

CHAPTER 16

Thought, Selections

Gilbert Harman

Knowledge and Probability

The lottery paradox

Some philosophers argue that we never simply believe anything that we do not take to be certain. Instead we believe it to a greater or lesser degree; we assign it a higher or lower “subjective probability.” If knowledge implies belief, on this view we never know anything that isn’t absolutely certain. That conflicts with ordinary views about knowledge, since our degree of belief in some things we think we know is greater than our degree of belief in other things we think we know.

We might count as believed anything whose “subjective probability” exceeds .99. But that would also conflict with ordinary views. We do not suppose that a man inconsistently believes of every participant in a fair lottery that the participant will lose, even though we suppose that the man assigns a subjective probability greater than .99 to each person’s losing. If ordinary views are to be preserved, belief must be distinguished from high degree of belief.

A rule of inductive inference is sometimes called a “rule of acceptance,” since it tells us what we can accept (i.e., believe), given other beliefs,

degrees of belief, etc. A purely probabilistic rule of acceptance says that we may accept something if and only if its probability is greater than .99. Kyburg points out that such a rule leads to a “lottery paradox” since it authorizes the acceptance of an inconsistent set of beliefs, each saying of a particular participant in a lottery that he will lose.¹

It is true that no contradiction arises if conclusions are added to the evidence on whose basis probabilities are calculated. Concluding that a particular person will lose changes the evidential probability that the next person will lose. When there are only 100 people left, we cannot infer the next person will lose, since the evidential probability of this no longer exceeds .99. But this does not eliminate paradox. The paradox is not just that use of a purely probabilistic rule leads to inconsistent beliefs. It is not obviously irrational to have inconsistent beliefs even when we know that they are inconsistent. It has occasionally been suggested² that a rational man believes that he has at least some (other) false beliefs. If so, it follows logically that at least one thing he believes is false (if nothing else, then his belief that he has other false beliefs); a rational man will know that. So a rational man knows that at least one thing he believes is false. Nevertheless it is paradoxical to suppose that we could rationally believe of every participant in a lottery that he will lose; and it is just as paradoxical to suppose that we could

Originally published in G. Harman, *Thought* (Princeton: Princeton University Press, 1973).

rationally believe this of all but 100 participants in a large lottery.

The lottery paradox can be avoided if a purely probabilistic rule of acceptance is taken to be relevant not to the acceptance of various individual hypotheses but rather to the set of what we accept. The idea is that the probability of the whole set must exceed .99. We are free to choose among various hypotheses saying that one or another participant in a lottery loses as long as the probability of the conjunction of all hypotheses accepted remains above .99. (The idea requires a distinction between what is simply accepted and what is accepted as evidence. If we could add new conclusions to the evidence, the lottery paradox would be generated as indicated in the previous paragraph.) However, although this version of a purely probabilistic rule does not yield the lottery paradox, it does not fit in with ordinary views, as I shall now argue.

Gettier examples and probabilistic rules of acceptance

In any Gettier example we are presented with similar cases in which someone infers *h* from things he knows, *h* is true, and he is equally justified in making the inference in either case.³ In the one case he comes to know that *h* and in the other case he does not. I have observed that a natural explanation of many Gettier examples is that the relevant inference involves not only the final conclusion *h* but also at least one intermediate conclusion true in the one case but not in the other. And I have suggested that any account of inductive inference should show why such intermediate conclusions are essentially involved in the relevant inferences. Gettier cases are thus to be explained by appeal to the principle

P Reasoning that essentially involves false conclusions, intermediate or final, cannot give one knowledge.

It is easy to see that purely probabilistic rules of acceptance do not permit an explanation of Gettier examples by means of principle *P*. Reasoning in accordance with a purely probabilistic rule involves essentially only its final conclusion. Since that conclusion is highly probable, it can be inferred without reference to any other

conclusions; in particular, there will be no intermediate conclusion essential to the inference that is true in one case and false in the other.

