THE JOURNAL OF PHILOSOPHY

VOLUME LXXV, NO. 8, AUGUST 1978

THE CONDITIONAL FALLACY IN CONTEMPORARY PHILOSOPHY *

MANY contemporary philosophers have committed a certain fallacy which, surprisingly, has not yet received a clear description. If one is provided we may be better able to avoid this mistake in the future. I shall develop a characterization of the "conditional fallacy" by illustrations drawn from the epistemological writings of Roderick Chisholm, Roderick Firth, Gilbert Harman, Keith Lehrer, Norman Malcolm, and Michael Slote. Moreover, I shall point out that this fallacy also occurs in ethical philosophy, in particular, in certain accounts of prima facie obligations, in John Rawls's definition of 'good', and in some varieties of the ideal-observer theory.

I. ONE VERSION OF THE CONDITIONAL FALLACY

An illustration of the mistake appears in Roderick Firth's account of noninferential warrant.[1] Firth says that some statements have a degree of warrant which is independent of any warrant that they may derive from their coherence with other statements. Firth maintains that a statement sometimes has the property of being noninferentially warranted for me even though I fail to believe the statement. Accordingly, he suggests that this property of a statement consists in the statement's "(1) purporting to characterize (and only to characterize) the content of my present experience, and (2) being a statement that I either now believe to be true or should now believe to be true if I had just decided whether it were true or false" (554).

* I am very grateful to George Boolos, Roderick Firth, and Michael Slote for helpful discussions of this paper.

[1] Cf. "Coherence, Certainty, and Epistemic Priority," this JOURNAL, LXI, 19 (Oct. 15, 1964): 545–547.

0022-362X/78/7508/0397$01.70 © 1978 The Journal of Philosophy, Inc.

By including the subjunctive conditional clause in (2), Firth has overlooked certain counterexamples that he could have detected by considering how the presence of the property he is analyzing depends on whether or not the antecedent of his conditional actually obtains. He might have considered a case similar to that of Mr. Earlybird, who in fact is sound asleep at *t* and for whom the following statement *h* is not at all warranted: "It sounds to me exactly as if I am hearing my wife telling me she was the first one up." We may nonetheless suppose that, on this particular morning, his wife is so surprised by the unusual fact of being the first one up that she can hardly wait to tell her husband—but is, out of consideration, holding her tongue. We may suppose, further, that *if*, contrary to fact, Mr. Earlybird *had* just begun the process mentioned in the antecedent of Firth's conditional, that is, the process of deciding whether or not *h* is true, he would consequently have shown signs of awakening and they would immediately have caused Mrs. Earlybird to announce that she was, indeed, the first one up. Mr. Earlybird's resulting sense experience would, in turn, both have led him to decide to believe *h* at *t* and have been a reason for *h*'s having the property of being noninferentially warranted for him. Thus, even though Mrs. Earlybird is *in fact* not speaking, Firth's conditional concerning the *hypothetical* situation that I have described is true, and it incorrectly leads one to claim that *h* is noninferentially warranted at *t* for the slumbering Mr. Earlybird.[2]

Since Firth's analysis of the property in question commits him to a corresponding analysis of the statement that a proposition has that property, it will prove useful to propose the following initial

[2] We can find counterexamples in waking experience. Women utilize Lamaze techniques during childbirth in order to reduce the severity of pain through concentration on breathing exercises and hand motions. As their thoughts stray back to possible pain, such discomfort is apt to increase. If a certain woman had attempted to decide whether or not some lengthy, complex statement correctly described her pain, rather than concentrating on her exercises, then the pain might, contrary to fact, have been of just such a nature. It is unacceptably paradoxical to maintain that the statement is actually warranted for her in spite of the fact that, awake and aware, she fails to believe that she has such pain, does not in fact have it, and may even lack any good inferential reason for believing that she has it.

Since these counterexamples depend upon what is due to the satisfaction of the antecedent of Firth's conditional, they are compatible with the view that subjunctive conditionals, when counterfactual, do not have to survive "backtracking" claims. See Jonathan Bennett, "Counterfactuals and Possible Worlds," *Canadian Journal of Philosophy*, IV, 2 (December 1974): 381–402, and David Lewis, *Counterfactuals* (Cambridge, Mass.: Harvard, 1973).

characterization of one version of the conditional fallacy:

> A mistake one makes in analyzing or defining a statement *p* by presenting its truth as dependent, in at least some specified situations, upon the truth (falsity) of a subjunctive conditional Ø [3] of the form: 'If state of affairs *a* were to occur, then state of affairs *b* would occur',[4] when
>
> (*Version 1*) one has failed to notice that the truth value of *p* sometimes depends on whether *a* actually occurs and does not depend merely upon the truth value of the analysans or definiens; moreover, one has failed to notice this because one has overlooked the fact that in some of the specified situations (i) conditional Ø is true (false), (ii) the analysans or definiens is true, (iii) state of affairs *a* does not occur, and (iv) if *a* were to occur then the occurrence of *a* or the occurrence of *b* or their combination (the occurrence of *a* or the absence of *b* or their combination) would be at least part of the cause of something that would make *p* true, although *p* is actually false.

