ones.

on the basis of descriptive testimony are typically (or always) completely fine. Perhaps, they differ simply in terms of how likely beliefs are to achieve a certain standard if formed in this way. Perhaps there is asymmetry in terms of the relation between beliefs formed on the basis of testimony and beliefs formed firsthand. For example, one might say that the extent to which the epistemic standing of moral or aesthetic beliefs formed solely on the basis of testimony falls behind that of beliefs in the relevant domain formed firsthand is greater than the extent to which the epistemic standing of testimony-based descriptive beliefs fall behind firsthand descriptive beliefs. Moreover, there are various different kinds of ‘epistemic standing’ which might be important. There might be an asymmetry in terms of being properly formed, being justified, amounting to knowledge, constituting understanding, or something else. Among those who think there is an epistemic asymmetry between moral and descriptive beliefs there is no consensus as to the exact nature of that asymmetry.⁴ Likewise, among those who think there is a psychological asymmetry it is not always clear what the precise contours of the phenomenon are supposed to be.

This paper will be mainly focused on the psychological claim which has been used to make the case that there is an epistemic asymmetry: that there is an asymmetry (of some kind) in the way we treat moral and descriptive testimony. However it is important to note that the hope of my project is not simply that establishing whether there is a psychological asymmetry will help arbitrate in the debate about whether there is an epistemic asymmetry. As [Andow \(2018\)](#) demonstrates, the techniques of experimental philosophy can potentially be used to provide more subtle insights that can help advance debates about testimony.

The empirical project of this paper also hopes to inform debate about what *explains* the epistemic asymmetry (if there is one) by getting clearer on the nature of the psychological asymmetry (if there is one). There are various different explanations that have been offered each of which has some plausibility. Perhaps adequate moral expertise is simply in comparatively short supply, or more difficult to identify, and so the moral testimony one encounters tends to be less reliable across the board (discussed, e.g., by [Hills, 2009](#)), one sign of this might be the comparatively high levels of disagreement about moral matters perhaps among the experts (discussed, e.g., in [Enoch, 2014](#)). Perhaps people are less reliable sources of moral testimony because they are more likely to lie or mislead when it comes to moral matters.⁵ Perhaps, within unjust social structures, deference to the apparently reliable sources could lead to a systematically distorted moral beliefs (and perpetuating the unjust social hierarchies) ([Dular, 2017](#)). Perhaps moral realism is false and some form of constructivism, relativism, subjectivism, noncognitivism or similar is true and either of those seems potentially

⁴Throughout much of the philosophical debate some of the relevant distinctions here articulated in terms of a contrast between ‘optimism’ and ‘pessimism.’ I won’t use those terms as they are used in so many different ways that their use invites confusion (see [Andow, 2018](#), for some discussion in relation to aesthetic testimony).

⁵[Andow \(2018\)](#) considers a similar idea with respect to aesthetic testimony but the idea has some traction here too. One might suspect that people hold back their genuine moral opinions for fear of causing offense, interfering, or being subject to social sanctions. Or one might suspect that disingenuous virtue signaling is frequently in one’s self-interest.

able to provide some explanation (Mcgrath, 2011). Perhaps moral knowledge is knowledge-how which you can't get from testimony (discussed in Hopkins, 2007). Perhaps deference is unable to provide the sentiments or emotions that are in some way important for moral judgment (Enoch, 2014; Fletcher, 2016). Perhaps what we value in moral agents is moral understanding (or whether they have grasped the moral grounds for their beliefs) rather than simple moral knowledge and getting moral beliefs via testimony is only sufficient for the latter (whereas in the descriptive domain either understanding is unimportant or easy to get via testimony) (Callahan, 2018; Hills, 2009; Hopkins, 2007). There might also be other factors in play that are less obviously epistemic. Perhaps deciding moral questions on one's own is important for the desirable goal of authenticity (Mogensen, 2017). Perhaps acquiring moral beliefs via testimony gets into trouble because it doesn't have a strong enough link to virtue (Howell, 2014; Jones, 1999; Mcgrath, 2011). I say that these factors are 'less obviously epistemic' rather than 'non-epistemic' because these explanations might be coupled with the idea that *one doesn't really know* unless one has achieved the relevant state (this is the position Mcgrath, 2011, takes, for example). Finally, and importantly, it may well be that two or more of these factors are active in tandem. There are doubtless many more possible explanations one could come up with.

How might an empirical study help arbitrate between these explanations? To give you a sense of how that might go, consider the following: if ordinary folks' treatment of testimony were found to correlate with perceptions of the background levels of expertise in the relevant domains, that finding might be used in the case for the claim that a difference in available expertise is what underlies an epistemic asymmetry. Such an argument wouldn't be completely straightforward, of course, as perceptions of background levels of expertise across various domains could be wildly miscalibrated. Nonetheless, a good understand of the psychological asymmetry (if there is one) between moral and descriptive testimony will be a valuable tool in the debate.

There is some extant evidence relating to non-philosophers' attitudes about aesthetic matters which will be outlined in the next section. So far, however, no similar empirical investigation has been extended to moral testimony. This paper presents an empirical investigation into deference to others with respect all three domains of belief: *descriptive*, *moral*, and *aesthetic* beliefs. The aims of the study reported in this paper were to (i) replicate Andow (2018) findings concerning aesthetic testimony in a different design, (ii) extend this new design to examine non-philosophers' judgments about moral testimony, and (iii) investigate the extent to which these asymmetries can be explained by three specific factors.

2 Previous empirical work on aesthetic testimony

There is no extant empirical work that speaks directly to the issue of the claimed psychological asymmetry between moral and descriptive testimony. Some studies, however, have been conducted that look at the analogous issue of a claimed psychological asymmetry be-

tween aesthetic and descriptive testimony. Although one should not necessarily expect the two issues to align exactly in the final analysis, the two philosophical debates have a similar structure and so these previous studies are a helpful starting point nonetheless.