For example, Mary's friend Mr Nogot convinces her that he has a Ford. He tells her that he owns a Ford, he shows her his ownership certificate, and he reminds her that she saw him drive up in a Ford. On the basis of this and similar evidence, Mary concludes that Mr Nogot owns a Ford. From that she infers that one of her friends owns a Ford. In a normal case, Mary might in this way come to know that one of her friends owns a Ford. However, as it turns out in this case, Mary is wrong about Nogot. His car has just been repossessed and towed away. It is no longer his. On the other hand, Mary's friend Mr Havit does own a Ford, so she is right in thinking that one of her friends owns a Ford. However, she does not realize that Havit owns a Ford. Indeed, she hasn't been given the slightest reason to think that he owns a Ford. It is false that Mr Nogot owns a Ford, but it is true that one of Mary's friends owns a Ford. Mary has a justified true belief that one of her friends owns a Ford but she does not know that one of her friends owns a Ford. She does not know this because principle *P* has been violated. Mary's reasoning essentially involves the false conclusion that Mr. Nogot owns a Ford.⁴

But, if there were probabilistic rules of acceptance, there would be no way to exhibit the relevance of Mary's intermediate conclusion. For Mary could then have inferred her final conclusion (that one of her friends owns a Ford) directly from her original evidence, all of which is true. Mr Nogot *is* her friend, he *did* say he owns a Ford, he *did* show Mary an ownership certificate, she *did* see him drive up in a Ford, etc. If a purely probabilistic rule would permit Mary to infer from that evidence that her friend Nogot owns a Ford, it would also permit her to infer directly that one of her friends owns a Ford, since the latter conclusion is at least as probable on the evidence as the former. Given a purely probabilistic rule of acceptance, Mary need not first infer an intermediate conclusion and then deduce her final conclusion, since by means of such a rule she could directly infer her final conclusion. The intermediate conclusion would not be essential to her inference, and her failure to know that one of her friends owns a Ford could not be explained by appeal to principle *P*.

A defender of purely probabilistic rules might reply that what has gone wrong in this case is not that Mary *must* infer her conclusion from something false but rather that, from the evidence that supports her conclusion, she *could* also infer something false, namely that Mr Nogot owns a Ford. In terms of principle *P*, this would be to count as essential to Mary's inference any conclusion the probabilistic rule would authorize from her starting point. But given any evidence, some false conclusion will be highly probable on that evidence. This follows, e.g., from the existence of lotteries. For example, let *s* be a conclusion saying under what conditions the New Jersey State Lottery was most recently held. Let *q* say what ticket won the grand prize. Then consider the conclusion, *not both s and q*. Call that conclusion *r*. The conclusion *r* is highly probable, given evidence having nothing to do with the outcome of the recent lottery, but *r* is false. If such highly probable false conclusions were always considered essential to an inference, Mary could never come to know anything.

The problem is that purely probabilistic considerations do not suffice to account for the peculiar relevance of Mary's conclusion about Nogot. Various principles might be suggested; but none of them work. For example, we might suspect that the trouble with *r* is that it has nothing to do with whether any of Mary's friends owns a Ford. Even if Mary were to assume that *r* is false, her original conclusion would continue to be highly probable on her evidence. So we might suggest that an inferable conclusion *t* is essential to an inference only if the assumption that *t* was false would block the inference. That would distinguish Mary's relevant intermediate conclusion, that Nogot owns a Ford, from the irrelevant conclusion *r*, since if Mary assumed that Nogot does not own a Ford she could not conclude that one of her friends owns a Ford.

But again, if there is a purely probabilistic rule of acceptance, there will always be an inferable false *t* such that the assumption that it is false would block even inferences that give us knowledge. For let *h* be the conclusion of any inference not concerned with the New Jersey Lottery and let *r* be as above. Then we can let *t* be the conjunction *h* & *r*. This *t* is highly probable on the same evidence *e* on which *h* is highly probable; *t* is false; and *h* is not highly probable relative to the

evidence *e* & (*not t*). Any inference would be undermined by such a *t*, given a purely probabilistic rule of acceptance along with the suggested criterion of essential conclusions.

The trouble is that purely probabilistic rules are incompatible with the natural account of Gettier examples by means of principle *P*. The solution is not to attempt to modify *P* but rather to modify our account of inference.

Knowledge and Explanation

A causal theory

Goldman suggests that we know only if there is the proper sort of causal connection between our belief and what we know.⁵ For example, we perceive that there has been an automobile accident only if the accident is relevantly causally responsible, by way of our sense organs, for our belief that there has been an accident. Similarly, we remember doing something only if having done it is relevantly causally responsible for our current memory of having done it. Although in some cases the fact that we know thus simply begins a causal chain that leads to our belief, in other cases the causal connection is more complicated. If Mary learns that Mr Havit owns a Ford, Havit's past ownership is causally responsible for the evidence she has and also responsible (at least in part) for Havit's present ownership. Here the relevant causal connection consists in there being a common cause of the belief and of the state of affairs believed in.

