Moral Explanations of Moral Beliefs

DON LOEB
The University of Vermont

Gilbert Harman and Judith Thomson have argued that moral facts cannot explain our moral beliefs, claiming that such facts could not play a causal role in the formation of those beliefs. This paper shows these arguments to be misguided, for they would require that we abandon any number of intuitively plausible explanations in non-moral contexts as well. But abandoning the causal strand in the argument over moral explanations does not spell immediate victory for the moral realist, since it must still be shown that moral facts *do* figure in our best global explanatory theory.

I. Introduction

Although Gilbert Harman may not have intended to challenge moral realism itself, doubts he raised about whether moral facts figure in our best explanatory theory have widely been taken as the basis for an important objection to this form of realism. Recent contributions to the philosophical literature have done much to reinvigorate this debate.[1] In this paper, I focus mainly on moral explanations of moral beliefs themselves, and argue that, recent claims to the contrary notwithstanding, there is nothing in principle objectionable about them. Still, I claim, this does not amount to a full-scale defense against Harman's challenge. Along the way, I try to untangle some of the threads in this fascinating and much-debated area.

Interestingly enough, Harman's argument may have helped to spur the development of one of the most sophisticated forms of moral realism to emerge in the Twentieth Century, a central project of which is to show that facts about morality *do* figure in the best overall explanatory theories we can come up with, and thus that we are justified in believing in these facts in

---

[1]   Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), Chs. 1 and 2; Gilbert Harman and Judith Jarvis Thomson, *Moral Relativism and Moral Objectivity* (Malden, MA: Blackwell Publishers, 1996); and "Book Symposium: *Moral Relativism and Moral Objectivity,*" *Philosophy and Phenomenological Research,* Vol. LVIII, No. 1, March 1998, pp. 161-222. What is meant by *moral realism* and *moral irrealism* is a matter of some dispute. To simplify here, I'll treat as a moral realist anyone who believes in moral facts (leaving the notion of a fact undefined for now) and as an irrealist anyone who doesn't. If we talk this way, of course, Harman's moral relativism comes out as a form of moral realism, and that might suggest that a more complex account would eventually be needed.

much the same way that we are justified in believing in, for example, facts about protons.[2] Typically the explanations put forward are moral explanations of *non*-moral facts. Moral irrealists, after all, would quite reasonably be unimpressed with moral explanations of purported *moral* facts, since such explanations presuppose the very facts irrealists wish to deny.[3]

In some cases the explanation for a person's moral belief is thought to be found in the state of affairs that makes that belief true. If a putative moral fact could indeed be shown to explain someone's believing *that very fact,* we would have reason to believe in the moral fact in question. And that, in turn, would support a more general claim of moral objectivity.

Recently, however, Judith Thomson has argued that moral facts never do explain our beliefs in them. Thomson's argument comes as a response to a number of putative moral explanations provided by Nicholas Sturgeon and others. In some of these examples, a person believes in a moral fact because of having observed a non-moral fact which appears to support it. Indeed, Sturgeon claims, Harman's central example is of precisely this sort. Famously, Harman asks us to imagine rounding a corner, seeing some boys setting fire to a cat, and observing on that basis that something wrong is being done. Sturgeon sees this as a good example of a moral explanation; one sees the behavior as wrong because it *is* wrong. But Harman argues that it is unreasonable to think that the *wrongness* of this behavior explains one's belief that something wrong is being done, in part because one's belief in the wrongness of causing great suffering to animals (together with one's recognition that someone *was* causing great suffering to an animal) are sufficient to explain one's belief that something wrong is being done.[4] This strand in Harman's argument appears to be an appeal to some sort of *parsimony*. The idea is that we can, and therefore should, explain the observed world without hypothesizing moral facts.[5]

---

[2] This family of views is sometimes referred to (perhaps a bit misleadingly) as "Cornell Realism". See, for example, Richard Boyd, "How to Be a Moral Realist" in Geoffrey Sayre-McCord, ed., *Essays on Moral Realism* (Ithaca, NY: Cornell University Press, 1988), pp. 181-228; David Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989); Richard Miller, "Ways of Moral Learning," *Philosophical Review* 94 (1985), pp. 507-56; Peter Railton, "Moral Realism," *Philosophical Review* 95 (1986), pp. 163-207; and Nicholas Sturgeon, "Moral Explanations" in *Morality, Reason and Truth,* David Copp and David Zimmerman, eds. (Totowa, NJ: Rowman and Allanheld, 1984), pp. 49-78.

[3] Brink briefly considers the possibility of moral explanations of moral facts as evidence for moral realism. *Moral Realism and the Foundations of Ethics,* pp. 182-83.