II. A SECOND VERSION OF THE CONDITIONAL FALLACY

An illustration of a second version of this mistake appears in the following definition, offered by Keith Lehrer: "But what does it mean to say that reasons give a man knowledge? It means that if he were asked, 'How do you know that?' and he were to give those reasons, his answer would be correct. Those reasons explain how he knows." [5] But suppose that Mr. Silent, the only friend of Mr. Faker, knows that they will remain friends in the immediate future. Yet Mr. Faker pretends to all others that he himself is a misanthrope, and the continuation of the friendship depends on Mr. Silent's keeping the secret. Mr. Nosey, who suspects that the former two are friends, asks Mr. Silent in front of Mr. Faker, "How do you know that you will remain friends with Mr. Faker in the immediate future?" Mr. Silent does know this, but would not if he were to state his actual reasons.

In this example, giving an answer would cause the end of the friendship and make the answer incorrect. A person does not give

[3] The qualification, "in at least some specified situations," allows for cases in which 'Ø' is embedded in another phrase. For example, when *p* is the statement that some given statement has the property of being non-inferentially warranted, Firth's analysans has the form: '*q* and (*r* or Ø)'. We may say that it presents the truth of *p* as dependent upon the truth of Ø in the "specified situations" where *r* is false.

[4] I shall follow the common practice of construing subjunctive conditionals so that they may, but need not, be contrary to fact.

[5] "How Reasons Give Us Knowledge, or the Case of the Gypsy Lawyer," this JOURNAL, LXVIII, 10 (May 20, 1971): 311–313; p. 312.

a correct answer to the question, "How do you know that *p*?" by citing certain reasons when in the situation where he gives those reasons he does *not* know that *p*.[6] Thus, on the present reading, Lehrer's definition illustrates a second version of the conditional fallacy:

> (*Version 2*) one has overlooked the fact that, in some of the specified situations, statement *p* is actually true, but, if *a* were to occur, then it would be at least a partial cause of something that would make *b* fail to occur (make *b* occur).

III. A GENERAL DESCRIPTION OF THE CONDITIONAL FALLACY

The illustrations and descriptions of the fallacy have thus far spoken only of overlooked causal connections. But causal considerations may not always be the pertinent ones, and there are other types of connections which a philosopher may have disregarded. This will require us to broaden the characterization of the fallacy and to explain what the two versions have in common.

In the first edition of *Theory of Knowledge*,[7] Roderick Chisholm distinguishes three attitudes toward propositions: believing, disbelieving, and "withholding," that is, neither believing nor disbelieving. He maintains that

> . . . what is suggested when we say of one of these attitudes that it is more reasonable than another is this: If the person in question were a rational being, if his concerns were purely intellectual, and if he were to choose between the two attitudes, then he would choose the more reasonable in preference to the less reasonable (21/2).

This account may be construed in different ways, depending upon what Chisholm intends by the term 'concerns'. Michael Slote interprets the term as covering the person's desires, and concurs with the resulting definition.[8] However, the definition is then subject to the following counterexample: Consider Professor Chisholm's own epistemic attitudes regarding proposition *h*: 'Chisholm's con-

[6] In his book, *Knowledge* (Oxford: Clarendon, 1974), Lehrer says that "we may say that a justification based on evidence explains how a man knows that *p* if and only if that justification would be a correct answer for the man to the question 'How do you know that *p*?'" (126). The crucial question is, instead, of the form, 'How does *S* know that *p*?' and a correct answer may be able to be *given* only by someone other than *S*. Yet it still may justify *S*'s believing that *p*, even when *S* does not *say* that he believes that *p*. Thus, in his earlier paper, the consequent of Lehrer's conditional should have read, 'his answer would explain his actual knowledge that *p*,' and the whole conditional could have simply been replaced by '*r* explains how *S* knows that *p*.'

[7] Englewood Cliffs, N.J.: Prentice-Hall, 1966.

