

# EXPLANATION AND JUSTIFICATION IN MORAL EPISTEMOLOGY

Walter Sinnott-Armstrong

Recent exchanges among Harman, Thomson, and their critics about moral explanations have done much to clarify this two-decades-old debate. I discuss some points in these exchanges along with five different kinds of moral explanations that have been proposed. I conclude that moral explanations cannot provide evidence within an unlimited contrast class that includes moral nihilism, but some moral explanations can still provide evidence within limited contrast classes where all competitors accept the necessary presuppositions. This points towards a limited version of moral skepticism.

Gilbert Harman's discussion of moral explanations at the beginning of his textbook, *The Nature of Morality*,<sup>1</sup> is rightly famous for raising a profound "problem with ethics." In his central example, Harman argues that a scientist's belief that a proton is in a cloud chamber at a time can be justified by an inference to the best explanation of the scientist's observations of a cloud in that chamber at that time. This justification works, because the actual presence of a proton and hence the truth of the belief is a necessary part of the best complete explanation of why the scientist forms the belief at that time rather than earlier or later. More generally, protons are needed to explain experimental results and thereby to explain why scientists come to believe in protons on the basis of those experiments.

The problem for ethics is that moral facts and properties never seem to be needed for explanations of this kind.<sup>2</sup> Harman tries to solve this problem by arguing that moral facts can be reduced to psychological facts that are needed for the best explanation of observations.

However, without such a reduction of morality to science, morality seems more problematic than science.

At first, it was not completely clear which kind of problem was being raised. On one interpretation, with some basis, Harman's argument raises a *metaphysical* or *ontological* problem about whether or not moral facts exist. However, certain background beliefs and interests affect which explanation is best, but these beliefs and interests do not affect which facts exist. This makes Harman's argument dubious if it is directly about moral ontology.

Harman's argument becomes more plausible if it is interpreted so that it raises a problem in moral *epistemology*. In recent exchanges, Harman makes it clear that epistemology is his main concern: "[M]y interest in moral explanations has always been an interest in the extent to which uncontroversial data might be better explained by one rather than another competing moral framework, so that the data could provide evidential support for the one framework over the other."<sup>3</sup> Harman's conclusion is that "[w]hether or not [Thomson's examples] are cases in which a moral claim explains some uncontroversial facts, they are not cases that support one moral framework as compared with another."<sup>4</sup> Thus, Harman now seems to allow the possibility of moral explanations, but he still denies that they have any epistemological force.

I will argue that Harman is correct to some extent but not completely. Moral explanations can be used to adjudicate between some rival moral frameworks if, but only if, these frameworks share certain assumptions.

To show this, the first step must be to clarify the basic principle in Harman's argument. It seems to be this:

An observation *O* is evidence for hypothesis  $H_1$  above hypothesis  $H_2$  if the truth of hypothesis  $H_1$  is a necessary part of the best complete explanation of observation *O*, and if neither observation *O* nor any presupposition of the explanation is in dispute between defenders of hypotheses  $H_1$  and  $H_2$ .

The truth of  $H_1$  need not be the complete explanation, since subsidiary principles will be needed; but the truth of  $H_1$  must be *necessary* and not just an optional add on.<sup>5</sup> Also, this kind of evidence requires not just any explanation but the *best* explanation, where the best explanation is, roughly, the simplest and most conservative that explains the most data. Most importantly, the observation to be explained must be *independent*, that is, not controversial or in dispute at present. If one theology provides the best explanation of why God told Abraham to kill his son, that is not evidence for that theology against an atheist who denies that God told Abraham anything.

This independence requirement makes such evidence relative to a contrast class. An observation that provides evidence for hypothesis  $H_1$  above hypothesis  $H_2$  might not provide evidence for hypothesis  $H_1$  above hypothesis  $H_3$  if the observation is admitted by proponents of  $H_1$  and  $H_2$  but is denied by proponents of  $H_3$ . We cannot, then, say simply that the observation is evidence for  $H_1$ . We need to say that it is evidence for  $H_1$  as opposed to a certain contrast class of competitors. This should come as no surprise. Explanations are relative to contrast classes, so this relativity is inherited by evidence based on inferences to the best explanation.

Given this relativity, one issue is whether these requirements are met by any moral explanation out of any *limited* contrast class of moral beliefs or theories. If so, the explained observations provide evidence for one moral belief or theory above others in the contrast class. If not, they do not.