Mary fails to know in the original Nogot–Havit case because the causal connection is lacking. Nogot's past ownership is responsible for her evidence but is not responsible for the fact that one of her friends owns a Ford. Havit's past ownership at least partly accounts for why one of her friends now owns a Ford, but it is not responsible for her evidence. Similarly, the man who is told something true by a speaker who does not believe what he says fails to know because the truth of what is said is not causally responsible for the fact that it is said.

General knowledge does not fit into this simple framework. That all emeralds are green neither causes nor is caused by the existence of the particular green emeralds examined when we come

to know that all emeralds are green. Goldman handles such examples by counting logical connections among the causal connections. The belief that all emeralds are green is, in an extended sense, relevantly causally connected to the fact that all emeralds are green, since the evidence causes the belief and is logically entailed by what is believed.

It is obvious that not every causal connection, especially in this extended sense, is relevant to knowledge. Any two states of affairs are logically connected simply because both are entailed by their conjunction. If every such connection were relevant, the analysis Goldman suggests would have us identify knowledge with true belief, since there would always be a relevant "causal connection" between any state of true belief and the state of affairs believed in. Goldman avoids this reduction of his analysis to justified true belief by saying that when knowledge is based on inference relevant causal connections must be "reconstructed" in the inference. Mary knows that one of her friends owns a Ford only if her inference reconstructs the relevant causal connection between evidence and conclusion.

But what does it mean to say that her inference must "reconstruct" the relevant causal connection? Presumably it means that she must infer or be able to infer something about the causal connection between her conclusion and the evidence for it. And this suggests that Mary must make at least two inferences. First she must infer her original conclusion and second she must infer something about the causal connection between the conclusion and her evidence. Her second conclusion is her "reconstruction" of the causal connection. But how detailed must her reconstruction be? If she must reconstruct every detail of the causal connection between evidence and conclusion, she will never gain knowledge by way of inference. If she need only reconstruct some "causal connection," she will always know, since she will always be able to infer that evidence and conclusion are both entailed by their conjunction.

I suggest that it is a mistake to approach the problem as a problem about what else Mary needs to infer before she has knowledge of her original conclusion. Goldman's remark about reconstructing the causal connection makes more sense as a remark about the kind of inference

Mary needs to reach her original conclusion in the first place. It has something to do with principle *P* and the natural account of the Gettier examples.

Nogot presents Mary with evidence that he owns a Ford. She infers that one of her friends owns a Ford. She is justified in reaching that conclusion and it is true. However, since it is true, not because Nogot owns a Ford, but because Havit does, Mary fails to come to know that one of her friends owns a Ford. The natural explanation is that she must infer that Nogot owns a Ford and does not know her final conclusion unless her intermediate conclusion is true. According to this natural explanation, Mary's inference essentially involves the conclusion that Nogot owns a Ford. According to Goldman, her inference essentially involves a conclusion concerning a causal connection. In order to put these ideas together, we must turn Goldman's theory of knowledge into a theory of inference.

As a first approximation, let us take his remarks about causal connections literally, forgetting for the moment that they include logical connections. Then let us transmute his causal theory of knowing into the theory that inductive conclusions always take the form *X causes Y*, where further conclusions are reached by additional steps of inductive or deductive reasoning. In particular, we may deduce either *X* or *Y* from *X causes Y*.

This causal theory of inferring provides the following account of why knowledge requires that we be right about an appropriate causal connection. A person knows by inference only if all conclusions essential to that inference are true. That is, his inference must satisfy principle *P*. Since he can legitimately infer his conclusion only if he can first infer certain causal statements, he can know only if he is right about the causal connection expressed by those statements. First, Mary infers that her evidence is a causal result of Nogot's past ownership of the Ford. From that she deduces that Nogot has owned a Ford. Then she infers that his past ownership has been causally responsible for present ownership; and she deduces that Nogot owns a Ford. Finally, she deduces that one of her friends owns a Ford. She fails to know because she is wrong when she infers that Nogot's past ownership is responsible for Nogot's present ownership.

Inference to the best explanatory statement

A better account of inference emerges if we replace “cause” with “because.” On the revised account, we infer not just statements of the form *X causes Y* but, more generally, statements of the form *Y because X* or *X explains Y*. Inductive inference is conceived as inference to the best of competing explanatory statements. Inference to a causal explanation is a special case.

The revised account squares better with ordinary usage. Nogot’s past ownership helps to explain Mary’s evidence, but it would sound odd to say that it caused that evidence. Similarly, the detective infers that activities of the butler explain these footprints; does he infer that those activities caused the footprints? A scientist explains the properties of water by means of a hypothesis about unobservable particles that make up the water, but it does not seem right to say that facts about those particles cause the properties of water. An observer infers that certain mental states best explain someone’s behavior; but such explanation by reasons might not be causal explanation.