In the first studies on aesthetic testimony, [Meskin et al. \(in preparation\)](#) ask to what extent participants treat aesthetic and descriptive testimony as having different value. This question is pursued across a number of studies. Participants read movie reviews and rated their confidence in various claims the review made about the movie (confidence being a proxy for judgments about epistemic justification). Confidence ratings were higher for purely descriptive claims than aesthetic claims. Participants read vignettes in which one person makes a claim and there is a question about how confident another person (either the participant or a third party) should be about that claim, or about whether a belief formed on the basis of this testimony would count as knowledge. [Meskin et al. \(in preparation\)](#) found that ratings for beliefs formed via testimony in aesthetic cases were consistently lower than those for descriptive cases; in the case of knowledge ratings the mean rating tended to fall below the midpoint of the relevant scale, but it did not for ratings of confidence ('neither confident nor unconfident').

There are various questions remaining following [Meskin et al. \(in preparation\)](#)'s studies. [Andow \(2018\)](#) attempts to address some of them. A first question concerns a different kind of epistemic judgment—whether a judgment is properly formed, legitimately formed, whether it is permissible to form a judgment—that also plays a role in the philosophical literature on aesthetic judgment (prominently in [Hopkins, 2007](#)). Meskin et al.'s studies don't speak to whether such judgments display a similar asymmetry. So, in a first study, [Andow \(2018\)](#) asked participants direct questions such as:

Suppose [*an expert/a friend/firsthand experience*] tells you that a particular painting [*is beautiful/is ugly/is large/cost \$14 million to create*]. Is it legitimate to rely on what [*the expert/the friend/firsthand experience*] tells you when forming your own opinion about whether the painting [*is beautiful/is ugly/is large/cost \$14 million to create*]?

The italics in the above indicate the levels of the two independent variables, domain and source, that were manipulated between subjects. Andow finds that participants' mean responses for aesthetic cases are well below those for descriptive cases and, notably, the midpoint of the relevant scale.

Another issue Meskin et al.'s studies don't speak to is that of why participants display these various asymmetries in their treatment of aesthetic and descriptive testimony. Andow explores three potential explanations. In a second study, he finds no evidence that the extent to which participants endorse a position called 'normativism' is related to the size of the asymmetry in their judgments about knowledge or legitimacy. Normativism is the position that judgments in a particular domain "'claim" for universal validity in the sense that when one formulates [a]...judgement, she considers her judgement as being either correct or incorrect so that, in a contradictory...debate, each side believes that at most one of them is right" ([Cova and Pain, 2012](#)). Andow's study measured beliefs about attitudes to norma-

tivism using a scale adapted from [Cova and Pain \(2012\)](#) in which participants consider a case of aesthetic disagreement (in which two protagonists hold p and not- p respectively) and then choose between the following options: (a) one party is right and the other is not; (b) both are right; (c) both are wrong; (d) Neither is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to their own opinion. Option (a) is the only one consistent with Normativism. In [Andow \(2018\)](#)'s study participants rate all four options, a normativism-rejection composite score was created (subtracting the ratings for a from the sum of b , c , and d), and are split into high and low normativism rejection groups on the basis of this composite score. The two main independent variables, domain and source, were manipulated between subjects. Legitimacy was measured as in the first study and knowledge using an similar form of question. No important differences between the high and low normativism rejection groups were observed.

In a third study, [Andow \(2018\)](#) does find evidence that participants' perceptions of the relative background levels of disagreement and deception are related to the size of the asymmetry between domains for knowledge ratings. Participants' perceptions of these background levels were measured using the following items:

1. People disagree about whether a particular painting is beautiful
2. People disagree about whether a particular object is large
3. One person sincerely thinks a painting is good while someone else sincerely thinks the very same painting is terrible
4. One person sincerely thinks a painting is large while someone else sincerely thinks the very same painting is tiny
5. Some people like a particular work of art while other people don't like it
6. Some people think a particular object is large while other people think it isn't
7. People present themselves as being a good judge whether something is beautiful, when in fact they are not a good judge
8. People present themselves as being a good judge of whether something is large, when in fact they are not a good judge
9. People present themselves as knowing a lot about art/book/music when in fact they do not know a lot about art
10. People present themselves as knowing a lot about the size of objects when in fact they do not know a lot about the size of objects
11. People lie about liking art
12. People lie about thinking things are large

Two composite scores were created, one for *disagreement difference*⁶ and one for *deception difference*.⁷ Again, domain and source were manipulated between subjects and knowledge and legitimacy were measured in the same way. High levels on both DisDiff and DecDiff were associated with larger two way interactions between domain (aesthetic vs descriptive) and source (firsthand vs testimony) for knowledge ratings indicating that the perception that levels of disagreement and deception are higher with respect to aesthetic than descriptive matters may be an important part of reason for the asymmetry in judgments about the epistemic status of beliefs formed on the basis of testimony across the two domains.

3 New empirical work extending to moral testimony

The current study represents the first attempt to get empirical traction on the claim of a psychological asymmetry between moral and descriptive testimony. Since one might expect the asymmetries between descriptive and (i) aesthetic and (ii) moral testimony to behave somewhat similarly, the main question the current study aims to address is as follows. Will the patterns observed in previous research concerning the aesthetic-descriptive contrast extend to the moral-descriptive contrast? Since there are various aspects to the patterns observed for the aesthetic-descriptive contrast, it is worth itemizing them clearly in the form of hypotheses.

3.1 Hypotheses

The first set of predictions is that judgments about the epistemic status of beliefs will be influenced by (a) *source* of the beliefs or the basis on which the beliefs are formed (with beliefs formed on the basis of testimony being judged to be in worse standing than those formed on the basis of firsthand experience), (b) the *property* the beliefs concern (with descriptive beliefs being considered to be in better epistemic standing than moral beliefs), and (c) the *interaction* of (a) and (b) (such that the status of beliefs formed on the basis of testimony and moral matters will be judged to be even worse than is explained by the combination of the two main effects).