A separate issue is whether any moral explanation can provide evidence for any moral belief when the contrast class is *unlimited*, so it includes *all* competitors. To address the latter issue, I will focus on one extreme competitor, *moral nihilism*, which is the view that there are no moral facts, including facts about which acts are morally required or wrong, good or bad, just or unjust, etc.

I will argue that no moral explanation can provide evidence against moral nihilism, since every such explanation depends on some moral presupposition that nihilists deny, so it fails the independence requirement. Consequently, moral explanations cannot provide evidence within an unlimited contrast class. Nonetheless, some moral explanations can still provide evidence within limited contrast classes where all of the competitors accept the necessary presuppositions. There will be other limited contrast classes within which the moral explanation cannot provide evidence, but my point here is just that moral explanations work for *some* limited contrast classes but *not* for an unlimited contrast class.

Many moral explanations have been proposed. Nobody claims that moral facts directly explain purely physical events, but moral facts are claimed to explain various aspects of human life, including acts and beliefs of individuals or groups of humans.<sup>6</sup> So there are four main patterns:

| WHAT IS EXPLAINED? | Individual    | Group        |
|--------------------|---------------|--------------|
| Act                | Woodworth     | Abolitionism |
| Belief             | Amazing Grace | Convergence  |

The first pattern occurs when a moral judgment about a person or a person's character trait is cited to explain that person's acts. A classic example, given by Nicholas Sturgeon, is that of Passed Midshipman Selim Woodworth, as described by Bernard DeVoto. Woodworth accepted command of the mission to rescue the Donner party, who were trapped and dying in the High Sierra mountains. Instead of leading rescue parties into the mountains, Woodworth "spent time arranging comforts for himself in camp, preening himself on the importance of his position."<sup>7</sup> As a result, many people died who could have been saved. DeVoto concludes, "Woodworth was just no damned good."<sup>8</sup> Sturgeon concurs. Their idea seems to be that Woodworth would have led a rescue if he had had any goodness at all in him. More specifically, his moral vices of cowardice and vainglory are supposed to explain his actions.

To assess this explanation, we need to specify the relevant contrast classes. If the question is why Woodworth did bad acts instead of good acts, then it seems natural to cite the moral badness of Woodworth's character. But this question obviously assumes that Woodworth's acts were bad. So it cannot provide evidence against anyone who denies that assumption, as do moral nihilists. To get evidence that does not beg the question against moral nihilism, we need the contrast class to be specified in morally neutral terms. Presumably, then, what is to be explained is why Woodworth stayed in the base camp instead of going into the mountains to save the Donner party.

Of this, there are several possible explanations: Maybe (1) he was coerced (by being held at gun point in the base camp). Maybe (2) he did not know that the Donner party needed him (because he thought someone else had already rescued them). Maybe (3) he thought he had to stay in the base camp to prevent some greater evil (such as the spread of an infectious disease carried by the Donner party). But none of this was true. Woodworth had no justification or excuse for his acts. So why did Woodworth do what he did? In contrast with explanations (1)–(3), it seems better to explain his actions by the fact that (4) Woodworth was no damned good.

Opponents of moral explanations will insist on considering other alternatives, such as (5) Woodworth was more concerned about his own comfort and ambition than about the lives of those whom he had promised to rescue. If some non-moral explanation like (5) is available, then (4) does not seem necessary.

However, defenders of this moral explanation will respond that (5) is a reason why (4) is true on the assumption that anyone is no damned good if he is more concerned about his own comfort and ambition than about the lives of others whom he promised to rescue. So (5) is not a

competitor to (4). Whereas (1)–(3) exclude (4) (and also (5)) as the best explanation, (5) does not undermine but supports (4), at least on certain assumptions.

Defenders of moral explanations might go further and claim that (4) is better than (5), because (4) is more general, so (4) would explain more actions than (5) would explain. Moreover, suppose Woodworth would have stayed in the base camp if he had been no damned good in some way other than that described by (5), such as if he had been more concerned about his physical appearance than about the Donner party. Then (4) rather than (5) would capture the level of generality that best explains Woodworth's overall pattern of actions.