[4] Harman's focus is on moral *observations,* but as Sturgeon points out, many moral beliefs are not plausibly treated as observations, and nothing of importance to the argument hinges on the restriction to observations. "Moral Explanations," pp. 54-55.

[5] "It would seem that *all we need assume* is that you have certain more or less well articulated moral principles that are reflected in the judgments you make, based on your moral sensibility." *The Nature of Morality,* p. 7 (emphasis added).

There is a second strand in Harman's argument, however, one not often carefully distinguished from the first. Harman worries that moral facts could not possibly have an *impact* on our beliefs: "But there does not seem to be any way in which the actual rightness or wrongness of a given situation can have any effect on your perceptual apparatus.… The explanatory chain from principle to observation seems to be broken in morality."[6] Call this the *causal* strand of his argument. Thomson's argument is largely a development of this causal strand. In what follows, I argue that the causal strand does not succeed in casting any doubt on moral explanations of moral beliefs (nor, by analogy, on any other moral explanations). Even so, abandoning the causal strand does not mean conceding that moral explanations of moral beliefs are good ones. Moral facts, like facts about protons, can be defended only by a showing that they *do* figure in the best overall explanatory theory we can come up with. They must *earn* their place in our understanding of the global economy.

II. Thomson's Epiphenomenality Argument

Sturgeon rejects Harman's critique of moral explanations, claiming that we can evaluate moral explanations by considering counterfactuals involving the denial of the moral claims underlying those explanations. With respect to Harman's cat-burning case, for example, we are asked to consider something like:

> **CF** If the boys hadn't done something wrong, the observer wouldn't have formed the belief that they had done something wrong.[7]

Since the wrongness of the boys' conduct supervenes on its being an act of deliberately causing great suffering to an animal, Sturgeon argues, CF is true. If, for example, the boys had been petting the cat, the observer would not have formed the belief that they were doing something wrong (barring unusual circumstances not troubling to the argument at hand).

Thomson objects to Sturgeon's use of counterfactuals like CF to test putative moral explanations, arguing that moral facts appear *epiphenomenal* in contexts such as this.[8] At the core of her response is a purported counterexample, a moral explanation she thinks passes the counterfactual test, but in which (according to Thomson) the state of affairs denied in the counterfactual's antecedent seems clearly epiphenomenal (and hence not part of the explanation for the state of affairs denied in the consequent). Imagine that

---

6   *The Nature of Morality,* p. 8.
7   *Ibid.,* p. 66.
8   Harman makes a related point in "Moral Explanations of Natural Facts—Can Moral Claims be Tested Against Moral Reality?" *The Southern Journal of Philosophy* Vol. XXIV, Supplement (1986), p. 63.

Donald suddenly shouted "Boo!" in the middle of someone's speech. Surely the rudeness of Donald's behavior supervenes on its being an act of shouting "Boo!" during someone's speech.[9] So if Donald's behavior had not been rude then it wouldn't have been an instance of shouting "Boo!" during someone's speech, and no sound would have appeared on a tape recording of the speech. And it is therefore true that:

> **TCF** If Donald's behavior hadn't been rude, then no sound would have appeared on the tape.

But just as surely, Thomson argues, the fact that Donald's behavior was rude did not *cause* the "Boo!" sound on the tape. The appearance of the "Boo!" sound is explained entirely by the shout itself. Likewise, even if CF is true, the wrongness of the boys' conduct could still be causally (and thus explanatorily) irrelevant to the observer's belief that the boys did something wrong.

Sturgeon's reply is that Thomson's argument proves too much. Even if the supervenience base (the fact that Donald shouted when he did) is sufficient to explain the outcome (the fact that there is a sound on the tape) that would not show that the supervening property is epiphenomenal. For similar things could be said about any number of other explanations in which the supervenience base is sufficient to explain a certain outcome but the supervening property seems explanatory nonetheless. No doubt Steve's anger supervenes on his brain being in some particular microphysical state, a state that would be sufficient to explain his having raised his voice at a faculty meeting. But even Thomson would not want to deny that his anger explains his raised voice. Unless she can provide some plausible criterion for distinguishing these two "apparently parallel cases", Sturgeon claims, her epiphenomenality argument fails.[10]

Thomson's response is that Sturgeon has misinterpreted her argument. She is not claiming that the fact that the supervenience base properties are sufficient to explain the outcome *shows* that the supervening properties cannot be explanatory. After all, sometimes supervening properties *are* explanatory. Despite accepting that the mental supervenes on the physical, for example, Thomson does not dispute that facts about the mental are explanatory and non-epiphenomenal. What she's arguing is that Sturgeon's counterfactual test for non-epiphenomenality (or explanatory relevance) is incorrect, for the test can be passed in cases like Donald's, in which the supervening property

---

9   The supervenience claims employed throughout this paper would undoubtedly have to be refined a bit in order to make them precise. But the rough formulations used here should be sufficient for our purposes.