Furthermore, the switch from “cause” to “because” avoids Goldman’s *ad hoc* treatment of knowledge of generalizations. Although there is no causal relation between a generalization and those observed instances which provide us with evidence for the generalization, there is an obvious explanatory relationship. That all emeralds are green does not cause a particular emerald to be green; but it can explain why that emerald is green. And, other things being equal, we can infer a generalization only if it provides the most plausible way to explain our evidence.

We often infer generalizations that explain but do not logically entail their instances, since they are of the form, *In circumstances C, X’s tend to be Y’s*. Such generalizations may be inferred if they provide a sufficiently plausible account of observed instances all things considered. For example, from the fact that doctors have generally been right in the past when they have said that someone is going to get measles, I infer that doctors can normally tell from certain symptoms that someone is going to get measles. More precisely, I infer that doctors have generally been right in the past because they can normally tell from certain symptoms that someone is going to

get measles. This is a very weak explanation, but it is a genuine one. Compare it with the pseudo-explanation, “Doctors are generally right when they say someone has measles because they can normally tell from certain symptoms that someone is going to get measles.”

Similarly, I infer that a substance is soluble in water from the fact that it dissolved when I stirred it into some water. That is a real explanation, to be distinguished from the pseudo-explanation, “That substance dissolves in water because it is soluble in water.” Here too a generalization explains an instance without entailing that instance, since water-soluble substances do not always dissolve in water.

Although we cannot simply deduce instances from this sort of generalization, we can often infer that the generalization will explain some new instance. The inference is warranted if the explanatory claim *that X’s tend to be Y’s will explain why the next X will be Y* is sufficiently more plausible than competitors such as *interfering factor Q will prevent the next X from being a Y*. For example, the doctor says that you will get measles. Because doctors are normally right about that sort of thing, I infer that you will. More precisely, I infer that doctors’ normally being able to tell when someone will get measles will explain the doctor’s being right in this case. The competing explanatory statements here are not other explanations of the doctor’s being right but rather explanations of his being wrong – e.g., because he has misperceived the symptoms, or because you have faked the symptoms of measles, or because these symptoms are the result of some other disease, etc. Similarly, I infer that this sugar will dissolve in my tea. That is, I infer that the solubility of sugar in tea will explain this sugar’s dissolving in the present case. Competing explanations would explain the sugar’s not dissolving – e.g., because there is already a saturated sugar solution there, because the tea is ice-cold, etc.

Further examples⁶

I infer that when I scratch this match it will light. My evidence is that this is a Sure-Fire brand match, and in the past Sure-Fire matches have always lit when scratched. However, unbeknownst to me, this particular match is defective. It will not light unless its surface temperature can be

raised to six hundred degrees, which is more than can be attained by scratching. Fortunately, as I scratch the match, a burst of Q-radiation (from the sun) strikes the tip, raising surface temperature to six hundred degrees and igniting the match. Did I know that the match would light? Presumably I did not know. I had justified true belief, but not knowledge. On the present account, the explanation of my failure to know is this: I infer that the match will light in the next instance because Sure-Fire matches generally light when scratched. I am wrong about that; that is not why the match will light this time. Therefore, I do not know that it will light.

It is important that our justification can appeal to a simple generalization even when we have false views about the explanation of that generalization. Consider the man who thinks that barometers fall before a rainstorm because of an increase in the force of gravity. He thinks the gravity pulls the mercury down the tube and then, when the force is great enough, pulls rain out of the sky. Although he is wrong about this explanation, the man in question can come to know that it is going to rain when he sees the barometer falling in a particular case. That a man's belief is based on an inference that cannot give him knowledge (because it infers a false explanation) does not mean that it is not also based on an inference that does give him knowledge (because it infers a true explanation). The man in question has knowledge because he infers not only the stronger explanation involving gravity but also the weaker explanation. He infers that the explanation of the past correlation between falling barometer and rain is that the falling barometer is normally associated with rain. Then he infers that this weak generalization will be what will explain the correlation between the falling barometer and rain in the next instance.

Notice that if the man is wrong about that last point, because the barometer is broken and is leaking mercury, so that it is just a coincidence that rain is correlated with the falling barometer in the next instance, he does not come to know that it is going to rain.