The second set of predictions concerns the influence of three other factors. The three factors are beliefs about the background level of deception, beliefs about the background level of disagreement and metaphysical beliefs. The ‘metaphysical’ questions in the current study are intended to gauge the extent to which participants favour realism or anti-realism about the relevant domain. These questions are similar to the measures of ‘normativism’ used in previous work (Andow, 2018; Cova and Pain, 2012). Although Andow (2018) did not find evidence that (as he called it) attitudes toward normativism were an important influence in the case of the contrast between aesthetic and descriptive testimony, the prediction tested

⁶DisDiff: 2 + 4 + 6 - 1 - 3 - 5

⁷DecDiff: 8 + 10 + 12 - 7 - 9 - 11

in the current study is that each of these three attitudes will explain some of the asymmetries predicted to be observed between legitimacy and knowledge ratings for beliefs formed via moral testimony.

The study also examines the aesthetic-descriptive contrast to facilitate a better comparison between the two in the different design of the current study.

3.2 Participants

Participants took part in a web-based study created using Qualtrics in October 2016. Participants were recruited using Amazon MTurk. The survey was kept open until 301 respondents had successfully completed the survey without failing attention checks (an accidental over-recruitment of one participant relative to a set target of 300 participants). The number of participants who initially consented to take part was 373. Of these, 72 participants failed to complete the survey or else failed an attention check.⁸ Of the remaining participants, 178 (40.1%) were Male and the rest Female, 90 (29.9%) indicated some experience of studying philosophy at university level. The average age of participants was 37.81 years.

3.3 Method and Materials

Participants were randomly assigned to receive questions about one of three types of property: descriptive, moral, or aesthetic.⁹ Each participant then received three blocks of questions in a randomized order: deception and disagreement questions; metaphysical questions; and epistemic questions. Finally, participants were asked to provide basic demographic information.

The disagreement and deception questions were designed to measure the extent to which participants thought it was common for people to be deceptive with respect to the relevant type of property and for people to disagree about whether things had the relevant type of property. Participants were asked to rate their level of agreement three items concerning deception and three about disagreement on a 200-point sliding scale from ‘Almost always happens’ to ‘Almost never happens’. The order of items was randomized. The full items were similar to those used in [Andow \(2018\)](#).

⁸The attention checks were as follows. Check 1: ‘Suppose that firsthand experience tells you that a particular painting is green. Please ignore the question below and select the answer on the far right of the scale. Do you know whether the painting is large?’ Check 2: ‘Suppose that an expert tells you that a particular painting is beautiful. Please ignore the question below and select the answer on the far left of the scale. Do you know whether the painting is large?’ Participants’ responses were excluded if their response to the attention checks was not within 10 points of the relevant end of the scale.

⁹The descriptive property used is ‘size’ which is context-sensitive. One worry one might have had I not chosen a context-sensitive term is that any asymmetry found between moral and descriptive testimony—such that beliefs formed on the basis of moral testimony are in poorer epistemic standing—was attributable to a contrast between non-context-sensitive descriptive matters and (putatively) context-sensitive moral matters. The choice of ‘size’ guards against this worry.

- dec1** Someone lies about whether [they like or dislike a particular piece of art / they think a particular object is large or not large / they think a particular act is moral or immoral]
- dec2** Someone presents themselves as knowing a lot about [art, books and music / the size of objects / the morality of actions] when in fact they do not know a lot about [those things / the size of objects / the morality of actions]
- dec3** Someone presents themselves as being a good judge of whether [something is beautiful / something is large / an action is moral] when in fact they are not a good judge
- dis1** Many people think a particular [work of art is beautiful / a particular object is large / a particular action is moral] while many other people think it isn't
- dis2** One person sincerely thinks [a painting is good / a painting is large / an action is moral] while someone else sincerely thinks [the very same painting is terrible / the very same painting is tiny / the very same action is immoral]
- dis3** People disagree about whether [a particular painting is beautiful / a particular object is large / a particular action is moral]

The metaphysical questions were designed to measure the extent to which participants are favour realism or anti-realism about the relevant domain, in particular whether they thought that there was a fact of the matter about which party is correct in cases of disagreement about the relevant type of property. Participants were presented with the appropriate version of the disagreement case:

John and Fred are in an argument. John says, ["The Empire State building is large," / "Van Gogh's sunflower paintings are beautiful," / "It's okay to hit people just because you feel like it,"] and Fred says, ["No, The Empire State building is not large." / "No, Van Gogh's sunflower paintings are not beautiful." / "No, it is not okay to hit people just because you feel like it."]

Participants were then asked to indicate which of the following best characterized John and Fred's views, before being asked to rate each item individually on sliding scale from 'Completely incorrect' (-100) to 'Completely correct' (100). The order of these items was as below in order to avoid confusion between the two tasks. These response options were altered appropriately according which case of disagreement participants considered.

- meta1** [Van Gogh's sunflower paintings are beautiful], so John is right and Fred is wrong.
- meta2** [Van Gogh's sunflower paintings are not beautiful], so Fred is right and John is wrong.
- meta3** I don't know whether [Van Gogh's sunflower paintings are beautiful], so I don't know which of them is right and which is wrong, but one of them is wrong.

meta4 There is no fact of the matter about unqualified claims like [“Van Gogh’s sunflower paintings are beautiful.”] Different people believe different things, and it is not absolutely true or false that [Van Gogh’s sunflower paintings are beautiful].¹⁰

The epistemic questions were designed to measure participants’ attitudes to the epistemic status of beliefs formed on the basis of first hand experience versus on the basis of testimony. Participants were presented with four cases and the two attention checks in a randomized order and asked to rate the extent to which they thought the relevant beliefs were (a) legitimate, and (b) counted as knowledge on similar sliding scale from either ‘Completely illegitimate’ or ‘Definitely don’t know’ to ‘Completely legitimate’ or ‘Definitely do know’. The cases took the following form:

Legitimate (L) Suppose that [a friend / firsthand experience] tells you that a particular [painting is large / painting is beautiful / action is immoral]. Is it legitimate to rely solely on what [the friend / firsthand experience] tells you when forming your own opinion about whether the [painting is large / painting is beautiful / action is immoral]?