10   Nicholas Sturgeon, "Thomson Against Moral Explanations," *Philosophy and Phenomenological Research,* Vol. LVIII, No. 1 (March 1998), pp. 204-05. (The example used here is my own.)

seems *paradigmatically* epiphenomenal. *Both* she and Sturgeon need a way of telling epiphenomenal from non-epiphenomenal supervening properties, she thinks.[11] But although she is confident about the epiphenomenality of Donald's rudeness, Thomson believes that there is no clear general test for epiphenomenality. In most cases, she thinks, we must simply examine a proposed explanation and consult our intuitions about whether it is more like the *Donald* case (and hence involves epiphenomenality) or more like cases involving mental states (and hence does not involve epiphenomenality).

Unfortunately, Sturgeon does not even share Thomson's intuitions about the epiphenomenality of Donald's rudeness. "If we were wondering whether rudeness could ever produce a sound on a tape, this example would seem a convincing argument that it could have such an effect, rather than an argument that it could not have it," he says.[12] It is not easy to see how this battle of intuitions could be settled. But it is worth noting that our intuitions can be affected by the way the putative explanation is described. If we ask whether the fact that Donald's shouting 'Boo!' loudly during the speech was rude could explain the sound on the tape, the answer seems to be no. But if instead we ask whether the fact that Donald did something rude could explain the sound on the tape, the answer might seem to be yes.[13]

In any event, Thomson's position has interesting implications for the parsimony strand in Harman's argument. To the extent that we agree with her about the non-epiphenomenality of the mental, we must reject any simple application of the parsimony strand to facts about the moral. If indeed it is true, as Thomson thinks, that the physical base facts on which some mental facts supervene are sufficiently explanatory, then it appears that the mental facts are not needed for explanation. If we nevertheless hold that mental facts are explanatory in such cases, then we cannot consistently reject *moral* facts

---

[11] "What Sturgeon has not noticed is that he too needs an account of the difference between the epiphenomenal and the non-epiphenomenal supervenient. I need an account according to which the moral goes along with Donald on the epiphenomenality side, and the mental goes on the non-epiphenomenality side. Sturgeon needs an account according to which Donald is on the epiphenomenality side, and the moral goes along with the mental on the non-epiphenomenality side." Judith Thomson, "Reply to Critics," *Philosophy and Phenomenological Research,* Vol. LVIII, No. 1 (March 1998), p 219.

[12] "Thomson Against Moral Explanations," p. 204. Thomson's response is unlikely to convince someone like Sturgeon: "What *can* he mean?…[H]e had better mean…no more than that instances of rudeness can produce sounds on a tape, for that is all that is shown by the case of Donald. But [that] is compatible with rudeness being everywhere epiphenomenal. So [Sturgeon's point] escapes me." "Reply to Critics," pp. 219-20. But although instances of rudeness causing sounds on a tape may be *compatible* with rudeness being epiphenomenal, Sturgeon has the intuition that they are *not* epiphenomenal.

[13] Compare the following *non-moral* case: The fact that Larry's hitting Moe over the head with a frying pan was a violent act does not appear to explain why Moe wound up in the hospital. But the fact that Larry did something violent to Moe certainly seems to.

on the ground that we can do without them. That we can do without them, on this view, is an insufficient reason for rejecting their role in explanations.[14]

III. A Test for Epiphenomenality?

Although Thomson does not claim to have a general test for epiphenomenality, in one class of cases she thinks we *can* say definitively that a moral explanation of the non-moral is unsuccessful. To demonstrate this, she brings forward a general principle for contexts such as these. I'll call Thomson's principle:

> **TP** "If a person believes that there is such a fact as F because he takes the fact X to be a reason for believing that there is such a fact as F, then F explains the person's belief only if F explains X."[15]

Thomson's example involves Jones's belief that the apple is red. According to TP, if Jones believes that (F) the apple is red because (X) the apple appears red to Jones, then (F) the apple's *being* red explains Jones's belief only if it explains (X) the apple's appearing red. And indeed it does. We can illustrate as follows (using the arrow as a symbol for "explains"):

The apple's being $\longrightarrow$ The apple's seeming $\longrightarrow$ Jones's belief that the
red   red   apple is red
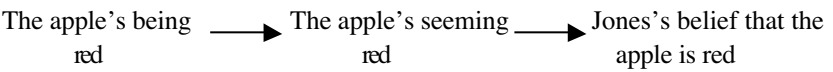
CONNECTED EXPLANATORY CHAIN

