

# MORAL REALISM AND FAULTLESS DISAGREEMENT

*Daniel Cohnitz*

## *Abstract*

Is moral realism compatible with the existence of moral disagreements? Since moral realism requires that if two persons are in disagreement over some moral question (that has a determinate answer) at least one must be objectively mistaken, it seems difficult to uphold that there can be moral disagreements without fault. Alison Hills argued that moral realism can accommodate such disagreements. Her strategy is to argue that moral reasoners can be faultless in making an objectively false moral judgement if they followed the relevant epistemic norm, i.e. *follow your conscience*, when making their judgement.

I will argue that Hills' strategy does not work. The putative epistemic norm *follow your conscience* does not trump moral truth, because believing something wrong for the wrong reasons is worse than believing something right for the wrong reasons.<sup>1</sup>

## 1. Introduction

Is moral realism compatible with the existence of moral disagreements? Since moral realism requires that if two persons are in disagreement over some moral question (that has a determinate answer) at least one must be objectively mistaken, it seems difficult to uphold that there can be moral disagreements without fault. Recently, Alison Hills argued that moral realism can accommodate such disagreements.<sup>2</sup> Her strategy is to argue that moral reasoners can be faultless in making an objectively false moral judgement if they followed the relevant epistemic norm, i.e. *follow your conscience*, when making their judgement.

I will argue that Hills' strategy does not work. The putative epistemic norm *follow your conscience* does not trump moral truth, because believing something wrong for the wrong reasons is worse

<sup>1</sup> I would like to thank Alexander Davies, Toomas Lott, Francesco Orsi, the reviewers for *Ratio*, and the audience at my presentation of an earlier version of this paper at ECAP9 in Bucharest for helpful comments. The research for this paper was supported by the grants IUT20-5 and ETF9083 of the Estonian Research Council.

<sup>2</sup> Alison Hills: 'Faultless Moral Disagreement', *Ratio* XXVI (2013), pp. 410–427.

than believing something right for the wrong reasons. Thus, Hills' strategy is not a way for the moral realist to accommodate the phenomenon of faultless moral disagreement.

In section 2 I will reconstruct Hills' argument. In section 3 I will explain why her argument requires a certain notion of 'faultless disagreement', which explains why we consider certain areas of discourse (aesthetics, matters of taste, ethics) to be special. What is required for the relevant notion of cognitive faultlessness is that you objectively could not have done better when you made your judgement. As I will argue in section 4, the norm *follow your conscience* does not guarantee that you could not do better *as a moral reasoner*. Thus, if you followed your conscience when judging that *p*, that does not mean that you therefore can faultlessly believe that *p* when *p* is false. In the final section, I will discuss whether the realist could amend the epistemic norm that Hills suggested in order to accommodate faultless disagreements. I will argue that only a change in strategy seems promising. Either the realist should deny that there are cases of real faultless disagreement, or she should accommodate it by assuming moral indeterminacy.

## 2. Hills' Argument

Hills begins with the observation that just as there seem to be cases of faultless disagreement in aesthetics, there similarly seem to be cases of faultless disagreement in ethics. Cases in which two parties to a dispute genuinely disagree, although neither participant is at fault, and neither participant might improve in giving up her belief.

Hills argues that the moral realist can accommodate faultless disagreements in the following sense:

While it must be true of [two disagreeing parties in a moral dispute] that at least one of them has a false belief, it may also be true that neither of them has made a mistake or is at fault in the way that [she] formed [her] belief, for [she] perfectly followed the appropriate epistemic norm: follow your conscience.<sup>3</sup>

Thus, the moral realist accepts that some mistake has been made, the belief of at least one of the disputants is *false*. But since the

<sup>3</sup> Hills, 'Faultless Moral Disagreement', p. 418.

norm *follow your conscience* is the relevant epistemic norm, and since both disputants followed it, they are both epistemically blameless.