Another example is the mad-fiend case. Omar falls down drunk in the street. An hour later he suffers a fatal heart attack not connected with his recent drinking. After another hour a mad fiend comes down the street, spies Omar lying in the

gutter, cuts off his head, and runs away. Some time later still, you walk down the street, see Omar lying there, and observe that his head has been cut off. You infer that Omar is dead; and in this way you come to know that he is dead. Now there is no causal connection between Omar's being dead and his head's having been cut off. The fact that Omar is dead is not causally responsible for his head's having been cut off, since if he had not suffered that fatal heart attack he still would have been lying there drunk when the mad fiend came along. And having his head cut off did not cause Omar's death, since he was already dead. Nor is there a straightforward logical connection between Omar's being dead and his having his head cut off. (Given the right sorts of tubes, one might survive decapitation.) So it is doubtful that Goldman's causal theory of knowing can account for your knowledge that Omar is dead.

If inductive inference is inference to the best explanatory statement, your inference might be parsed as follows: "Normally, if someone's head is cut off, that person is dead. This generalization accounts for the fact that Omar's having his head cut off is correlated here with Omar's being dead." Relevant competing explanatory statements in this case would not be competing explanations of Omar's being dead. Instead they would seek to explain Omar's not being dead despite his head's having been cut off. One possibility would be that doctors have carefully connected head and body with special tubes so that blood and air get from body to head and back again. You rule out that hypothesis on grounds of explanatory complications: too many questions left unanswered (why can't you see the tubes? why wasn't it done in the hospital? etc.). If you cannot rule such possibilities out, then you cannot come to know that Omar is dead. And if you do rule them out but they turn out to be true, again you do not come to know. For example, if it is all an elaborate psychological philosophical experiment, which however fails, then you do not come to know that Omar is dead even though he is dead.

Statistical inference

Statistical inference, and knowledge obtained from it, is also better explicated by way of the notion of statistical explanation than by way of

the notion of cause or logical entailment. A person may infer that a particular coin is biased because that provides the best statistical explanation of the observed fraction of heads. His conclusion explains his evidence but neither causes nor entails it.

The relevant kind of statistical explanation does not always make what it explains very probable. For example, suppose that I want to know whether I have the fair coin or the weighted coin. It is equally likely that I have either; the probability of getting heads on a toss of the fair coin is $1/2$; and the probability of getting heads on a toss of the weighted coin is $6/10$. I toss the coin 10,000 times. It comes up heads 4,983 times and tails 5,017. I correctly conclude that the coin is the fair one. You would ordinarily think that I could in this way come to know that I have the fair coin. On the theory of inference we have adopted, I infer the best explanation of the observed distribution of heads and tails. But the explanation, that these were random tosses of a fair coin, does not make it probable that the coin comes up heads exactly 4,983 times and tails exactly 5,017 times in 10,000 tosses. The probability of this happening with a fair coin is very small. If we want to accept the idea that inference is inference to the best explanatory statement, we must agree that statistical explanation can cite an explanation that makes what it explains less probable than it makes its denial. In the present case, I do not explain why 4,983 heads have come up rather than some other number of heads. Instead I explain how it happened that 4,983 heads came up, what led to this happening. I do not explain why this happened rather than something else, since the same thing could easily have led to something else.

To return to an example I have used elsewhere, you walk into a casino and see the roulette wheel stop at red fifty times in a row. The explanation may be that the wheel is fixed. It may also be that the wheel is fair and this is one of those times when fifty reds come up on a fair wheel. Given a fair wheel we may expect that to happen sometimes (but not very often). But if the explanation is that the wheel is fair and that this is just one of those times, it says what the sequence of reds is the result of, the "outcome" of. It does not say why fifty reds in a row occurred this time rather than some other time, nor why that particular

series occurred rather than any of the $2^{50}-1$ other possible series.

This kind of statistical explanation explains something as the outcome of a chance set-up. The statistical probability of getting the explained outcome is irrelevant to whether or not we explain that outcome, since this kind of explanation is essentially pure nondeterministic explanation. All that is relevant is that the outcome to be explained is one possible outcome given that chance set-up. That is not to say that the statistical probability of an outcome is irrelevant to the explanation of that outcome. It is relevant in this sense: the greater the statistical probability an observed outcome has in a particular chance set-up, the better that set-up explains that outcome.