Moral explanations of moral beliefs fail this test, however. Suppose, for example, John believes that (F) Mary acted rightly, because he takes the fact that (X) she kept her promise to be reason for believing that (F) Mary acted rightly. Does the fact that Mary acted rightly explain John's belief that she did?[16] Not if TP is correct. Mary's having kept her promise explains John's belief that Mary acted rightly. But the fact that Mary acted rightly does not explain the fact that Mary kept her promise. On the contrary, Mary's having kept her promise explains her having acted rightly. And something like this is true, it would seem, for any case in which someone believes that a moral

---

[14] Parsimony, of course, raises complex issues. Conceivably it could be shown that we *cannot* do without mental facts, but that moral facts are after all dispensable. But at least the *simple* version of the parsimony strand—according to which we should reject putatively supervening properties if the properties on which they are thought to supervene would be sufficient for explanatory purposes—is unavailable to anyone sharing Thomson's (intuitively appealing) view on the non-epiphenomenality of mental facts and the explanatory power of the physical facts on which those mental facts supervene.

[15] *Moral Relativism and Moral Objectivity,* p. 84.

[16] Like Thomson, I take it for granted here that the fact that Mary acted rightly is equivalent to Mary's having acted rightly.

proposition is true on the basis of a non-moral observation. We can illustrate as follows:

| Mary's having acted rightly | **X** → | Mary's having kept her promise | → | John's belief that Mary acted rightly |
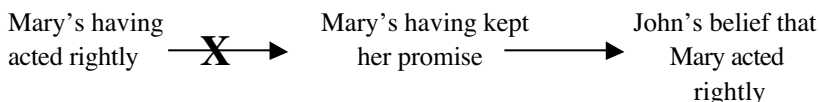
BROKEN EXPLANATORY CHAIN 1

Thomson does not say why she accepts TP, but it seems clear that her reason for accepting it is related to her concerns about epiphenomenality. If Mary's having done the right thing didn't *cause* her to keep her promise, but her having kept her promise caused John to believe that she did the right thing, then it appears that her doing the right thing wasn't part of the causal chain resulting in John's belief that she did the right thing. No doubt a fuller account of the causal chain would include a role for John's *belief* that Mary kept her promise. But a causal chain involving *only* this belief (and Mary's having kept her promise) seems to have no room for the moral fact.

IV. Sturgeon's Reply to Thomson

In reply to Thomson, Sturgeon claims that TP is too narrow. When X is someone's reason for believing the moral proposition that F, he says, we should also treat F as explaining the person's belief that F when X "advert(s) to the very facts constituting the moral one."[17] Thus, we should take the fact that (F) Mary did the right thing as explaining John's belief that she did the right thing, since John's reason for believing that she did is that (X) she kept her promise, and X is the very fact that constitutes her having done the right thing in this case. More generally, Sturgeon appears to accept:

TP′ If a person believes that there is such a fact as F because he takes the fact X to be a reason for believing that there is such a fact as F, then F explains the person's belief only if *either:*

1) F explains X, *or*

2) X *constitutes* F

Why?

Sturgeon asks us to consider the following moral explanatory claim about the origins of abolitionism:

ME Abolitionism was a response to the evils of slavery,

---

17   "Thomson Against Moral Explanations," p. 205.

and two possible non-moral explanations for abolitionism:

> **NMA** Abolitionism resulted from the abolitionists' economic resentment of slaveholders, and

> **NMB** Abolitionism resulted from the abolitionists' recognition that slavery caused immense suffering and their belief that slavery was for that reason wrong.

NMA appears to undermine ME. But NMB does not seem to undermine ME. In fact NMB seems to *support* ME. Yet if TP were correct, Sturgeon argues, we would have to treat *both* NMA and NMB as (if true) equally undermining of ME. And that, he thinks, is implausible.[18]

As it stands, Sturgeon's example does not appear to make contact with TP, since the explananda in ME, NMA, and NMB make no (explicit) reference to anyone's belief. But we can amend the example slightly to avoid this problem:

> **ME'** The abolitionists' belief that slavery should be abolished was a response to the evils of slavery.

And likewise:

> **NMA'** The abolitionists' belief that slavery should be abolished resulted from their belief that slavery was costing them money, and

> **NMB'** The abolitionists' belief that slavery should be abolished resulted from their recognition that slavery caused immense suffering and their belief that slavery was for that reason wrong.

If TP were correct, Sturgeon's argument seems to be, *both* non-moral explanations would (if true) undermine ME equally, since neither the abolitionists' belief that slavery was costing them money nor their belief that slavery caused suffering (and was therefore wrong) would be explained by slavery's actually being wrong. But NMB' does *not* undermine ME'. On the contrary, NMB' *supports* ME', while NMA' undermines it, Sturgeon thinks. Thus TP appears too narrow.