Knowledge (K) Suppose that [a friend / firsthand experience] tells you that a particular [painting is large / painting is beautiful / action is immoral] and that you base your opinion about whether the [painting is large / painting is beautiful / action is immoral] solely on what [the friend / firsthand experience] tells you. Do you know whether the [painting is large / painting is beautiful / action is immoral]?

3.4 Descriptive Results

Descriptive results for all questions by condition are in [Table 1](#).¹¹ The descriptive results for knowledge and legitimacy ratings for all conditions are displayed in [Figure 1](#). There was good internal reliability among the three deception ($\alpha = .77$) and three disagreement items ($\alpha = .88$). Two composite scores for deception and disagreement were created by taking the mean across the relevant three items.

The distribution of responses for the metaphysical items of interest were not suitable for linear modeling (containing many extreme values). A composite score was created by taking the highest response for each participant across M1–3 minus their response to M4, representing their overall attitude towards the idea that there is an objective fact of the matter in disagreement cases. The result was a variable which was still not suitable as it retained too many extreme values and had signs of a bimodal distribution. For the purposes of the mediation analyses reported below, this variable was transformed (using an inverse normal function applied to percentile ranked composite scores).

¹⁰The items here are slightly different than in previous work ([Andow, 2018](#); [Cova and Pain, 2012](#)) in order to accommodate (a) the fact that the disagreement concerns something with which the participants are likely familiar, and (b) worries about number of different themes captured by the fourth item (the fourth option here is based instead on an item from [Cohen and Nichols, 2010](#), a study on the folk metaphysics of colour).

¹¹All analysis conducted in SPSS 22 using the PROCESS 3 macro for the mediation analyses.

Table 1: Means and SDs for all questions by condition

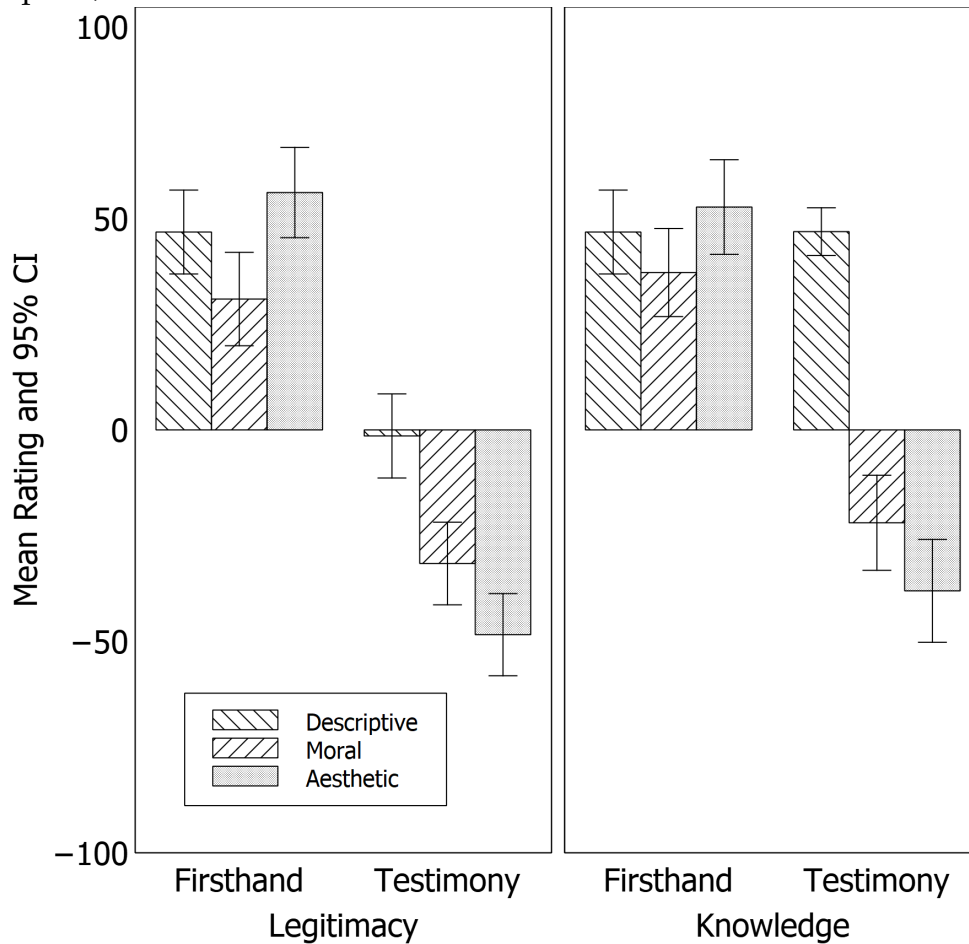
Question	Descriptive	Moral	Aesthetic
dec1	-21.25 (55.12)	15.42 (47.89)	17.44 (52.38)
dec2	0.50 (54.04)	36.31 (42.28)	30.74 (44.64)
dec3	6.79 (52.06)	42.02 (41.18)	25.25 (47.36)
dis1	-0.96 (50.05)	29.73 (49.52)	47.29 (46.06)
dis2	-15.63 (56.44)	35.00 (46.31)	56.54 (41.37)
dis3	7.79 (52.17)	46.39 (43.29)	57.63 (40.19)
meta1	43.02 (59.65)	-83.38 (36.93)	-44.47 (58.72)
meta2	-62.18 (40.37)	84.51 (32.53)	-54.36 (51.15)
meta3	-51.32 (57.15)	-56.44 (60.86)	-37.56 (66.53)
meta4	19.29 (73.19)	-42.57 (66.01)	89.00 (26.08)
firsthandL	46.79 (49.59)	30.94 (57.74)	56.13 (54.45)
testimonyL	-1.43 (49.72)	-31.59 (51.09)	-48.42 (49.58)
firsthandK	45.86 (54.06)	37.21 (54.43)	52.70 (57.07)
testimonyK	46.89 (28.18)	-21.96 (58.76)	-38.06 (62.03)