The point is less a point about statistical explanation than a point about statistical inference. I wish to infer the best of competing statistical explanations of the observed distribution of heads. This observed outcome has different statistical probabilities in the two hypothetical chance set-ups, fair coin or weighted coin. The higher this statistical probability, the better, from the point of view of inference (other things being equal). The statistical probability of an outcome in a particular hypothetical chance set-up is relevant to how good an explanation that chance set-up provides. Here a better explanation is one that is more likely to be inferable. For example, I infer that I have the fair coin. The statistical probability of 4,983 heads on 10,000 tosses of a fair coin is much greater than the statistical probability of that number of heads on 10,000 tosses of the weighted coin. From the point of view of statistical probability, the hypothesis that the coin is fair offers a better explanation of the observed distribution than the hypothesis that the coin is biased. So statistical probability is relevant to statistical explanation. Not that there is no explanation unless statistical probability is greater than $1/2$. Rather that statistical probability provides a measure of the inferability of a statistical explanation.

According to probability theory, if initially the coin is just as likely to be the fair one or the weighted one and the statistical probability of the observed outcome is much greater for the fair coin than for the weighted coin, the probability that the coin is fair, given the observed evidence, will be very high. We might conclude that the

statistical probability of the observed outcome given the fair or weighted coin is only indirectly relevant to my inference, relevant only because of the theoretical connections between those statistical probabilities and the evidential probabilities of the two hypotheses about the coin, given the observed evidence. But that would be to get things exactly backward. No doubt there is a connection between high evidential probability and inference; but, as we have seen, it is not because there is a purely probabilistic rule of acceptance. High probability by itself does not warrant inference. Only explanatory considerations can do that; and the probability relevant to explanation is statistical probability, the probability that is involved in statistical explanation. It is the statistical probabilities of the observed outcome, given the fair and weighted coins, that is directly relevant to inference. The evidential probabilities of the two hypotheses are only indirectly relevant in that they in some sense reflect the inferability of the hypotheses, where that is determined directly by considerations of statistical probability.

Suppose that at first you do not know which of the two coins I have selected. I toss it 10,000 times, getting 4,983 heads and 5,017 tails. You infer that I have the fair coin, and you are right. But the reason for the 4,983 heads is that I am very good at tossing coins to come up whichever way I desire and I deliberately tossed the coin so as to get roughly half heads and half tails. So, even though you have justified true belief, you do not know that I have the fair coin.

If statistical inference were merely a matter of inferring something that has a high probability on the evidence, there would be no way to account for this sort of Gettier example. And if we are to appeal to principle *P*, it must be a conclusion essential to your inference that the observed outcome is the result of a chance set-up involving the fair coin in such a way that the probability of heads is $1/2$. Given a purely probabilistic rule, that conclusion could not be essential, for reasons similar to those that have already been discussed concerning the Nogot–Havit case. On the other hand, if statistical inference is inference to the best explanation and there is such a thing as statistical explanation even where the statistical probability of what is explained is quite low, then your conclusion about the reason for my getting 4,983 heads is seen to be essential to your

inference. Since your explanation of the observed outcome is false, principle *P* accounts for the fact that you do not come to know that the coin is the fair coin even though you have justified true belief.

Conclusion

We are led to construe induction as inference to the best explanation, or more precisely as inference to the best of competing explanatory statements. The conclusion of any single step of such inference is always of the form *Y because X* (or *X explains Y*), from which we may deduce either *X* or *Y*. Inductive reasoning is seen to consist in a sequence of such explanatory conclusions.

We have been led to this conception of induction in an attempt to account for Gettier examples that show something wrong with the idea that knowledge is justified true belief. We have tried to find principles of inference which, together with principle *P*, would explain Gettier's deviant cases. Purely probabilistic rules were easily seen to be inadequate. Goldman's causal theory of knowing, which promised answers to some of Gettier's questions, suggested a causal theory of induction: inductive inference as inference to the best of competing causal statements. Our present version is simply a modification of that, with *explanatory* replacing *causal*. Its strength lies in the fact that it accounts for a variety of inferences, including inferences that involve weak generalizations or statistical hypotheses, in a way that explains Gettier examples by means of principle *P*.

Evidence One Does Not Possess

Three examples

Example (1)

While I am watching him, Tom takes a library book from the shelf and conceals it beneath his coat. Since I am the library detective, I follow him as he walks brazenly past the guard at the front door. Outside I see him take out the book and smile. As I approach he notices me and suddenly runs away. But I am sure that it was Tom, for I know him well. I saw Tom steal a book from the library and that is the testimony I give before the University Judicial Council. After testifying, I leave the hearing room and return to my post in

the library. Later that day, Tom's mother testifies that Tom has an identical twin, Buck. Tom, she says, was thousands of miles away at the time of the theft. She hopes that Buck did not do it; but she admits that he has a bad character.