If this is Sturgeon's argument against Thomson, however, it is unpersuasive. Thomson can grant that the two non-moral explanations *look* different, in that one appears to support the moral explanation and the other doesn't. But she can claim that this is *merely* an appearance, as evidenced by the fact

---

18   *Ibid.*, pp. 205-05. Sturgeon develops this point at greater length in "Nonmoral Explanations," in James E. Tomberlin, ed., *Philosophical Perspectives* 6 (Atascadero, CA: Ridgeview Press, 1992), pp. 97-101.

that the moral explanation for the moral belief fails to satisfy TP. The moral facts *don't* cause the moral belief in question, she can argue; the facts they supervene on do. The moral facts themselves, she can continue to maintain, are epiphenomenal. Unless we are given some separate reason for rejecting TP, Sturgeon's assumption that the moral explanation goes through appears question-begging in this context.[19]

V. A More Convincing Reply to Thomson

Is there a way to resolve this dispute over moral explanations of moral beliefs? We might try to assimilate the moral cases to the example Thomson began with, Jones's belief that the apple is red. Thomson illustrates TP by claiming that if Jones believes that (F) the apple is red because (X) the apple appears red to Jones, then (F) the apple's *being* red explains Jones's belief only if it explains (X) the apple's appearing red. Perhaps we can explain moral beliefs in a similar way. For example, I believe that (F) Laura is a good person because (X) Laura *seems* like a good person. My belief that she is good is explained by her seeming to be good, and that in turn is explained by her being good.[20] If so, then moral beliefs can be explained by moral facts in a way that does not violate TP.

There is something fishy about this sort of explanation, however. One wants to know *how* Laura's seeming to be good is explained by her being good.[21] What exactly is the connection between her goodness and her seeming to be good? If it is not possible to say more about the causal mechanisms involved, one might be suspicious of explanations like this. Moreover, examples of this kind do nothing to help with cases like John's belief that Mary did the right thing because she kept her promise. The fact that she kept

---

19   Thomson's response to Sturgeon suggests that she may have something like this reply in mind. According to Thomson, when historians say that (ME) abolitionism was a response to the evils of slavery (or that people believed that slavery was wrong because it *was* wrong) what they *mean* is that people thought that slavery was wrong because they believed that (X) slavery had certain features (such as causing suffering) which *made* it wrong and that (Y) slavery was wrong because it had those features. If that *is* what they mean, Thomson says, then she has no objection. But on that interpretation of the historians' explanatory claim, it *does* distinguish between the two non-moral explanations; it conflicts with NMA but not with NMB. What Thomson won't accept is the *further* claim that people believed that slavery was wrong not just because they believed that slavery had those wrong-making features, but also because slavery *was* wrong. The wrongness of slavery still seems epiphenomenal to her. "Reply to Critics," pp. 221-22.

20   Likewise: "John believes that Mary did the right thing because it seemed as though she did the right thing." For a similar case, see William Tolhurst, "Supervenience, Externalism and Moral Knowledge," *The Southern Journal of Philosophy* Vol. XXIV, Supplement (1986), p. 46.

21   The worry expressed here is related to one Harman expresses in filling out the causal strand of his argument. See "Moral Explanations of Natural Facts—Can Moral Claims be Tested Against Moral Reality?" pp. 62-63. See also Thomson, *Moral Relativism and Moral Objectivity,* p. 83.

her promise is not explained by the fact that she did the right thing. Thus, even if some moral beliefs (such as my belief that Laura is good) *could* be explained by their own truth according to TP, others still violate it.

Instead of looking for moral explanations that satisfy TP, then, perhaps we should challenge TP itself. A counterexample to TP from *outside* of ethics is more likely to be convincing. Suppose that I posses ordinary expertise about marine animals, and I believe that (F) Moby is a mammal. It's no *accident* that my belief that Moby is a mammal is true; I believe it because Moby *is* a mammal and I know this. Thus, the fact that (F) Moby *is* a mammal is part of the best explanation for my *believing* that F. But in ordinary cases, this explanation will fail to satisfy TP. Suppose I take the fact that Moby is a *whale* to be a reason for believing that Moby is a mammal. The fact that (X) Moby is a whale explains my belief that (F) Moby is a mammal. But the fact that Moby is a mammal does not explain the fact that Moby is a whale. Thus TP is violated, as illustrated below:

Moby's being a mammal   **X**➤   Moby's being a whale   ➝   My belief that Moby is a mammal
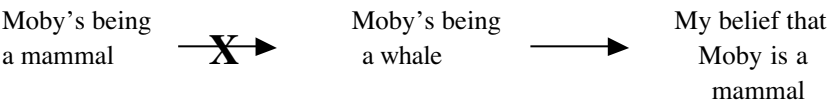
BROKEN EXPLANATORY CHAIN 2

On the contrary, the fact that Moby is a whale explains the fact that Moby is a mammal (in the same way that Mary's having kept her promise explains her having acted rightly).[22] Moreover, this example is easily generalized.[23] That something *is* the case is *usually* a central part of the explanation for our truly believing it to be the case.[24]

Recall our earlier characterization of Thomson's argument for TP. If (F) Mary's having done the right thing didn't *cause* (X) her keeping her promise, but Mary's having kept her promise caused John to believe that Mary did the right thing, then it appears that her doing the right thing wasn't part of the

---

[22] This way of putting things may seem odd because the explanations in question are not *causal* explanations. But not all explanations are causal. I return to this point in Sections VI and VII, below.