3.5 Moral versus Descriptive

Is there an asymmetry in responses between the Moral and Descriptive domains? A 2×2 MANOVA was conducted to investigate the main effects of the between-subjects factor property (moral versus descriptive) and the within-subjects source (testimony versus firsthand) and the interaction effect on legitimacy and knowledge ratings. The interaction was significant ($F(2, 198) = 25.80, p < .0005, \eta_p^2 = .21$). There were significant main effects of source ($F(2, 198) = 68.41, p < .0005, \eta_p^2 = .41$) and property ($F(2, 198) = 23.95, p < .0005, \eta_p^2 = .20$). The nature of these main effects and interaction was then explored using the following univariate tests and follow-up t-tests. The reports of these univariate tests and follow-up tests explain the nature of the main effects and interactions. You can also use the visual representation of the results [Figure 1](#) to guide interpretation. Readers interested in the qualitative nature of the findings rather than the statistical details can skip (if they wish) ahead to the discussion.

Legitimacy Univariate tests for legitimacy ratings find no significant interaction ($p = .13$). However, they do find a significant main effect of source ($F(1, 199) = 137.42, p < .0005, \eta_p^2 = .41$) and property ($F(1, 199) = 16.48, p < .0005, \eta_p^2 = .08$). Paired samples t-tests reveal that legitimacy ratings for firsthand are significantly higher than ratings for testimony for both moral ($t(104) = 9.37, p < .0005, d = 1.146$) and descriptive properties ($t(95) = 7.24, p < .0005, d = 0.971$). Pairwise comparisons using independent samples t-tests reveal that legitimacy ratings are lower for moral properties than descriptive for both testimony ($t(199) = 4.24, p < .0005, d = .598$) and firsthand ($t(198.25) = 2.09, p = .038, d = .294$).

Figure 1: Mean ratings and 95% confidence intervals for legitimacy and knowledge ratings across descriptive, moral and aesthetic domains.



Knowledge Univariate tests for knowledge ratings reveal a significant interaction ($F(1, 199) = 45.63, p < .0005, \eta_p^2 = .19$), as well as a main effect of source ($F(1, 199) = 42.59, p < .0005, \eta_p^2 = .18$) and property ($F(1, 199) = 48.02, p < .0005, \eta_p^2 = .19$). Pairwise comparisons were also used to explore the nature of the interaction between property and source for knowledge ratings. Pairwise comparisons using paired samples t-tests reveal that knowledge ratings for firsthand are significantly higher than ratings for testimony for moral properties ($t(104) = 8.61, p < .0005, d = 1.044$), but not for descriptive properties ($p = .85, d = 0.023$). Pairwise comparisons using independent samples t-tests reveal that knowledge ratings for testimony are lower for moral properties than descriptive ($t(152.35) = 10.73, p < .0005, d = 1.473$), the same is not true for firsthand ($p = .26, d = .159$).

3.6 Aesthetic versus Descriptive

Does the current study replicate the patterns of asymmetry found by previous work? A 2×2 mixed MANOVA was conducted to investigate the main effects of property (aesthetic versus descriptive) and source (testimony versus firsthand) and the interaction effect on legitimacy

and knowledge ratings. The interaction was significant ($F(2, 193) = 44.54, p < .0005, \eta_p^2 = .32$). There were significant main effects of source ($F(2, 193) = 121.79, p < .0005, \eta_p^2 = .56$) and property ($F(2, 193) = 23.99, p < .0005, \eta_p^2 = .20$).

Legitimacy Univariate tests for legitimacy ratings reveal a significant interaction ($F(1, 194) = 32.99, p < .0005, \eta_p^2 = .15$), as well as a main effect of source ($F(1, 194) = 242.66, p < .0005, \eta_p^2 = .56$) and property ($F(1, 194) = 12.29, p = .001, \eta_p^2 = .06$). Pairwise comparisons were used to explore the nature of the interaction between property and source for legitimacy ratings. Paired samples t-tests reveal that legitimacy ratings for firsthand are significantly higher than ratings for testimony for aesthetic properties ($t(99) = 14.57, p < .0005, d = 2.008$) and for descriptive properties (as reported above). Pairwise comparisons using independent samples t-tests reveal that legitimacy ratings for testimony are lower for aesthetic properties than descriptive ($t(194) = 6.63, p < .0005, d = .947$), there is no difference between properties for firsthand ($p = .211, d = .179$).

Knowledge Univariate tests for knowledge ratings reveal a significant interaction ($F(1, 194) = 88.27, p < .0005, \eta_p^2 = .31$), as well as a main effect of source ($F(1, 194) = 84.39, p < .0005, \eta_p^2 = .30$) and property ($F(1, 194) = 48.06, p < .0005, \eta_p^2 = .20$). Pairwise comparisons were also used to explore the nature of the interaction between property and source for knowledge ratings. Paired samples t-tests reveal that knowledge ratings for firsthand are significant higher than ratings for testimony for aesthetic properties ($t(99) = 11.38, p < .0005, d = 1.522$), but not for descriptive properties (as reported above). Pairwise comparisons using independent samples t-tests reveal that knowledge ratings for testimony are lower for aesthetic properties than descriptive ($t(139.43) = 12.42, p < .0005, d = 1.751$), there is no difference between properties for firsthand ($p = .391, d = .123$).

3.7 Discussion

The first set of predictions was that judgments about the epistemic status of beliefs would be influenced by (a) *source* of the beliefs or the basis on which the beliefs are formed (with beliefs formed on the basis of testimony being judged to be in worse standing than those formed on the basis of firsthand experience), (b) the *property* the beliefs concern (with descriptive beliefs being considered to be in better epistemic standing than moral beliefs, and aesthetic beliefs), and (c) the *interaction* of (a) and (b) (such that the status of beliefs formed on the basis of testimony and about moral or aesthetic matters will be judged to be even worse than is explained by the combination of the two main effects).