[8] Cf. *Reason and Scepticism* (London: Allen & Unwin, 1970), pp. 85/6.

cerns are purely intellectual'. In reality, it is false that believing *h* is more reasonable for Chisholm than disbelieving *h*, since he surely knows that his concerns are *not* purely intellectual. But in a hypothetical situation where satisfaction of the antecedent in Chisholm's conditional made it true that his concerns were purely intellectual, he (or at least a subject of the example rather like Chisholm) would similarly know these concerns to be purely intellectual. Examining his concerns would lead him to believe *h* (assuming that he would have no Freudian scruples about the existence of lurking nonintellectual desires) and such a belief would become more reasonable. Chisholm's definition incorrectly treats it as being more reasonable in his actual situation.

It is possible that, by speaking of *S*'s "concerns," Chisholm actually meant instead to refer to *S*'s duties or responsibilities. For, in a later account,[9] Chisholm takes 'epistemic preferability' as the fundamental epistemic term to be explained, and suggests that, if we let *p* and *q* be epistemic attitudes including just believing or withholding propositions, then

> . . . we might paraphrase the locution "*p* is epistemically preferable to *q* for *S* at *t*" in somewhat the following way: If *S* were a purely intellectual being, a being capable only of believing and of withholding belief, and if at *t* he had just the duty of trying his best to bring it about that, for every proposition, *h*, he then believe *h* if and only if *h* is true, then it would be more fitting to the situation in which he finds himself at *t*[10] for him to bring about *p* at *t* than for him to bring about *q* at *t* (225).

Yet a counterexample to this paraphrase, and to the previous definition given in *Theory of Knowledge*, arises when *p* is Chisholm's believing the proposition *h'*: 'Chisholm is a purely intellectual being (in the present sense)' and *q* is his withholding *h'*. In reality, believing is not epistemically preferable for Chisholm, who is aware of the fact that he is not this type of purely intellectual being. But in a hypothetical situation where the antecedent of the conditional obtained, he (or at least a subject of the example rather like Chisholm) would be aware of his different capacities

[9] Cf. "On the Nature of Empirical Evidence," in Roderick Chisholm and Robert J. Swartz, eds., *Empirical Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1973), pp. 224–249.

[10] I read this as meaning: fitting to the hypothetical situation in which he would find himself at *t*. Otherwise, Chisholm's account fails to say anything intelligible about the epistemic attitudes involved in my counterexamples.

and would believe *h'*, which would be more fitting and epistemically preferable.

We must notice, nonetheless, that it is inappropriate to speak of the preceding counterexamples as examples where the occurrence of the state of affairs mentioned in the antecedent of the conditional *causes* Chisholm to believe *h* (or *h'*), or *causes* certain propositional attitudes to be fitting for him. Moreover, we shall later consider examples of the conditional fallacy in ethical philosophy where there are no relevant causal links involved.

Thus, we need to broaden our former description of the conditional fallacy. There are further reasons for doing so. For example, Chisholm's account in his more recent paper is intended not as an analysis but as a paraphrase. In addition, our description of version 1 of the fallacy should allow for cases where the truth value of the statement in question would change from true to false upon the occurrence of *a*. We may, accordingly, characterize the conditional fallacy as follows:

> A mistake one makes in analyzing, defining, or paraphrasing a statement *p* or in giving necessary and sufficient conditions for the truth of that statement, by presenting its truth as dependent, in at least some specified situations, upon the truth (falsity) of a subjunctive conditional, *Ø*, of the form: 'If state of affairs *a* were to occur then state of affairs *b* would occur', when
>
> (*Version 1*) one has failed to notice that the truth value of *p* sometimes depends on whether *a* actually occurs and does not merely depend upon the truth value of the analysans, definiens, paraphrase or list of necessary and sufficient conditions; moreover, one has failed to notice this because one has overlooked the fact that, in some of the specified situations, (i) conditional *Ø* is true (false), (ii) the analysans, definiens, or paraphrase is true or the necessary and sufficient conditions are fulfilled, (iii) state of affairs *a* does not occur, and (iv) if *a* were to occur then the occurrence of *a* or the occurrence of *b* or their combination (the occurrence of *a* or the absence of *b* or their combination) would help make *p* true, although it is actually false [or help make *p* false, although it is actually true and although the analysans, definiens, or paraphrase would remain true or the necessary and sufficient conditions would remain fulfilled] or the states of affairs in question would occur at least partly because *p* would be true [or would be false] or because of what would make *p* true [or would make *p* false] or would together with the truth [or falsity] of *p* form at least part of a reason for some other occurrence; [11] or

[11] In the interests of generality, it should be noted that, even with this expansion, what I have just described is itself but one form that the first

(*Version 2*) one has overlooked the fact that, in some of the specified situations, statement *p* is actually true but if *a* were to occur then the occurrence of *a* would be at least part of what would make *b* absent (make *b* occur) or *a* would occur at least partly because of the absence of *b* (occurrence of *b*) or because of what makes *b* absent (makes *b* occur),[12] or would together with the absence of *b* (occurrence of *b*) form at least part of a reason for some other occurrence.[13]

I believe that it is helpful to specify separately the two versions of this fallacy. But, with the above range of relations in mind, we could include both versions of the conditional fallacy under the following succinct description:

One has overlooked the fact that in some of the specified situations the occurrence of certain relations involving factors that are mentioned in *p* or in the analysans (definiens, paraphrase, or list of necessary and sufficient conditions) is connected either with the occurrence of *a* or with the absence of *a* in such a way as to be responsible for a disparity between the truth value of *p* and the truth value of the analysans (definiens, etc.) in those situations.