[23] Does Moby's being a whale *constitute* his being a mammal (in which case the counterexample would fit under TP')? I don't have clear intuitions about constitution in this case. But the case works as a counterexample to TP regardless of how we understand that difficult notion.

[24] Undoubtedly my belief that whales are mammals is not explained *directly* by the fact that whales are mammals. But suppose we were to fill out the explanation further, pointing to the fact that I learned in school that whales are mammals, that the claim that they are mammals was in a textbook I read, that the textbook was written by a biologist, and so on, down the causal/explanatory line. Eventually we would come to beliefs that are explained by the states of affairs they hold to be true (for example, the beliefs of the biologist who, after studying the physiological structure of whales, concluded that they are mammals).

causal chain resulting in John's belief. The explanatory chain (to use Harman's phrase) appears to be broken. Would Thomson say the same about the *Moby* case? In principle, she could argue that if the fact that Moby is a mammal didn't *cause* Moby's being a whale, but the fact that Moby is a whale caused me to believe that he is a mammal, then the fact that he is a mammal wasn't part of the causal chain resulting in my belief that he is a mammal. If this were right, then Moby's being a mammal would be epiphenomenal to my belief that he is a mammal, and could not explain it.

But while Thomson could in principle take this line she'd still need to explain why I believe that Moby is a mammal. Why did I fasten upon *that* belief, and not one of the innumerable false beliefs available to me (that Moby is a reptile, for example) or on no belief at all? To answer, we should begin by noting that explanations like "I believe that Moby is a mammal because he is a whale," like most explanations we would give in everyday contexts, are somewhat truncated. In fact, I don't believe that Moby is a mammal *solely* because Moby is a whale. I also *believe* that Moby is a whale and that whales are mammals. And a fuller explanation for my belief that Moby is a mammal would involve both of these beliefs. But my belief that whales are mammals *is* explained by the fact that whales are mammals. Likewise, my belief that Moby is a whale is explained by the fact that Moby is a whale. We can illustrate as follows:
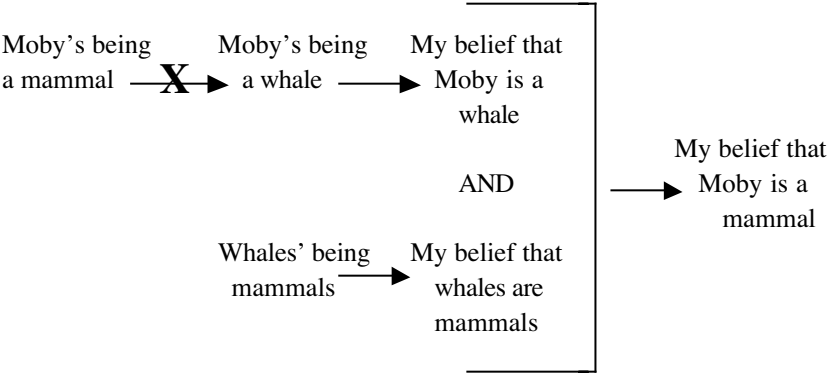
MORE COMPLETE EXPLANATORY CHAIN

Each of my two beliefs, then, is explained by the state of affairs that makes it true. And together, these imply that Moby is a mammal. They also provide an explanation for my belief that Moby is a mammal—an explanation we can reasonably gloss by saying that I believe Moby is a mammal because he is one.

Similar things could be said about John's belief that Mary did the right thing. If we assume that keeping one's promises is morally required, then it seems very likely that John believes that this is so because it *is* so. Combine

that with John's belief that Mary kept her promise (and her *having* kept her promise) and we have an explanation for John's belief that Mary did the right thing—an explanation we can gloss by saying that John believes Mary did the right thing because she *did* do the right thing.