These predictions were borne out. There were main effects of source and property, as well as an interaction of the kind described. To illustrate, beliefs formed on the basis of firsthand experience tended to be judged to be in better epistemic standing than beliefs formed on the basis of testimony across all three domains both with respect to legitimacy and knowledge

(with one exception for knowledge ratings concerning descriptive beliefs which were not significantly different between testimony and firsthand). Beliefs about morality and aesthetics tended to be judged to be in poorer epistemic standing than beliefs about descriptive matters in testimony cases with respect to both legitimacy and knowledge ratings, but not in firsthand cases (exception: firsthand legitimacy ratings are lower for morality than descriptive).

So far, there are important consistencies in the patterns of results in the current study and those previously found by [Meskin et al. \(in preparation\)](#) and [Andow \(2018\)](#). In particular, one can note the following: (i) judgments about legitimacy and knowledge are consistently below the mid-line for aesthetic beliefs formed via testimony, and this pattern extends to moral beliefs in the current study; (ii) judgments about legitimacy and knowledge are consistency around or above the mid-line for descriptive beliefs; (iii) legitimacy and knowledge ratings are consistently above the mid-line for all domains in the firsthand case.

3.8 Mediation Analyses

What explains the observed effects of domain? To what extent are they explained by differences in beliefs about (a) background levels levels of disagreement, (b) background levels of deception, (c) metaphysics of the relevant properties?

A series of mediation analyses were conducted in order to examine the extent to which the asymmetries observed between domains in terms of ratings of beliefs formed via testimony can be explained by three potential mediators (disagreement, deception, metaphysics).¹² Mediation models were run separately for each of the two relevant contrasts between domains of belief (aesthetic vs descriptive, moral vs descriptive) and for the two epistemic questions concerning testimony from a friend (knowledge, legitimacy).¹³ In total twelve regression models

¹²See [MacKinnon et al. \(2007\)](#) for an accessible introduction to mediation effects. To help out the uninitiated, let me give a quick example. Suppose it is found studying philosophy (X) is related to dying prematurely (Y). There are various hypotheses one might formulate. One says that those who are very ill take up philosophy (that X and Y have a common cause). One says that studying philosophy is itself bad for your health (that X causes Y). Other propose an effect of studying philosophy on life expectancy which is mediated by some third variable (X causes M which causes Y). Perhaps studying philosophy is a cause of a sedentary life-style and it is this life-style that reduces life-expectancy. Such mediation models can be tested by running a series of regressions.

¹³A quick note should be made on the choice of models. Intuitively one might expect any mediating effect of, for example, disagreement to go via ratings of the firsthand case. The thought might be that domain affects disagreement ratings which in turn influence ratings of beliefs in the relevant including straightforward firsthand beliefs. Accordingly, initially, mediation models were created that contained the firsthand case as a second mediator. However, the importance of the second mediator was negligible in such models; notably the total indirect effect in those models was not larger than in the simpler single-mediator models which are the focus of the analysis here.

Table 2: Results of twelve mediation analyses. The first three columns indicate the independent (X), mediator (M), and dependent (Y) variables in the model. In each model, M and Y were standardized. ‘Aes’, ‘Des’ and ‘Mor’ stand for Aesthetic, Descriptive, and Moral. ‘Dis’ and ‘Dec’ stand for Disagreement and Deception composite scores. ‘L’ and ‘K’ stand for knowledge and legitimacy ratings. The next five columns contain path weights and respective p-values, 95% CI (bootstrapped using 5000 samples) for the indirect effect, where an asterisk indicates no overlap with zero. The final column contains the proportion of the total effect of X on Y explained by the indirect effect via M.

Variables in model			Total effect	Path coefficients			Indirect (95% CI)	Indirect (as % of total)
X	M	Y		X→M (p)	Direct (p)	M→Y (p)		
Aes/Des ¹⁵	Dis	L	.44 (< .01)	- .60 (< .01)	.27 (< .01)	-.29 (< .01)	.17* (.075, .267)	39.15
	Dec			-.35 (< .01)	.35 (< .01)	-.27 (< .01)	.09* (.040, .159)	21.29
	Meta			.54 (< .01)	.29 (< .01)	.28 (< .01)	.15* (.054, .255)	34.75
	Dis	K	.67 (< .01)	- .60 (< .01)	.59 (< .01)	-.13 (.03)	.08 (.000, .155)	11.58
	Dec			-.35 (< .01)	.64 (< .01)	-.09 (.09)	.03 (-.004, .081)	4.90
	Meta			.54 (< .01)	.55 (< .01)	.21 (< .01)	.11* (.037, .188)	16.95
Mor/Des ¹⁶	Dis	L	.28 (< .01)	- .42 (< .01)	.18 (.01)	-.24 (< .01)	.10* (.035, .173)	35.83
	Dec			-.43 (< .01)	.18 (.01)	-.23 (< .01)	.10* (.038, .171)	35.40
	Meta			-.42 (< .01)	.35 (< .01)	.16 (.04)	-.07 (-.147, .001)	-24.43
	Dis	K	.54 (< .01)	- .42 (< .01)	.51 (< .01)	-.08 (.17)	.03 (-.014, .082)	6.15
	Dec			-.43 (< .01)	.51 (< .01)	-.06 (.24)	.03 (-.022, .076)	5.02
	Meta			-.42 (< .01)	.57 (< .01)	.07 (.27)	-.03 (-.082, .029)	-5.27

were run. The results of the mediation analyses can be seen in [Table 2](#).¹⁴

Before taking a more fine-grained look, there are three coarse-grained observations it is helpful to make straight away. First, there is much clearer evidence of mediation in the case of legitimacy ratings than knowledge ratings. Second, the contribution of metaphysical beliefs seems to be rather different across the two contrasts between domains. Third, there is very little evidence of mediation in the case of moral vs descriptive knowledge. Now let's take a bit more of a fine-grained look.