Do I know that Tom stole the book? Let us suppose that I am right. It was Tom that took the book. His mother was lying when she said that Tom was thousands of miles away. I do not know that she was lying, of course, since I do not know anything about her, even that she exists. Nor does anyone at the hearing know that she is lying, although some may suspect that she is. In these circumstances I do not know that Tom stole the book. My knowledge is undermined by evidence I do not possess.⁷

Example (2)

Donald has gone off to Italy. He told you ahead of time that he was going; and you saw him off at the airport. He said he was to stay for the entire summer. That was in June. It is now July. Then you might know that he is in Italy. It is the sort of thing one often claims to know. However, for reasons of his own Donald wants you to believe that he is not in Italy but in California. He writes several letters saying that he has gone to San Francisco and has decided to stay there for the summer. He wants you to think that these letters were written by him in San Francisco, so he sends them to someone he knows there and has that person mail them to you with a San Francisco postmark, one at a time. You have been out of town for a couple of days and have not read any of the letters. You are now standing before the pile of mail that arrived while you were away. Two of the phony letters are in the pile. You are about to open your mail. I ask you, "Do you know where Donald is?" "Yes," you reply, "I know that he is in Italy." You are right about where Donald is and it would seem that your justification for believing that Donald is in Italy makes no reference to letters from San Francisco. But you do not know that Donald is in Italy. Your knowledge is undermined by evidence you do not as yet possess.

Example (3)

A political leader is assassinated. His associates, fearing a coup, decide to pretend that the bullet hit someone else. On nationwide television they announce that an assassination attempt has failed

to kill the leader but has killed a secret service man by mistake. However, before the announcement is made, an enterprising reporter on the scene telephones the real story to his newspaper, which has included the story in its final edition. Jill buys a copy of that paper and reads the story of the assassination. What she reads is true and so are her assumptions about how the story came to be in the paper. The reporter, whose by-line appears, saw the assassination and dictated his report, which is now printed just as he dictated it. Jill has justified true belief and, it would seem, all her intermediate conclusions are true. But she does not know that the political leader has been assassinated. For everyone else has heard about the televised announcement. They may also have seen the story in the paper and, perhaps, do not know what to believe; and it is highly implausible that Jill should know simply because she lacks evidence everyone else has. Jill does not know. Her knowledge is undermined by evidence she does not possess.

These examples pose a problem for my strategy. They are Gettier examples and my strategy is to make assumptions about inference that will account for Gettier examples by means of principle *P*. But these particular examples appear to bring in considerations that have nothing to do with conclusions essential to the inference on which belief is based.

Some readers may have trouble evaluating these examples. Like other Gettier examples, these require attention to subtle facts about ordinary usage; it is easy to miss subtle differences if, as in the present instance, it is very difficult to formulate a theory that would account for these differences. We must compare what it would be natural to say about these cases if there were no additional evidence one does not possess (no testimony from Tom's mother, no letters from San Francisco, and no televised announcement) with what it would be natural to say about the cases in which there is the additional evidence one does not possess. We must take care not to adopt a very skeptical attitude nor become too lenient about what is to count as knowledge. If we become skeptically inclined, we will deny there is knowledge in either case. If we become too lenient, we will allow that there is knowledge in both cases. It is tempting to go in one or the other of these directions, toward skepticism or leniency, because it proves so

difficult to see what general principles are involved that would mark the difference. But at least some difference between the cases is revealed by the fact that we are *more inclined* to say that there is knowledge in the examples where there is no undermining evidence a person does not possess than in the examples where there is such evidence. The problem, then, is to account for this difference in our inclination to ascribe knowledge to someone.

Evidence against what one knows

If I had known about Tom's mother's testimony, I would not have been justified in thinking that it was Tom I saw steal the book. Once you read the letters from Donald in which he says he is in San Francisco, you are no longer justified in thinking that he is in Italy. If Jill knew about the television announcement, she would not be justified in believing that the political leader has been assassinated. This suggests that we can account for the preceding examples by means of the following principle.

One knows only if there is no evidence such that if one knew about the evidence one would not be justified in believing one's conclusion.

However, by modifying the three examples it can be shown that this principle is too strong.