It might be objected that the fact that Moby is a mammal is not part of the explanation for my belief that he is a mammal. It is, it could be claimed, only an *implication* of the facts that explain that belief. And likewise, it could be claimed, the fact that Mary did the right thing is not part of the explanation for John's belief that she did the right thing, but only an implication of the facts that explain his belief. But even if this were correct, it would pose no problem for the moral realist. Suppose, for example that it would be more accurate to say only that John's belief that Mary did the right thing is explained by both his belief that she kept her promise and his belief that keeping promises is the right thing to do, each of which is in turn explained by the fact it represents as true. Still, the best explanation for John's belief that Mary did the right thing would require the hypothesis that there are moral facts, in particular the fact that keeping promises is the right thing to do. And the best explanation for John's belief that Mary did the right thing would give us reason to accept that belief as true, since the claim that she did the right thing is an implication of the facts that explain it.[25]

So Thomson is faced with a dilemma. If my belief that Moby is a mammal can be explained by the fact that Moby *is* a mammal, then John's belief that Mary did the right thing can, at least in principle, be explained by the fact that Mary *did* do the right thing, and moral explanations of non-moral beliefs can, in principle, succeed. To avoid this result, Thomson would have to reject a vast number of intuitively correct explanations like the one offered on behalf of my belief that Moby is a mammal.

VI. Why Abandoning the Two Strands is not Fatal to Harman's Argument

Although our true beliefs can usually be explained by the states of affairs that make them true, not every purported explanation of the sort I've been discussing is a good one. Gary's belief that Moby is a fish is not explained by Moby's being a fish, even if Gary believes that all whales are fish. They aren't. There is always room, in principle, to question whether a given belief is *in fact* explained by the state of affairs it holds to be true, and it's obvious that false beliefs are not explained by their own truth. Moreover, if one is trying to establish the existence of moral facts, one cannot simply assume that they *do* figure in such explanations. That would be to make a mistake similar to the one we noted at the outset, when we put aside attempts to

---

[25] In what follows, I will continue to gloss the more complex explanations so: "John believes that Mary did the right thing because it is true."

prove that moral facts exist by showing how they could figure in the best explanations of *other* moral facts. In the same way, moral facts could not be proven to exist merely by citing their *possible* role in explanations of our beliefs in them.

Readers of "Moral Explanations" might be tempted to think that Sturgeon was being too modest when he said that he did not intend to show that moral properties *do in fact* figure in the best explanation of anything, but only to rebut Harman's claim that they *couldn't* explain (the explanatory chain being broken).[26] We are now in a position to see why this cannot be right. Sturgeon's decision to focus on the causal strand of Harman's argument is responsible for a crucial methodological assumption he makes. For Sturgeon interprets Harman's claim that moral facts *couldn't* figure in the best explanations of the nonmoral to mean that they couldn't *even if they existed*. To see whether they could, Sturgeon says, he needs to *assume* that they exist, and ask whether on that assumption they could do any explaining. But showing that moral facts could explain *if* they existed is not showing that they *do* explain anything, nor that we have reason to believe in them. That, as Sturgeon himself acknowledges, is a much more difficult undertaking.[27]

If Sturgeon's accomplishment is only to show that moral facts *could* explain if they existed, then it is not adequate to defeat the core of Harman's argument. Initially it looked as though the argument had two strands. But both the concern about epiphenomenality and the concern about parsimony arose within the context of a background methodological assumption shared by all of the parties discussed here: that we are not justified in believing in facts of *any* kind unless they figure in our best overall explanatory theory.[28] The two strands we have been discussing were meant to show that moral facts

---

[26] "Now fine discriminations among competing explanations of almost anything are likely to be difficult, controversial, and provisional. Fortunately, however, my discussion of Harman's argument will not require any fine discriminations. This is because Harman's thesis, as we have seen, is *not* that moral explanations lose out by a small margin; nor is it that moral explanations, although sometimes initially promising, always turn out on further examination to be inferior to nonmoral ones. It is, rather, that reference to moral facts always looks, right from the start, to be 'completely irrelevant' to the explanation of any of our observations and beliefs." "Moral Explanations," p. 56. My colleague, Arthur Kuflik, has referred to Sturgeon's strategy as *Argumentum ad Harminem*.

[27] By analogy, suppose someone claimed that schmuarks could not figure in best explanations because nothing so small could figure in best explanations. And suppose that we had a proof that there is no reason in principle to object to explanations involving things of that size (because the hypothesized schmuarks are no smaller than quarks, say). Still, that in itself would give us no reason at all to believe in schmuarks. For that, we'd need a showing that schmuarks *do* figure in the best explanation of something in the observed world.