Even if so, however, this same level of generality can be achieved by describing the non-moral supervenience base for this moral property. Let's stipulate that "badness" is the complex non-moral base property that underlies the supervening moral property of being bad (or no damned good). Then we can cite Woodworth's badness to explain everything that could be explained by his badness (or by his being no damned good). One might object that there might be nothing that unifies badness.<sup>9</sup> But there also might be nothing that unifies badness (or being no damned good). So there is no reason to prefer the moral explanation over the subvening non-moral explanation. Admittedly, there is also no reason to prefer the non-moral explanation. Moreover, if supervenience ruled out all moral explanations and evidence, it would also rule out all biological and economic explanations and evidence; insofar as biological and economic facts also supervene on physical facts, just as moral facts do. So this argument is too strong to be plausible.<sup>10</sup>

This makes it seem legitimate to cite the moral fact that Woodworth was no damned good to explain why he stayed in the base camp instead of saving the Donner party, but only if we assume that his kind of character is bad. If this assumption is shared by all of the moral views in a limited contrast class, then DeVoto's moral explanation can be cited as evidence out of that contrast class for the moral fact that Woodworth was no damned good.

Moreover, contrary to Harman's claim, this moral explanation can "support one moral framework as compared with another."<sup>11</sup> Consider a utilitarian moral theory on which one is morally good if one's character has good consequences. Compare this with a religious moral theory on which one is morally good if one is connected to God by faith. The religious view still might hold that God would never let anyone with faith be more concerned about his own comfort and ambition than about the lives of others whom he promised to rescue. But there might be strong independent evidence that Woodworth had faith in God. That

together with the evidence that Woodworth was no damned good would support the utilitarian moral framework over the religious framework, as long as both continue to share the moral assumption that a person is no damned good if he is more concerned about his own comfort and ambition than about the lives of others whom he promised to rescue.

In contrast, this assumption would be denied by moral nihilists. Since moral nihilists deny all moral facts, they deny (4). Instead, they would cite something like (5) (or his badness) as the best explanation of Woodworth's acts. And (5) does compete with (4) on their view, since they believe (4) cannot be true but (5) can, so (5) is not a reason for (4). Consequently, the independence requirement for evidence is not met by this moral explanation when moral nihilism is in the contrast class, so this purported moral explanation cannot provide any evidence against moral nihilism.<sup>12</sup>

Similar points apply to the second pattern, where a moral fact is supposed to explain some group action or event. In the standard example, we observe that the anti-slavery abolitionist movement grew in the United States during the 1830's. We wonder why. One possible explanation is that (1) The abolitionists were largely a displaced elite expressing resentment at industrialists who had replaced them.<sup>13</sup> Another possible explanation is that (2) The abolitionists reacted to the real evils in slavery (which became worse during the 1820s).

The latter explanation seems better than the former if enough abolitionists were industrialists. Opponents of moral explanations will insist on considering other alternatives, such as (3) Abolitionists disapproved of the pains and losses of freedom in slavery (which became worse during the 1820s). But explanations (2) and (3) are not competitors if pain and loss of freedom are the evils in slavery. This assumption would be denied by a nihilist. So the rise of abolitionism cannot be evidence against moral nihilism. But it can still be evidence within some limited contrast classes that share common moral assumptions.

Things get trickier with the third pattern, where an individual's moral belief is explained by its truth.<sup>14</sup> If there are moral facts, then it is possible that someone called Amazing Grace is a perfectly reliable detector of all moral facts. Grace believes that an act is morally wrong when and only when the act is morally wrong, and similarly for right, bad, good, and so on. The real moral wrongness of an act would then be the best explanation of Grace's belief that the act is morally wrong.<sup>15</sup>

Such an explanation of a moral belief is criticized by Judith Jarvis Thomson on the grounds that Grace believes that an act is wrong because she observes non-moral features of the act, which explain but are *not* explained by the moral fact.<sup>16</sup> This is said to distinguish moral perception

from visual perception, where one's belief that an apple is red is based on the fact that it appears red, which *is* explained by the fact that the apple is red. However, symmetry can be reintroduced simply by saying that Grace believes that an act is wrong because it seems wrong to her, and that seeming *is* explained by the moral fact that it is wrong. So Thomson's criticism does not cut very deep.

A deeper (although more obvious) problem is that this example is fictional. In the actual world, one could cite Grace's belief as evidence for a moral fact only if one already had adequate independent evidence that Grace is a reliable detector of moral facts. Such evidence would have to include observations that Grace's past moral beliefs were true.

This works fine in some cases. I might have known Grace for a long time and believe that she has always gotten moral matters right in the past. Then, if you and I are having a moral dispute, you can cite Grace's moral belief as some evidence against my moral view. Our shared moral assumptions about the truth of Grace's past moral beliefs can make evidence out of our observations of Grace's current moral beliefs. The best explanation of why she believes what she does might be that her beliefs are true once again.