I do not think it useful to abbreviate this description even further by dropping all mention of the relations in question and by saying merely that the philosopher has made the mistake of overlooking the fact that there are situations where statement *p* has the opposite truth value to that of the analysans, definiens, etc. For, in the examples I am discussing, the philosopher can be expected to know that to be a fact only because he can be expected

---

version of the conditional fallacy may take. We might call it the *counterfactual form*, since the relevant counterexamples treat conditional Ø as counterfactual. In what might be called the *factual form* of version 1, *a* actually occurs, and, if *a* were absent, then the absence of *a* or the absence of *b* or their combination (the absence of *a* or the occurrence of *b* or their combination) would be at least part of what would make the truth value of *p* differ from its actual value.

12 The phrase, 'or because of what makes *b* absent (makes *b* occur)', should be understood so as to be compatible with the point about back-tracking claims mentioned in fn 2.

13 We can further extend the description of the fallacy by allowing that version 1 can be committed even when a philosopher's account purports to provide only a *sufficient* condition for the truth of statement *p*, and by allowing that version 2 can be committed even when his account purports to provide only a *necessary* condition for the truth of *p*.

A further qualification is needed in this description of the fallacy if the philosopher is attempting to provide only a set of *materially* necessary or sufficient conditions for the truth of *p*. In such cases, the putative counterexample must be not merely an imaginable situation but a situation that actually occurs.

to bear in mind the relations in question. It is appropriate to say that a philosopher has made a mistake or committed a fallacy when that philosopher ought to have known better; otherwise, although his account itself might be said to be mistaken, it manifests an unlucky effort rather than the commission of a fallacy. (Moreover, the mistake just mentioned can be made when giving accounts that do not involve conditionals.)

If we now consider a definition offered by Norman Malcolm[14] we can explain the point of having included clause (ii) in our description of version 1, as well as the point of including a similar qualification within the first set of square brackets:

> Our definition of factual memory can now be stated in full as follows: A person, *B*, remembers that *p* from a time, *t*, if and only if *B* knows that *p*, *and B* knew that *p* at *t*, *and* if *B* had not known at *t* that *p* he would not now know that *p* (236).

Consider a set of circumstances in which the antecedent of the conditional contained in this definition obtains and in which the conditional is true. This will be a situation where the statement that *B* remembers that *p* from *t* is actually false. But such an example does not show that Malcolm has committed a fallacy. For in the situation described, the requirements in the definiens that *B* knew at *t* that *p*, and that *B* knows that *p* are both unsatisfied, and Malcolm naturally wishes the falsity of the definiens to correspond to the falsity of the statement being defined.

IV. OTHER FALLACIES CONCERNING CONDITIONALS

Appropriate objections may nonetheless be made to Malcolm's definition in order to illustrate several fallacies that a philosopher may commit when utilizing conditionals which need to be distinguished from the conditional fallacy. In what I shall call the *fallacy of irrelevant conditionals*, one overlooks the fact that there is a subset of the specified situations in which conditional $\emptyset$ is simply not relevant to a correct account of statement *p*.

Stanley Munsat criticizes Malcolm in a manner which suggests that he regards Malcolm as having committed this fallacy. Munsat points out that I may remember from time *t* that, for example, I killed a deer while driving at *t*, but subsequently come across the same information: at a later time, t_1, someone tells me that I killed a deer while driving at *t*. Thus, "I at least might know it now anyway, because of the remark made to me at t_1."[15] Munsat con-

[14] *Knowledge and Certainty* (Englewood Cliffs, N.J.: Prentice-Hall, 1963).
[15] *The Concept of Memory* (New York: Random House, 1966), p. 23.

cludes that a suitable definition of factual memory simply "does not have a hypothetical as one of the conditions" (33).

A related counterexample will show that Malcolm's definition also commits the conditional fallacy, and thus that it is possible for a philosopher to commit both fallacies at the same time. Let t_1 again be a time between t and the present. Suppose that astronaut Fallible does correctly remember having performed a certain procedure at t. Nonetheless, astronaut Backup was prepared to tell Fallible at t_1 that the procedure had been performed, if Fallible had shown by his behavior that he really did not know whether he had run through it (perhaps because it has become so automatic). In the present example, in contrast to that provided by Munsat, the satisfaction of the antecedent in Malcolm's conditional helps to explain the falsity of the consequent, and shows that Malcolm has committed version 2 of the conditional fallacy.