[28] For a challenge to the claim that moral facts can be justified only by their role in *causal* explanations, see Thomas Nagel, *The View From Nowhere* (New York: Oxford University Press, 1986), p. 144-46. It will become clear shortly why I agree with Nagel, if only in part.

cannot play such a role. But even if these strands are unsuccessful, there is more to the argument over moral explanations. Defenders of moral facts are unlikely to convince moral irrealists until they provide some positive reason for believing that moral properties *do* play this explanatory role.[29] Each sort of entity, the irrealist will claim, needs to fight for its own place in the global economy.[30]

But although the global economy is an explanatory economy, it is not an entirely *causal* explanatory economy. The fact that Moby is a whale explains the fact that he is a mammal, but his being a whale does not *cause* him to be a mammal. Likewise, we believe certain mathematical claims because they are true, but most of us would think it implausible to claim that we interact *causally* with mathematical entities that make them true. Similarly, if there are moral facts, it could be that we believe that they are facts because they *are* facts, even if they do not cause us to believe in them. The pervasive focus on whether moral facts could play a causal role in the generation of our moral beliefs draws our attention away from the possibility of non-causal explanations of this sort.

VII. Moral Facts Explained *by* Non-Moral Facts?

Surprisingly, Thomson herself offers an explanatory claim as a defense of moral objectivity.[31] Although no moral claim explains a non-moral claim, she thinks, we have reason to believe a certain moral claim if the truth of that claim would *be explained by* the existence of a certain non-moral fact. Thus, if Alice's having done something morally right would be explained by her having kept her promise, then we have reason to accept the claim that Alice did something morally right. The explanation would not be a *causal* explanation, but Thomson notes what we have just seen—that not all (good) explanations are causal.

---

[29] Sturgeon seems not to disagree: "[T]he procedure of inferring plausible moral explanations of our nonmoral evidence is not a way of throwing, across an is-ought gap, a bridge that would satisfy a certain sort of determined skeptic—the sort Thomson seems to have set out to answer." "Thomson Against Moral Explanations," p. 203. But some moral realists have appeared to maintain that we are warranted in believing in moral facts even apart from any showing that they figure in our best global explanatory picture, claiming that the realistic nature of moral experience requires us to treat moral realism as our default position. See, for example, David Brink, *Moral Realism and the Foundations of Ethics*, pp. 23-24; Jonathan Dancy, "Two Conceptions of Moral Realism," *Proceedings of the Aristotelian Society*, Vol. LX, Supplement (1986); David McNaughton, *Moral Vision* (Oxford: Basil Blackwell, 1988, Ch. 3); and Nagel, *The View from Nowhere*, Ch. VIII. For serious doubts about this argumentative strategy, see my "The Argument from Moral Experience" (unpublished manuscript).

[30] Just how that fight should proceed is a difficult question—one I don't address here.

[31] The claim of moral objectivity Thomson defends is essentially epistemological, that "It is possible to find out about some moral sentences that they are true." *Moral Relativism and Moral Objectivity*, p. 68. Thus she focuses on moral *skepticism*, rather than moral irrealism, though she recognizes important connections between the two.

What is surprising about this defense of moral objectivity is that it relies on our having a certain background theory and applying it to the case at hand in just the way that the *Moby* case does. My belief that Moby is a mammal is mediated by my belief that whales are mammals—that he is a whale explains his being a mammal. Similarly Thomson's belief that Alice's having done something morally right is explained by her having kept her promise and is itself mediated by Thomson's belief that keeping one's promises is the right thing to do. If we can rely on background theory in the *Moby* and *Alice* cases and others like them, there is no reason in principle why we can't rely on background theory in explaining people's moral beliefs themselves. John's belief that Mary acted rightly is mediated by his belief that she kept her promise *and* his belief that keeping one's promises is the right thing to do.

In all of these cases, at least part of the explanation is non-causal. Moby's being a whale does not cause his being a mammal, though it does *explain* his being a mammal. Mary's having kept her promise does not cause her having acted rightly, though it could explain her having acted rightly. Her having acted rightly doesn't cause John's belief that she acted rightly, but might help to explain it. And, presumably, the fact that keeping promises is right could not cause John to believe that keeping promises is right, though it could explain his belief. If we restrict our attention to things that can play a causal role, we will overlook any number of non-causal, but possibly best, explanations.

Once again, of course, explanations of this sort are only as good as the background theories they employ. So the worry about explanations that presuppose the existence of the very moral facts in question remains. Indeed, as Harman points out, Thomson herself objects to certain types of moral explanations of non-moral facts on something very much like this ground.[32] The objection may be even more potent with respect to the explanatory claim Thomson puts forward in defense of moral objectivity, since a moral fact appears in the *explanandum* and not the *explanans*. The supposition that there *are* any moral facts in need of explanation begs the question against moral irrealism.

For something like the same reason, showing that moral facts *could* explain our beliefs regarding those very facts is insufficient to clinch the moral realist's claim that there are moral facts. In isolation from the non-moral aspects of our best overall theory, moral explanations assume the very thing their defenders wish to establish. A proper defense of their place in the global economy requires a showing that they figure in some more robust way

---

[32] *Moral Relativism and Moral Objectivity,* pp. 86, 170-71.

in the best explanatory story we can tell. That being so, moral realists still have some explaining to do.[33]