What does it mean to follow your conscience? Hills offers the following gloss:

All that I mean by that in this context is: you use your own judgement to form your moral belief. I do not intend to give a full account, indeed any specific account, of what is involved in that. But I take it that it could include at least the following ways of coming to a moral judgement. You have a strong intuition that some action or practice is wrong (or that it is permissible). You weigh up the reasons for and against a particular action and decide that it is wrong. You assess arguments for or against an action and judge some to be much stronger than others. You have a strong emotional reaction to an action (disgust, admiration . . .) and make a moral judgement on that basis. [. . .] *The important thing [. . .] is that [. . .] it is you who makes the moral assessment.*<sup>4</sup>

Hills argues that we should follow the epistemic norm *follow your conscience* because this is the only way to gain *moral understanding*. You have moral understanding, if and only if you believe for some moral fact *p*, that *p* because of *q*, and *p* and *q* and the explanatory relationship all obtain.<sup>5</sup> Moral understanding might not give you moral knowledge (because Hills allows that your belief is not safe, or not reliably formed, or not merely accidentally true), but it is the most important epistemic state, because only moral understanding allows you to get to moral truth through your 'own sensitivity to moral reasons' (in a way in which believing that *p* on the basis of testimony does not).

### 3. Faultless Disagreement

If we follow the terminology introduced by Max Kölbel<sup>6</sup> then Hills' strategy seems to be that of the *mitigated realist*. Strictly speaking, faultless disagreements are impossible; at least one of

<sup>4</sup> Hills, 'Faultless Moral Disagreement', p. 417, my emphasis.

<sup>5</sup> Hills, 'Faultless Moral Disagreement', p. 424.

<sup>6</sup> Max Kölbel: 'Faultless Disagreement', *Proceedings of the Aristotelian Society* 104 (2003), pp. 53–73.

the disagreeing parties must be mistaken. But unlike the *unmitigated realist* who would explain the appearance of faultless disagreement away as a case in which we merely cannot tell who made the cognitive mistake that led to the false belief, the mitigated realist tries to identify areas of discourse in which disagreeing parties can both be, in fact, *cognitively* faultless:

This means that both have followed the correct methods for arriving at a view on the matter, methods that are in accordance with the concepts involved in the view in question. Ignorance of relevant evidence or faulty reasoning is not an issue.<sup>7</sup>

It is worth emphasizing that this strategy should not count just any case as cognitively faultless in which the cognitive agent possesses defeasible justification for her beliefs because she responsibly formed her beliefs in accordance with the evidence available to her and some relevant epistemic norm or other. Otherwise all areas of discourse would allow for faultless disagreements, and there was no *prima facie* problem for the *moral* realist to accommodate this phenomenon. In other words, it does not suffice for being blameless that you subjectively could not have done better in arriving at your judgement, but that it is *impossible* to do better.

Contrast the case of Ann and Betty with that of Carry and Darcy:

TASTE: Ann and Betty both had Italian gorgonzola with honey for desert. Ann makes the judgement; 'This gorgonzola with honey is delicious', Betty makes the judgement; 'Gorgonzola with honey is disgusting'. Both sampled the desert under normal conditions, and would reliably come to the same judgement in similar circumstances.

COLOUR: Carry and Darcy are asked to sort objects by colour. Carry sorts them into six categories, Darcy sorts them into five, because she has a certain rare form of colour blindness, which makes it impossible for her to tell apart two of the colours involved. When asked how many different colour categories the objects fall into, Carry says; 'There are six categories', Darcy says; 'There are five categories'. Both looked at the objects under ideal lighting conditions and would reliably come to the same judgements under similar conditions.

<sup>7</sup> Kölbel, 'Faultless Disagreement', pp. 59–60.

In the relevant sense of 'faultless', we would say that Ann and Betty are having a (potentially) faultless disagreement. On the other hand, Carry and Darcy do not have a faultless disagreement.

Thus, when we observe that both parties to a dispute about matters of taste can be faultless – as in the case of Ann and Betty –, we do not just mean that they both carefully looked at the available evidence when making their judgement, as Carry and Darcy did. Rather, what we mean is that they both looked at all the evidence that's objectively relevant and formed their judgements as it was objectively appropriate.

Hills seems to agree with this, when she argues that moral realism can accommodate faultless disagreements in all the four ways in which the notion is defined by Kölbel:<sup>8</sup>

**DEFINITION 1:** A disagreement is faultless if A believes that p and B believes that not-p, and neither has made a mistake (is at fault).

**DEFINITION 2:** A disagreement is faultless if though A and B disagree, each has exactly the view that they ought to have, and giving up the belief in question would be an improvement for neither of them.