It is, of course, possible to commit the first version of the conditional fallacy without committing the fallacy of irrelevant conditionals; the definiens may be faulty because of a requirement outside the conditional, and the truth value of an appropriate replacement for that requirement might vary in a suitable fashion with the occurrence or absence of a. Similarly, an account might commit the second version of the conditional fallacy merely because it requires a truth value for the conditional opposite to the value it should require, so that the conditional remains relevant to a correct account.

The previous counterexamples to Malcolm's definition will also allow us to compare the conditional fallacy with what may be called the *ceteris paribus fallacy*, where one makes the mistake of overlooking the fact that the conditional that one's account presents as true (false) is nonetheless true (false) only other things equal. For example, one may have failed to notice that there is a subset of the specified situations where statement p is true and in which one's definiens requires conditional $\emptyset$ to be true, yet in which circumstances happen to arise which prevent b from occurring when a does.

Munsat's discussion shows that, for a person who remembers that p from time t, Malcolm's conditional is true other things equal, provided that we take "other things equal" as ruling out all alternative ways in which the person might have gained the present knowledge that p. Since Malcolm can be expected to be aware of

those alternatives, he commits the *ceteris paribus* fallacy as well as the two I have previously mentioned.

One can commit the second version of the conditional fallacy without committing the *ceteris paribus* fallacy. A type of example in which this happens is when one commits the conditional fallacy together with what may be called the *fallacy of contrary conditionals.* In the latter fallacy, one overlooks the fact that the conditional that one's account presents as true (false) is simply false (true) and cannot even be said to be true (false) other things equal. For instance, some unsophisticated phenomenalist might try to analyze the statement that Dr. Crippen murdered his wife when and where he did in terms of conditionals about the multitude of appearances that would have been manifested to hypothetical viewers at many different spots in the room, overlooking the fact that Crippen would not have committed the crime in front of a witness (other than the victim).[16]

A philosopher can commit the *ceteris paribus* fallacy without committing the conditional fallacy. For example, those factors which would block the occurrence of *b*, were *a* to occur, may themselves have no particular connection with the occurrence or absence of *a*. That is, it may just be a coincidence that they would be present if *a* were to occur. Nonetheless, we might expect the philosopher to know that this type of coincidence can obtain in the present universe and not merely in some logically possible world.[17]

V. THE DIFFICULTY OF AVOIDING THE CONDITIONAL FALLACY

Gilbert Harman points out an example of the second version of the conditional fallacy while discussing what it is for particular reasons to be those for which a person believes something and what it is for particular reasons to be those which give a person knowledge. He states that "a familiar suggestion is that relevant reasons are those a person would offer if asked to justify his belief. This suggestion cannot be correct. Albert may offer good reasons not because he thinks they are any good, but because he thinks they will convince his audience." [18] For instance, Albert may give his advisor not the real reasons for which Albert believes he will fail

[16] I owe this illustration to Roderick Firth.

[17] The "argument from perceptual relativity" is an attempt to show that phenomenalism cannot avoid the *ceteris paribus* fallacy, but not to show that it necessarily commits the second version of the conditional fallacy. [Cf. Firth, "Radical Empiricism and Perceptual Relativity," *Philosophical Review*, LIX, 2, 3 (April, July, 1950): 164–183, 319–331.]

[18] "Knowledge, Reasons, and Causes," this JOURNAL, LXVII, 21 (Nov. 5, 1970): 841–855; p. 843.

his ethics course, but those which he believes the advisor will find convincing.

Harman continues by offering another objection which, in effect, points out an instance of version 1 of the conditional fallacy: "Nor is the analysis to be rescued by requiring that Albert be sincere. Being asked to justify his belief might lead Albert to reassess his reasons. This could lead him for the first time to appreciate his good reasons. . . . Only then could he be said to know that he was going to fail" (843).

We shall see that, in spite of having detected several instances of the conditional fallacy, Harman goes on to offer an account that commits version 2 of the fallacy. This points up the need for an explicit description of the conditional fallacy. Without it, a philosopher may find himself occasionally committing the very same type of mistake he has been quick to spot in others. Thus, it is ironic but not entirely surprising that Harman, too, should fail to avoid the conditional fallacy.