Moral vs Descriptive: Legitimacy ratings Asking participants about moral cases rather than descriptive cases predicts higher levels of perceived disagreement, deception, and agreement that there is a fact of the matter about the relevant issue.¹⁷ Higher levels of disagreement and deception and lower agreement that there is a fact of the matter each predict lower ratings of the knowledge of forming a belief via testimony (controlling for direct effect of X). There is partial mediation in two of the models with the indirect path via disagreement accounting for 36% of the total effect, via deception accounting for 35%. Note that, in the case of judgments about metaphysics, although the confidence interval for the indirect path contains zero there are signs of possible inconsistent mediation such that asking questions about moral cases rather than descriptive has the direct effect of reducing legitimacy ratings, but it also raises metaphysics ratings which themselves increase legitimacy ratings.

Moral vs Descriptive: Knowledge ratings While asking participants about moral cases rather than descriptive cases does predict higher levels of perceived disagreement, deception and metaphysics (these are the same paths as in the previous paragraph), the effect of domain on knowledge doesn't seem to be mediated by any of these.

Aesthetic vs Descriptive: Legitimacy ratings Asking participants about aesthetic cases rather than descriptive cases predicts higher levels of perceived disagreement and deception, and lower agreement that there is a fact of the matter about the relevant issue. Higher levels of disagreement and deception and lower agreement that there is a fact of the matter

¹⁴In interpreting the results of these analyses, there are three main points of interest: (a) The path weight for the indirect effect, i.e., the total effect of domain on ratings (in the testimony case) minus the effect of domain on ratings controlling for the mediator. The larger the indirect effect the more domain has a mediated effect on ratings. (b) The bootstrapped 95% confidence interval for the indirect effect. When this confidence interval doesn't overlap zero the indirect effect is judged to be statistically significant. (c) The ratio of the indirect effect to the total effect. This is one way to think about effect size in mediation – What proportion of the total effect is accounted for by the effect via the mediator?

¹⁷The fact that moral matters invite higher agreement that there is a fact of the matter is perhaps surprising. However, note, it is likely an artifact of the particular items used: the morality of hitting people just because you feel like it (which may invite perceptions of objectivity which are unrepresentative of morality in general) vs the largeness of the Empire State building (which may invite perceptions of subjectivity which are unrepresentative of descriptive matters in general). If that interpretation is correct, then some caution should be exercised in interpreting mediation results involving the metaphysics item.

each predict lower ratings of the legitimacy of forming a belief via testimony (controlling for direct effect of X). There is partial mediation in all three models with the indirect path via disagreement accounting for 39% of the total effect, metaphysics for 35%, and deception 21%. Although, of course, note that these indirect paths shouldn't be interpreted as additive (it is not the case that together they account for 95%).

Aesthetic vs Descriptive: Knowledge ratings While asking participants about aesthetic cases rather than descriptive cases does predict higher levels of perceived disagreement and deception, and lower metaphysics (these are the same paths as in the previous paragraph), the effect of domain on knowledge doesn't seem to be mediated by either disagreement or deception. Lower agreement that there is a fact of the matter does appear to partially mediate the effect. Lower metaphysics predict lower ratings of the legitimacy of forming a belief via testimony (controlling for direct effect of X). The indirect path via metaphysics accounts for 17% of the total effect.

3.9 Discussion

The second set of predictions concerned the influence of the three further factors: beliefs about the background level of deception, beliefs about the background level of disagreement, and metaphysical beliefs. The prediction was that each of these three attitudes will explain some of the asymmetries predicted to be observed between legitimacy and knowledge ratings for beliefs formed via testimony. These predictions were also borne out at least in part. It was found that perceptions of background levels of disagreement and deception partially explain both asymmetries concerning judgments about legitimacy. In the case of the asymmetry between aesthetic and descriptive matters, metaphysical beliefs were found to partially explain the asymmetry both for judgments about legitimacy and knowledge.

In relation to the influence of these three factors, the findings of the current study don't cohere quite so well with the previous findings of [Andow \(2018\)](#) concerning the contrast between aesthetic and descriptive domains. That previous study found no role for judgments about metaphysics and found a role for disagreement and deception only in the case of knowledge judgments. The current study finds a role for judgments about metaphysics (with respect to the contrast between aesthetic and descriptive matters) and a role for disagreement and deception only in the case of legitimacy judgments. It is important to note that there is no great conflict here (remembering that the null results in either study should be interpreted cautiously). Together, they can be taken as suggesting that a decent portion of the asymmetry in the way participants treat aesthetic and moral testimony relative to descriptive testimony can be explained with reference to perceptions of the background level of disagreement and deception, and, in the case of the aesthetic-descriptive asymmetry (and perhaps only there), metaphysical beliefs. There is, nonetheless, further work to be done looking at how the influence of these three factors patterns across the various types of epistemic judgment which are relevant to the philosophical debates.

4 Implications

What are the implications of the current findings for debates about moral testimony (and about aesthetic testimony)? There are, of course, some important limitations to the current work. The range of descriptive, moral, and aesthetic beliefs featured in the materials is limited and the specific claims featured may not be fully representative of the wider domains. This means that the extent to which the current findings should be assumed to generalize to the wider phenomenon is limited. However, the current findings nonetheless represent an important first step in getting empirical traction on a empirical picture which is at the heart of an important philosophical issue. Debate about the virtues and vices of accepting moral claims on the basis of testimony, or on deferring to moral experts, has largely left the psychological claim that there is an asymmetry in the ways we treat moral and descriptive testimony unexamined.

The current study begins a project of empirically examining this intuitive foundation to the debate about moral testimony. In particular, the current findings raise three points: (i) moral knowledge does seem to be perceived to be particularly difficult to obtain via testimony and this may be in part because it is also thought to be less legitimate or somehow improper to form moral beliefs this way; (ii) this represents an asymmetry with descriptive matters where it is perceived to be relatively easy to obtain descriptive knowledge via testimony (but some question about how legitimate it is); (iii) there is some sign that this difference between the perceived legitimacy of forming moral beliefs via testimony as opposed to descriptive beliefs is attributable to the perceived background levels of disagreement and deception being higher with respect to moral matters than to descriptive.