**DEFINITION 3:** A disagreement is not faultless if at least one disputant is in error and further discussion or investigation is called for. It is faultless if this is not the case: 'there is no disputing matters of taste'.

**DEFINITION 4:** A disagreement is faultless if each disputant is blameless.

Hills argues that moral realism can accommodate faultless disagreements between parties A and B so defined, because if A and B followed their conscience, then in virtue of them thereby following *the* epistemic norm, they *objectively* did not make a mistake, are blameless and cannot possibly improve by changing their belief (because that would mean to violate the epistemic norm). Moral realism might also accommodate faultless disagreements as defined in 3, if the following scenario occurs:

For instance, if you follow your conscience and form a moral judgement, but then discover that many people disagree with

<sup>8</sup> Kölbel, 'Faultless Disagreement', cited here from Hills 'Faultless Moral Disagreement', p. 418.

you (who have also followed their consciences as well as they can) it may be appropriate to reflect again on the issue and check your initial judgement, taking into account any arguments or reasons they can adduce for their moral views. But if your conscience really does return the same judgement, perhaps at that stage it is legitimate not to reflect any further.<sup>9</sup>

In this case, you have already done all you can do in order to correctly form your moral judgement. This is why the moral case is special on Hills' account (and why her view is that of a mitigated realist). In other areas of discourse, the higher-order evidence provided by the disagreement of our peers might matter (as Hills allows, p. 422), such that dismissing it would be a cognitive mistake, but in the moral case we might be entitled to dismiss the higher-order evidence if the disagreeing view of our peers has no influence on the deliverances of our conscience.<sup>10</sup>

#### 4. Why you should not always follow your conscience

As already explained above, Hills argues that believing a moral claim *p* when one has followed one's conscience (which under normal circumstances leads to true belief) cannot be trumped or improved by any other reasoning, because following one's conscience in forming a moral belief is necessary for acquiring moral *understanding*, and the latter is the ultimate epistemic goal of moral reasoning. This is the reason why following your conscience is *the* epistemic norm when it comes to making moral judgements.

But even if we grant that moral understanding is the ultimate epistemic goal, and that following your conscience is the only way towards that goal, it does not mean that reasoning that does not follow that method cannot trump reasoning that follows the method. After all, when you are the party in the dispute that has the mistaken belief, then you believe a false claim *for the wrong reasons*. You believe that *p* (the moral claim) and that *p* because *q* (where you take *q* to be the relevant moral reason for *p*). Now, since *p* is false, either *q* will be false too, or not the relevant moral reason, or perhaps both, false *and* irrelevant. Changing your belief

<sup>9</sup> Hills, 'Faultless Moral Disagreement', n. 6.

<sup>10</sup> To be clear, when following your conscience, you can be influenced by others (their arguments, but also their judgements). You only need to make your own moral assessment in the final judgement. See Hills, 'Faultless Moral Disagreement', n. 5.

to a true moral belief via some other reason (e.g. on the basis of testimony) might not give you moral understanding, so you might only end up believing the right thing (the moral truth) for the wrong reasons (morally irrelevant reasons). But the latter seems still better than the result you acquired when only following your conscience. In what follows, I will try to show how a situation can arise in which not following your conscience is the epistemically and morally right thing to do.

It seems clear, and – as we saw above – conceded by Hills, that sometimes it is *rational* not to trust ones' own evaluation of all the arguments and all the evidence, even if one has the capacities to evaluate arguments and evidence in normal circumstances reliably. As we saw already, a case that Hills considers is that of peer disagreement. It might be, she concedes, that if the ultimate cognitive aim is knowledge or truth, it is rational to change your belief merely on the basis of peer disagreement.<sup>11</sup>

Other examples from the non-moral domain seem to be the kinds of cases that Bishop and Trout discuss in their *Epistemology and the Psychology of Human Judgement*.<sup>12</sup> Statistical prediction rules, 'SPRs' for short, have often been shown to outperform experts in making certain predictions even if the experts had more evidence to base their judgement on, and even when experts were also provided with the result of the SPR. In these cases, biases and other factors that the experts were unaware of made their own assessments of the evidence less reliable than that of the SPR. For example, an SPR that predicts progressive brain dysfunction on the basis of cues from intellectual tests, diagnoses 83% of cases correctly, while clinical experts predicting from the same data only got 63% right, and only 75% right if they also had the results of the prediction rule available.<sup>13</sup> If you know that, it is irrational and *irresponsible* not to follow the prediction of the SPR but rather to choose to trust your own judgement instead.