Harman is led into difficulty because he wishes to provide what he calls a "functional" rather than causal account of the connection between believing and the reasons that support one's believing, the reasons upon which it is based. According to Harman, explaining why someone believes something is like explaining why a nondeterministic automaton is in a particular state, where automata are sets of states functionally related to one another and to input and output. To specify an automaton as nondeterministic "is to specify the possible states of the automaton, possible input, and possible output, and what output can follow any given state and input" but not to employ causal language in the description (848, 850).

Accordingly, when discussing what it is for a person to believe something for a set of reasons, Harman wishes to allow that the person may have several sets of reasons, each supporting that belief, but also wishes to avoid construing this as a case of causal overdetermination or of multiple causation. Harman offers, instead, the following suggestion:

> Other things equal, if a person believes a conclusion for certain reasons and becomes doubtful about those reasons, he becomes doubtful about the conclusion. "Other things equal" is meant to rule out the possibility . . . in which one acquires new reasons as one comes to doubt the old. The phrase must also be used to rule out [the analogue of] overdetermination and the analogue of multiple causation. In case there are several sets of reasons for which someone be-

> lieves something, he must become doubtful about all sets before becoming doubtful about his belief. The relevance of any of the sets is this: if he became doubtful of all the other sets, his belief would rest crucially on that set so that, if he should then become doubtful of it, he will become doubtful of his belief, other things equal. Any set of reasons supports someone's belief in a way that a subset of its legs may support a table (847/8).

This suggests that Harman regards one or the other of the following conditionals as constituting an account of the claim that several sets of reasons each independently support a person's believing that *p*:

C_1 For any one of the sets *r*, if (a) the person were doubtful of the reasons in all the sets except *r*, then (b) both (i) other things equal he would believe that *p* and would not doubt that belief, and (ii) if he were doubtful of the reasons in set *r* then other things equal he would be doubtful of his belief that *p*.

C_2 For any one of the sets *r*, if (a) the person were doubtful of the reasons in all the sets except *r* and were still to believe that *p*, then (b′) both (i′) other things equal he would not doubt that belief and (ii) if he were doubtful of the reasons in set *r* then other things equal he would be doubtful of his belief that *p*.

The occurrence of the phrase 'other things equal' appears to be intended as an abbreviation, rather than as a device to avoid a completely "functional" account. But, in specifying when other things are not equal, Harman has overlooked the analogue of another type of causal context. If some of the real causes of a certain result had not been effective, then a factor that was not actually present in the situation might have occurred which would have prevented the remaining actual causes from producing the result. This is not overdetermination or multiple causation.

For example, it may be that, if the legs in one of the sets supporting an old table had begun to fall off, this would have led someone who has been restoring the table to pick it up in order to turn it over for repairs, thereby simultaneously canceling the support of the remaining legs. Yet, in actuality, no legs collapsed and he did not pick up the table.

Analogously, suppose that late last night, while pushing myself to prepare today's lecture in order to substitute for an ill colleague, I forced myself to work out four arguments presenting my reasons for a habitual atheism which I had not previously scrutinized. It may be that, if I were to begin doubting the premises of the first

three arguments as I am now about to present them to the class, my doubts would lead to a heightened sense of fallibility. In conjunction with my recollection of the frame of mind in which I pulled the lecture together and my awareness that philosophy of religion is really not my field, this might make me refrain from continuing to count the premises of the fourth argument as having any *force*, even though I still regarded them as true. Those premises would then not support my atheism. (We may suppose that this would happen even though I would not give up my long-standing atheism on the spot, but would only omit that part of the lecture, while nonetheless beginning to have nagging doubts about whether to remain an atheist.) Yet imagine that I am not actually struck by any doubts about any of the arguments, and so none of these contingencies arise, and I simply continue to rest my atheism on all four arguments while presenting the lecture. In this example, the satisfaction of clause (a) in C_1 and C_2 would make clauses (i) and (i′), respectively, fail to be satisfied. Therefore, Harman's own account of the way in which reasons support beliefs commits version 2 of the conditional fallacy.

In his paper on the gypsy lawyer, Keith Lehrer raises a counterexample against Harman which depends on the power of emotion to overwhelm the influence of r and to make the person fail to believe p once all the additional sets of reasons are removed (cf. 311/2). But this type of objection defeats only interpretation C_1.[19] Moreover, it can be avoided if we alter C_1 by bringing in

[19] In that earlier paper, Lehrer misrepresented clause (i) or clause (i′) as saying of the person that r "would explain his belief that p" (311). In *Knowledge*, Lehrer instead treats such a clause as saying that r "would then sustain the belief" (124). If Lehrer intends 'sustain' to mean support, he thereby renders Harman's definition circular, for it was offered as an account of what it is for reasons to support belief; if Lehrer intends 'sustain' in a causal sense, he overlooks Harman's express desire to avoid a causal account of these relations.