However, this cannot work against a moral nihilist for the simple reason that a moral nihilist would deny both that Grace's positive moral beliefs were ever true and that she is now a reliable detector of moral facts. Consequently, even if Grace is reliable and moral facts are essential to the best explanation of Grace's moral beliefs, Grace's beliefs still cannot be evidence against a moral nihilist because of the independence requirement for evidence.

The fourth pattern can be exemplified by the observations that almost everyone agrees that torturing babies just for fun is morally wrong, and modern societies are converging on the belief that slavery is morally wrong. The claim is *not* that so many people cannot possibly be mistaken. Instead, the argument is that the best explanation of why so many people agree is that they are reacting to a real moral wrong.<sup>17</sup>

The most common response is that such agreement or convergence can be explained at least as well by sociobiology or some other social science without any moral facts. The fact that such moral beliefs help society survive is enough to explain why we have evolved so that so many people hold those beliefs.

Moral explanationists might respond that this sociobiological explanation does not compete with the moral explanation, because the belief that such acts are morally wrong helps society by dissuading people from doing such acts. Preventing these acts helps society because such acts usually cause harm to society or to its members. And that is just what

makes those acts morally wrong, or so it is assumed. If this substantive moral assumption is accepted by both parties in a dispute, then observations of moral agreement and convergence might be best explained by moral facts, and this can be cited as evidence for one of the disputing moral frameworks above another.

The story is quite different with moral nihilism. As before, moral nihilists will insist that the explanation in terms of moral facts does compete with the sociobiological explanation, and nihilists will reject the moral assumption that acts are morally wrong when they usually reduce the survival value of society. Consequently, observations of moral agreement and convergence cannot be used as evidence against nihilism.

These four patterns all cite moral facts to explain non-moral facts. Thomson recently admitted that explanations in this direction never provide evidence for moral beliefs.<sup>18</sup> However, she went on to claim that moral evidence can be based on explanations in the other direction, that is, when non-moral facts explain moral truths. In her example, "Alice's giving Bert a banana was her keeping her word when [she knew] it cost her a lot to do so and she could have got away with not doing so. That, it seems plausible to think, would explain the truth of 'Alice's giving Bert a banana was just'."<sup>19</sup>

Harman responded, "Clearly, in a context of a dispute as to whether actions like Alice's constitute justice, [the claim that the cited features of Alice's act explain why it is just] is controversial and cannot show that there is evidence in favor of one moral framework rather than another."<sup>20</sup> In a different context, however, this kind of explanation might be evidence in favor of one moral framework over another. Suppose Carol and David agree that Alice's act was just, and the dispute is about whose moral framework can best explain why it is just. Carol is a utilitarian, so David argues forcefully that Alice's act does not maximize utility, since it will cost Alice a lot to give the banana to Bert, and nobody will be hurt if Alice does not give it, since Bert doesn't need the banana anyway. If Carol admits that utilitarians cannot explain why Alice's act is just, but she still believes that Alice's act is just, then David can offer a better explanation of its justice, namely, Thomson's. This explanation of a moral fact by non-moral facts can then provide evidence in favor of David's deontological framework above Carol's utilitarian framework.

Of course, none of this works unless Carol believes that Alice's act is just, so it cannot work against a utilitarian who sticks to her theory or against a moral nihilist. But that is the point. Moral explanations like this can work in some contexts against some opponents even if they cannot work in all contexts against all opponents.

The final question is: Does this refute moral skepticism? That depends on what moral skepticism claims. An *extreme* version of moral skepticism claims that there cannot ever be any evidence in favor of any moral belief or theory as opposed to *any* other. If I am right, moral explanations refute that extreme version of moral skepticism.

In contrast, a more *limited* version of moral skepticism claims only that there cannot ever be any evidence in favor of any moral belief or theory as opposed to *all* others. One other moral theory is moral nihilism. I argued that moral explanations can never provide evidence against moral nihilism, so they can never provide any evidence in favor of any moral belief or theory as opposed to moral nihilism. If nothing else can provide such evidence either, as I have argued elsewhere,<sup>21</sup> then limited moral skepticism is true.

Thus, the "no-explanation argument for moral skepticism" contains part of the truth, but only part.<sup>22</sup> And defenders of moral explanations have part of the truth, but only part. The only way to capture the whole truth is to distinguish different kinds of justification relative to different contrast classes and then to endorse limited moral skepticism.