However, you are an *expert*, and your judgement is to some extent reliable. Following your own assessment of the case, based on your experience and your theoretical knowledge, is the only way to gain *understanding* of why your prediction holds up. Just following the SPR does not deliver that understanding. That the SPR is more successful in predicting progressive brain dysfunction

<sup>11</sup> Hills, 'Faultless Moral Disagreement', pp. 422–423.

<sup>12</sup> Michael Bishop and J.D. Trout: *Epistemology and the Psychology of Human Judgement*, OUP 2005.

<sup>13</sup> Bishop and Trout, *Epistemology*, p. 13.

is, as far as you know, just a brute fact. Thus, also here, just getting the prediction right on the basis of following an SPR is less valuable than getting the prediction right *and* understanding why the prediction is right. But here it would be irrational and irresponsible to stick to your own assessment when you know the SPR does better. Is the moral case really different? Can you dismiss evidence that your moral judgement is false, without being blameworthy?

Let's take an example suggested by Hills. Francis believes that eating meat is permissible; Oscar believes that it is not. The moral realist should say that one of them is mistaken. Let's assume that Francis is mistaken. Now, the belief he has might of course not be without consequences. If many people are like Francis (as there are) and if they all believe something that is morally wrong, chances are that a lot of morally wrong things are done, because they are believed to be permissible when they are not.

Already at this point one might wonder why the personal achievement of Francis, viz. to arrive at moral understanding, should outweigh the potential moral costs it has to trust in your conscience when you have rational reason not to. Let's assume Francis has such reasons. Oscar has been right in the past and Francis knows that, as well as that – although both make reliable moral judgements – Oscar's track record is better. Let us stipulate that it would be rational for Francis to switch to Oscar's belief, if 'only' truth was at stake. Unfortunately, though, Francis does not find Oscar's arguments and reasons morally compelling. He appreciates that Oscar is probably right, but is unable to see *why* he is right, and he is not moved by the fact that he disagrees with Oscar. Even if one takes the idea seriously that Francis' arriving at moral understanding might outweigh moral knowledge, the problem is that if Francis ends up believing something false, there is just no personal achievement of moral understanding to be had either. You cannot get moral understanding for a false moral judgement. Thus, if Francis has rational reasons to change his belief, i.e. if he has reason to believe his moral judgement is false, he has *no* reason to follow his conscience. He should adopt Oscar's judgement, if he has reason to believe it is right, even if he does not understand why it is right. Therefore, in the case considered by Hills, if you learn that most people disagree with you on a moral judgement, and you have reason to consider them to be your epistemic peers, you cannot just stick to your guns, if it is in the non-moral case irrational to stick to your guns. Consequently,

'follow your conscience' cannot be *the* epistemic norm for moral reasoning, and following it will not make you blameless.

Note that Hills cannot just accommodate such cases by factoring the peer disagreement into the moral deliberation. It is true that an agent, in the situation I consider, has epistemic as well as moral reasons to follow the moral testimony of her peers. But the point is not that these reasons should get added to the rest of the moral deliberation of the agent, such that the decision to follow the judgement of your peers is a result of 'following your conscience' after all. If the situation is sufficiently similar to the SPR case, the agent has reasons to believe that her moral deliberation is inferior to the judgement suggested by the SPR, *even if* she takes the judgement of the SPR into account. Paulina Sliwa describes similar cases when arguing in defence of moral testimony.<sup>14</sup> Her cases involve situations in which the moral agent has reasons to believe that her own judgement is compromised by bias or self-interest, and that others do not suffer from that bias in the relevant situation. Thus, in such cases the agent has reasons not to just add her peers' judgements to her moral deliberation, but simply to adopt her peers' judgements:

[I]t seems that in cases like these, further deliberation after receiving the testimony may be exactly what the agent should not do. After all, if the agent is worried about her moral judgement being biased, there is no guarantee that her further deliberation won't be biased or overly impressed with her self-interested reasons as well. And once she starts reflecting further, she opens the way to succumbing to temptation and rationalizing the testimony away.<sup>15</sup>

Thus there are cases where the epistemically as well as morally right thing to do is *not* to follow your conscience. Hence *follow your conscience* cannot be the epistemic norm to follow in making moral judgements, and hence it cannot be that, because you followed the norm, you end up being epistemically faultless when arriving at a false moral judgement.