Moreover, Lehrer's remark clashes with his own claim to have proved by his counterexample that if the person doubted the reasons outside set r then "it might even be that he would no longer appreciate the evidence if the belief were to fade" (123). This incorrectly implies that the antecedent of Harman's conditional includes the phrase, "and the person were to begin to doubt his belief that p." Lehrer seems to have confused the passage I have been discussing with one in which Harman speaks of a time at which a set of reasons first comes to support one's belief that p, or at which one reviews the reasoning which led one to the conclusion that p. Harman says that when a conclusion is reached by reasoning then the mental or neurophysiological processes onto which the premises of the reasoning may be mapped themselves lead to belief in p "or would so lead if p were not already believed" (853/854). We may note, however, that Harman's desire to utilize a purely functional approach is once again thwarted. For the above account of reaching a conclusion by reason-

causal considerations (*pace* Harman) and adding to its antecedent the phrase, "and the person's believing that p were not determined by emotional factors (other than any which are claimed to exist in claiming that p)." But even this change fails to defeat my counterexample, since a heightened sense of fallibility can hardly be called an emotional factor.

One might try to evade my objection by adding to the antecedents of C_1 and C_2 either the phrase, "and the person were not to doubt that r is strong enough to support believing that p," or the phrase, "and the person were not to doubt that r provides good reason(s) for believing that p." But the former addition renders the account circular, inasmuch as it was intended as an account of support for one's believing that p. And either addition will reintroduce the difficulty that satisfying the condition might lead the person for the first time to appreciate his good reasons.

In his later book, *Thought*,[20] there is no passage corresponding to the above conditional account of support, and Harman's defense of a functional approach remains to that extent incomplete. Moreover, an instance of the second version of the conditional fallacy appears in *Thought* when Harman gives a list of necessary and sufficient conditions for coming to know something by inference. In keeping with a functional viewpoint, Harman suggests that inference is "a change that can be described simply by mentioning what beliefs are given up and what new beliefs are added" (169). "Our 'premises' are all our antecedent beliefs; our 'conclusion' is our total resulting view" (159). He then suggests as a necessary condition for a person, say, Mary, to come to know something by inference

> . . . not that the actual premises of the inference (everything Mary believes ahead of time) be known to be true but only that the inference remain warranted when the set of antecedent beliefs is limited to those Mary antecedently knows to be true and continues to know after the inference (170).

Harman appears to be requiring that the following conditional be true: If Mary were to have—among all the beliefs she actually has—only those which she antecedently knows and continues to

---

ing specifies what *would* happen and thus goes beyond merely specifying—as we do for nondeterministic automata—what *could* happen. So the account is susceptible to counterexamples similar to those which I brought against C_1 and C_2.

[20] Princeton, N.J.: University Press, 1973.

know after the actual inference and were to make a (substitute) inference to the conclusion(s) of her actual inference, then her (substitute) inference would be warranted.

As a counterexample, suppose that Mary is on a vacation safari and falsely believes *f*: 'Mary sees a dinosaur moving along the edge of a high plateau above her'. After forming this false belief, she comes both to believe the false statement *g*: 'Someone now has a true belief concerning the present whereabouts of dinosaurs now living on land', and simultaneously (or subsequently) comes to know *h*: 'Mary believes *g*'. Harman is committed to speaking of *h* as part of what Mary has "inferred," but his conditional incorrectly prevents us from saying that Mary knows *h*. For satisfaction of the antecedent would remove the false "premise" *f* (or *g*), and we may presume that the resulting substitute inference to "conclusion" *h* is not warranted by Mary's remaining true beliefs.[21]

VI. THE CONDITIONAL FALLACY IN ETHICS

Treatments of prima facie obligations sometimes commit the first version of the conditional fallacy. For example, philosophers sometimes explain the statement that person *S* has a prima facie moral obligation to do action *A* as follows: Doing *A* would be what *S* morally ought to do, all things considered (or would be the morally right thing for *S* to do), if *S* were to have no moral obligations to perform any alternative action. However, Socrates takes himself to have, among his various moral obligations, a moral obligation to teach Alcibiades during the symposium, as well as a moral obligation not to harm him physically during the proceedings, and we wish to speak of these as prima facie obligations (even if they are also part of what Socrates morally ought to do, all things considered). But if the obligation not to harm Alcibiades physically were missing, it would have to be absent for a reason, and this might very well be a reason that would remove the other obligation as well, e.g., Alcibiades' total absence from the occasion or his insanely attempting to assassinate Socrates.[22]

Another example of the conditional fallacy in ethics appears in the course of John Rawls's attempt to find a constant sense for the term 'good'. Rawls defines a person's real good by reference to what

[21] Harman might respond that, although Mary knows *h* upon inferring it, she does not come to know *h* *by* the inference. However, his functional description of inference offers no means of drawing such a distinction.