Walter Sinnott-Armstrong, Department of Philosophy, Dartmouth College, Hanover, NH 03755-3592; Walter.Sinnott-Armstrong@dartmouth.edu

## NOTES

1. Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), chap. 1.
2. In these debates, the term 'fact' is and will here be used in a minimalist sense in which it is a fact that *p* if and only if *p*. Moral beliefs can and will here be called 'true' in a similarly minimalist sense.
3. Gilbert Harman, "Responses to Critics," *Philosophy and Phenomenological Research* 58.1 (March 1998): 207f. Compare Harman, *op. cit.*, 3; Gilbert Harman, "Moral Explanations of Natural Facts—Can Moral Claims be Tested Against Moral Reality?" *The Southern Journal of Philosophy* 24, supplement (1986): 60; and Gilbert Harman and Judith Jarvis Thomson, *Moral Relativism and Moral Objectivity* (Oxford: Blackwell, 1996), 78, 168ff.
4. Harman and Thomson, *op. cit.*, 170.
5. In other words, the moral fact must be a necessary enabler of the explanation. I argue that all reasons are necessary enablers in "An Argument for Consequentialism," *Philosophical Perspectives* 6 (1992): 399–421.

6. One could also distinguish attitudes from beliefs, as does Thomson in Harman and Thomson, *op. cit.*, 87ff, but that distinction would not affect my main points.
7. Nicholas Sturgeon, "Moral Explanations" in *Morality, Reason, and Truth*, ed. D. Copp and D. Zimmerman (Totowa, NJ: Rowman and Allanheld, 1984), 63.
8. Bernard DeVoto, *The Year of Decision: 1846* (Boston: Houghton Mifflin, 1942), 442.
9. This objection might be suggested by Jonathan Dancy's claims about "natural shapelessness" in *Moral Reasons* (Oxford: Blackwell, 1993), 76. But see my "Some Varieties of Particularism" (forthcoming in *Metaphilosophy*).
10. See Harman and Thomson, *op. cit.*, 80, and Peter Railton, "Moral Explanation and Moral Objectivity" *Philosophy and Phenomenological Research* 58.1 (March 1998): 178-180.
11. Harman and Thomson, *op. cit.*, 170.
12. Another common example of this pattern cites Hitler's moral depravity as the best explanation of why he ordered the killings of millions of innocent people. However, this explanation assumes that Hitler's acts were evil (Harman and Thomson, *op. cit.*, 86). This assumption would, of course, be accepted by almost everyone, including me, but it would be rejected by moral nihilists. So this example provides no more evidence against moral nihilism than do Woodworth's acts.
13. This explanation is discussed by Nicholas Sturgeon, "Thomson Against Moral Explanations," 205, and Judith Jarvis Thomson, "Reply to Critics," 221, both in *Philosophy and Phenomenological Research* 58.1 (March 1998).
14. One example of the third pattern is someone who does not believe that slavery is morally wrong but converts to the belief that slavery is morally wrong as a result of reading an historical novel or watching a movie like *Roots*. By hypothesis, the conversion cannot be explained by any of this person's former moral beliefs, so the best explanation is supposed to be that his new belief results from the real evil in slavery. See Richard Werner, "Moral Realism," *Ethics* 93 (1983): 653-679. As before, this explanation would never be accepted by anyone who denies that slavery is evil, so it cannot be evidence against nihilism. The trickier example in the text comes from William Tolhurst, "Supervenience, Externalism, and Moral Knowledge," *The Southern Journal of Philosophy* 24, Supplement (1986), 43-55, esp. 46.
15. Indeed, if the supervenience base of moral wrongness has no shape, we might not be able to formulate any non-moral explanation of Grace's belief. See note 9.
16. Harman and Thomson, *op. cit.*, 84.
17. This argument does not deduce 'ought' from 'is', because it is inductive. Still, the premises are non-moral (despite 'best' explanation being evaluative),

- and the conclusion is moral (if it is not just about what we have evidence to believe), so this argument still might seem like pulling a rabbit out of a hat. Anyway, even if this argument works, it is not much like evidence in science.
18. Harman and Thomson, *op. cit.*, 91.
  19. *Ibid.*, 93.
  20. *Ibid.*, 170f.
  21. See my "Moral Skepticism and Justification" in *Moral Knowledge?*, ed. Walter Sinnott-Armstrong and Mark Timmons (New York: Oxford University Press, 1996), 3-48; and "Nihilism and Skepticism about Moral Obligations," *Utilitas* 7 (1995): 217-236.
  22. Harman and Thomson, *op. cit.*, 75.