<sup>14</sup> Paulina Sliwa: 'In defense of moral testimony', *Philosophical Studies* 158 (2012), pp. 175–195.

<sup>15</sup> Sliwa, 'In defense', p. 183.

## 5. How should the realist accommodate faultless disagreement?

One might object that I have been unfair in my construction of the case. The idea should rather be that there are situations in which Francis does not know that Oscar is more likely to be right, and that it is in such situations that you should follow your conscience, and that if you do, you have exactly the belief you should have. The latter might be right. If, say,  $p$  and not- $p$  are equally likely to be true, for all you know, then the fact that your conscience recommends  $p$ , and that following your conscience is the only way to moral understanding, will then make it rational to believe  $p$  rather than not- $p$  (because true belief + understanding is a more desirable state than mere true belief). But notice that it is only because you have no evidence whatsoever to believe that  $p$  is false. This construal of faultless disagreements is simply that of unmitigated realism; such an epistemic situation can occur in any area of inquiry. This view does not explain, as Kölbel already pointed out, why we seem to react differently to disagreements about moral matters than to disagreements about how many people are in the room.<sup>16</sup> Cases in which we are subjectively blameless cannot explain the difference, because they occur everywhere.<sup>17</sup>

However, as our discussion shows, unmitigated realism, i.e. the view that there are disagreements in some areas of discourse without cognitive shortcoming, even though not both parties in the disagreement can be right, is difficult to defend. Precisely, because it is difficult to see how someone could have been absolutely flawless in her thinking and fully informed in an objective way, but arrive at a false belief.<sup>18</sup> Hills' strategy requires that there

<sup>16</sup> Kölbel, 'Faultless Disagreement'.

<sup>17</sup> Karl Shafer also seems to recommend the strategy of the unmitigated realist. His 'epistemic' accommodation of faultless disagreement is based on the idea that a cognitive agent can be said to be blameless if she followed the right epistemic norm, although she might get things wrong because of ignorance or error. See Karl Shafer: 'Faultless Disagreement and Aesthetic Realism', *Philosophy and Phenomenological Research* LXXXII (2011), pp. 265–286. This position is much weaker than Hills' (although Hills refers to Shafer's paper as a source of inspiration for her own account, cf. Hills 'Faultless Moral Disagreement', n. 2). For example, Shafer does not consider the relevant epistemic norm to trump other ways of getting to know the truth (for example, unlike Hills, he defends the role of testimony of experts for gaining aesthetic knowledge, while on Hills' account basing your moral beliefs on testimony is a violation of the appropriate epistemic norm). This way the realist can accommodate faultlessness in 'some sense', but not in all the senses that Hills aspires to accommodate. For a thorough critique of unmitigated realism, see Kölbel 'Faultless Disagreement' and *Truth without Objectivity*, Routledge 2002.

<sup>18</sup> Kölbel, 'Faultless Disagreement', p. 60.

is an epistemic norm that does not guarantee true belief but is the rule to follow even in case it leads to false belief. But on a realist conception it is not clear what benefit such a norm would bring. Hills suggests that the benefit is that of gaining moral understanding, but as we have seen, moral understanding requires truth and thus cannot justify the norm when it leads one away from truth. There does not seem to be any other epistemic benefit that could trump truth on a moral realist conception. Therefore, I doubt that Hills' strategy could be amended to make the account more plausible.

Now, of course, moral realism could accommodate faultless disagreements if the realist would allow that there is some moral indeterminacy.<sup>19</sup> Perhaps everyone should give about 10% of their income to charity, but whether you should give exactly 10%, or rather 9.99999% or 10.000001% etc. is not settled by the moral facts. Thus there can be faultless disagreement in the sense that neither party is wrong. But once you allow that at least one of them has a false belief, you seem to have to say that at least one of them has also made a *cognitive* mistake, or the realist will not be able to accommodate faultless disagreements.

*Department of Philosophy  
University of Tartu  
Ülikooli 18  
50090 Tartu, Estonia  
cohnitz@ut.ee*

<sup>19</sup> Hills, 'Faultless Moral Disagreement', p. 416.