This first point poses a *prima facie* challenge to those in the debate who think that ‘the problem’ with moral deference is that it is not a good way to achieve moral understanding, or some other threshold which is in the problematic cases not achieved by mere knowledge (e.g., [Hills, 2009](#)). It also poses a challenge to those who claim that the intuitive inappropriateness of moral deference is moral alone and not epistemic ([Blanchard, 2018](#)). Participants seemed reluctant to accept one could acquire moral knowledge via testimony alone. It is possible that these responses are driven by a sense that knowledge is not enough, or participants may not really distinguish moral knowledge and understanding. It is also possible that a study with a more diverse battery of questions might be able to find reliable differences between occasions when participants are willing to accept that moral deference can lead to moral knowledge and those when they are not.

The second and third point together point towards a particular kind of explanation. Participants are more skeptical of moral deference when they perceive there to be higher levels of disagreement and deception about such matters. This supports the idea that at least some of the asymmetry is due to something other than a (perceived) in principle difference between domains. The current results certainly don’t support the idea that a main reason for the asymmetry is (that the folk believe in) moral anti-realism (although some caution should

be exercised in interpreting what is effectively a null result).¹⁸ However, it may well be that our intuitive understanding of moral testimony is shaped by numerous factors. And, while there is no evidence here that participants' understanding of morality is that it is special and somehow commands working things out for oneself rather than deferring to others, such evidence may yet be forthcoming from other studies. Indeed, in §2, I canvassed a large number of potential explanations which the current results don't speak to directly and future work should use these to generate further hypotheses to be tested. Nonetheless, the findings in this paper suggest that a decent proportion of the asymmetry can be put down to perceived higher levels of disagreement and deception. One important gap in the current research concerns epistemic evaluations beyond legitimacy (or whether a belief is properly formed) and knowledge. I noted that there could be an epistemic asymmetry in terms of other concepts such as justification and understanding. Further work could also examine judgments about these.

The findings in this paper also help to flesh out and develop our understanding of the asymmetry between aesthetic and descriptive testimony, and suggest that the two asymmetries take a similar form (with respect to the influence of domain and source). The one hint of an important difference concerned what accounts for that asymmetry: a difference in the role of judgments about whether there was a fact of the matter. At least in this study, moral matters actually invited a stronger sense that there was a fact of the matter than descriptive matters. Although this is perhaps due to the precise materials used, it does betray a potentially important difference as it did not prevent them from responding in strongly asymmetric ways to moral and descriptive issues. On the other hand, with respect to the asymmetry between aesthetic and descriptive matters, the extent to which participants thought there was a fact of the matter did seem to account for some of the asymmetry. This may provide some reason to be cautious about assuming that moral-descriptive asymmetries and moral-aesthetic asymmetries are analogous phenomena.

References

- Andow, J. (2018). Aesthetic testimony and experimental philosophy, in F. Cova and S. Réhault (eds.), *Advances in Experimental Philosophy of Aesthetics*, 1 ed., Bloomsbury Academic, chp. 2.
- Blanchard, J. (2018). Moral realism and reliance on moral testimony, *Philosophical Studies*, Online first.
- Callahan, L. F. (2018). Moral testimony: A re-conceived understanding explanation, *Philosophical Quarterly* 68(272).
- Cohen, J. and Pain, N. (2010). Colours, colour relationalism and the deliverances of introspection, *Analysis* 70(2):218–228.

¹⁸ Although, of course, an appeal to an actual systematic difference in levels of disagreement is used by many to argue in favour of moral anti-realism of some form (see [Loeb, 1998](#), for example)

- Cova, F. and Pain, N. (2012). Can folk aesthetics ground aesthetic realism?, *Monist* 95.
- Dular, N. (2017). Moral testimony under oppression, *Journal of Social Philosophy* 48(2):212–236.
- Enoch, D. (2014). A defense of moral deference, *Journal of Philosophy* 111(5): 229–258.
- Fletcher, G. (2016). Moral testimony: Once more with feeling, in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics (Vol. 11)*, Oxford University Press.
- Groll, D. and Decker, J. (2014). Moral testimony: One of these things is just like the others, *Analytic Philosophy* 55(1).
- Hills, A. (2009). Moral testimony and moral epistemology, *Ethics* 120(1): 94–127.
- Hills, A. (2013). Moral testimony, *Philosophy Compass* 8(6): 552–559.
- Hopkins, R. (2007). What is wrong with moral testimony?, *Philosophy and Phenomenological Research* 74(3): 611–634.
- Howell, Robert, J. (2014). Google morals, virtue, and the asymmetry of deference, *Noûs* 48(3): 389–415.
- Jones, K. (1999). Second-hand moral knowledge, *The Journal of Philosophy* 96(2): 55–78.
- Loeb, D. (1998). Moral Realism and the Argument from Disagreement, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 90(3): 281–303.
- MacKinnon, D. P., Fairchild, A. J. and Fritz, M. S. (2007). Mediation analysis, *Annual Review of Psychology* 58(1): 593–614.
- Mcgrath, S. (2011). Skepticism about moral expertise as a puzzle for moral realism, *The Journal of Philosophy* 108(3): 111–137.
- McShane, P. J. (2018). The non-remedial value of dependence on moral testimony, *Philosophical Studies* 175(3): 629–647.
- Meskin, A., Liao, S. and Andow, J. (in preparation). Aesthetic testimony: Some empirical evidence.
- Mogensen, A. L. (2017). Moral testimony pessimism and the uncertain value of authenticity, *Philosophy and Phenomenological Research* 95(2).
- Roberts, P., Andow, J. and Schmidtke, K. (2014). Colour relationalism and the real deliverances of introspection, *Erkenntnis* 79(5): 1173–1189.
- Robson, J. (2012). Aesthetic testimony, *Philosophy Compass* 7(1): 1–10.
- Sliwa, P. (2012). In defense of moral testimony, *Philosophical Studies* 158(2).