[22] An alternative account of prima facie obligation provided by W. D. Ross also commits the conditional fallacy. Cf. my "Prima Facie Duty," this JOURNAL, LXII, 11 (May 27, 1965): 279–287.

is for that person the most rational plan of life given reasonably favorable circumstances; he lists as a necessary condition for the most rational plan of life that it would be chosen by the person if that person were to have full deliberative rationality.[23] But in defining full deliberative rationality, Rawls requires "that there are no errors of calculation or reasoning, and that the facts are correctly assessed . . . also that the agent is under no misconceptions as to what he really wants" (417). Satisfaction of the antecedent in Rawls's conditional entails that, in the hypothetical situation, the person would *have* the competence involved in complete deliberative rationality and know that he has it, and would not, for example, be out of touch with his desires in a way that can be overcome only through psychotherapy. Since it is irrational to plan to obtain something when one knows that one already has it, the conditional incorrectly leads us to say that it is not part of a rational plan of life (and thus not part of anyone's real good) to come closer to deliberative rationality by, for example, seeking psychiatric help.

Rawls cannot answer that one's psychological limitation is a relevant fact concerning which one deliberates in the hypothetical situation. For it would no longer be a fact, thanks to the very description of that situation. Thus, the definition at most allows that it is good to *keep* the competence involved in full deliberative rationality when one already has it and incorrectly implies that it is not an intrinsic good to know what one really wants, and in that respect not an intrinsic good to know oneself.[24]

The subjective version of the ideal-observer theory presents definitions of evaluative terms which take forms similar to the following: "X is good" means that, if I were to have characteristics $C_1, \ldots, C_n$, then I would have reaction R_1 to X; and "X is bad" means that, if I were to have characteristics $C_1, \ldots, C_n$, then I would have reaction R_2 to X. Moreover, as a way of attempting to acknowledge that the two value claims are mutually exclusive, such a theory treats R_1 and R_2 as mutually exclusive reactions.

Given the way $C_1, \ldots, C_n$ are usually described in such accounts, it is possible to find occasions on which people would commonly grant that a certain inconvenience or suffering they undergo, s, is good, on balance, because it helps them to come closer to possessing certain of the personal characteristics, $C_1, \ldots, C_n$, for example, impartiality, awareness of the facts, vivid imaginativeness. They

[23] Cf. *A Theory of Justice* (Cambridge, Mass.: Harvard, 1971), pp. 404–409.

[24] For further discussion, see my "Rawls, Brandt, and the Definition of Rational Desires," *Canadian Journal of Philosophy*, VIII, 2 (June 1978).

would say this because they either value those characteristics or at least value coming closer to having them. Moreover, they would say that, if s is not required for those purposes, then it is simply bad. However, if the antecedent in the above conditionals already were to obtain for those people, then they would already have the characteristics in question, s would not be thus required, and that would account for their not having reaction R_1 to s in such hypothetical circumstances. So this type of ideal-observer theory commits the conditional fallacy. Of course, I have assumed that we are considering definitions that are at least plausible to the extent that R_1 and R_2 are attitudes, feelings, or other responses that are, indeed, commonly associated with judging something good or bad. Thus, reaction R_1 would be missing in at least some of the cases in question because the person would no longer see any point to s.[25]

In view of how often instances of the conditional fallacy have occurred in epistemology and ethical philosophy, and because there is no reason to suppose that the fallacy can be committed only in those areas, we may expect that it will continue to appear in the midst of foremost philosophical endeavors. But if we were to ignore the history of philosophy we would be doomed to repeat it—although not in *that* very conditional.

ROBERT K. SHOPE

University of Massachusetts/Boston

[25] Notice that, if the theory insists that R_1 would occur in this example, it is led into a contradiction. For the common value judgment that s is bad if s is not required, combined with the theory's definition of the latter statement, implies that R_2 would occur, yet the theory also maintains that R_1 and R_2 are mutually exclusive reactions.

Defenders of this version of the ideal-observer theory might try to save it in either of two ways: (1) they might attempt to defend the implausible view that phrases such as 'X is good' or 'X is something S morally ought to do, all things considered' have one sense when applied to ends but another when applied to means, and might attempt to provide a definition of the latter sense not subject to my objections; (2) they might speak of the observer as reacting, not to X itself, but merely to contemplation of the possibility of X, e.g., the possibility of suffering s in circumstances where one lacks the relevant characteristics among $C_1, \ldots, C_n$. But this fails to distinguish the statements: (i) 'X is good when in circumstances Y', (ii) 'Circumstances Y are good when accompanied by X', and (iii) 'The combination of X and Y is good'.