Suppose that Tom's mother was known to the Judicial Council as a pathological liar. Everyone at the hearing realizes that Buck, Tom's supposed twin, is a figment of her imagination. When she testifies no one believes her. Back at my post in the library, I still know nothing of Tom's mother or her testimony. In such a case, my knowledge would not be undermined by her testimony; but if I were told only that she had just testified that Tom has a twin brother and was himself thousands of miles away from the scene of the crime at the time the book was stolen, I would no longer be justified in believing as I now do that Tom stole the book. Here I know even though there is evidence which, if I knew about it, would cause me not to be justified in believing my conclusion.

Suppose that Donald had changed his mind and never mailed the letters to San Francisco. Then those letters no longer undermine your

knowledge. But it is very difficult to see what principle accounts for this fact. How can letters in the pile on the table in front of you undermine your knowledge while the same letters in a pile in front of Donald do not? If you knew that Donald had written letters to you saying that he was in San Francisco, you would not be justified in believing that he was still in Italy. But that fact by itself does not undermine your present knowledge that he is in Italy.

Suppose that as the political leader's associates are about to make their announcement, a saboteur cuts the wire leading to the television transmitter. The announcement is therefore heard only by those in the studio, all of whom are parties to the deception. Jill reads the real story in the newspaper as before. Now, she does come to know that the political leader has been assassinated. But if she had known that it had been announced that he was not assassinated, she would not have been justified in believing that he was, simply on the basis of the newspaper story. Here, a cut wire makes the difference between evidence that undermines knowledge and evidence that does not undermine knowledge.

We can know that h even though there is evidence e that we do not know about such that, if we did know about e , we would not be justified in believing h . If we know that h , it does not follow that we know that there is not any evidence like e . This can seem paradoxical, for it can seem obvious that, if we know that h , we know that any evidence against h can only be misleading. So, later if we get that evidence we ought to be able to know enough to disregard it.

A more explicit version of this interesting paradox goes like this.⁸ "If I know that h is true, I know that any evidence against h is evidence against something that is true; so I know that such evidence is misleading. But I should disregard evidence that I know is misleading. So, once I know that h is true, I am in a position to disregard any future evidence that seems to tell against h ." This is paradoxical, because I am never in a position simply to disregard any future evidence even though I do know a great many different things.

A skeptic might appeal to this paradox in order to argue that, since we are never in a position to disregard any further evidence, we never know anything. Some philosophers would turn the

argument around to say that, since we often know things, we are often in a position to disregard further evidence. But both of these responses go wrong in accepting the paradoxical argument in the first place.

I can know that Tom stole a book from the library without being able automatically to disregard evidence to the contrary. You can know that Donald is in Italy without having the right to ignore whatever further evidence may turn up. Jill may know that the political leader has been assassinated even though she would cease to know this if told that there was an announcement that only a secret service agent had been shot.

The argument for paradox overlooks the way actually having evidence can make a difference. Since I now know that Tom stole the book, I now know that any evidence that appears to indicate something else is misleading. That does not warrant me in simply disregarding any further evidence, since getting that further evidence can change what I know. In particular, after I get such further evidence I may no longer know that it is misleading. For having the new evidence can make it true that I no longer know that Tom stole the book; if I no longer know that, I no longer know that the new evidence is misleading.

Therefore, we cannot account for the problems posed by evidence one does not possess by appeal to the principle, which I now repeat:

One knows only if there is no evidence such that if one knew about the evidence one would not be justified in believing one's conclusion.

For one can know even though such evidence exists.

A result concerning inference

When does evidence one doesn't have keep one from having knowledge? I have described three cases, each in two versions, in which there is misleading evidence one does not possess. In the first version of each case the misleading evidence undermines someone's knowledge. In the second version it does not. What makes the difference?

My strategy is to account for Gettier examples by means of principle *P*. This strategy has led us to conceive of induction as inference to the best explanation. But that conception of inference

does not by itself seem able to explain these examples. So I want to use the examples in order to learn something more about inference, in particular about what other conclusions are essential to the inference that Tom stole the book, that Donald is in Italy, or that the political leader has been assassinated.

It is not plausible that the relevant inferences should contain essential intermediate conclusions that refer explicitly to Tom's mother, to letters from San Francisco, or to special television programs. For it is very likely that there is an infinite number of ways a particular inference might be undermined by misleading evidence one does not possess. If there must be a separate essential conclusion ruling out each of these ways, inferences would have to be infinitely inclusive – and that is implausible.

Therefore it would seem that the relevant inferences must rule out undermining evidence one does not possess by means of a single conclusion, essential to the inference, that characterizes all such evidence. But how might this be done? It is not at all clear what distinguishes evidence that undermines knowledge from evidence that does not. How is my inference to involve an essential conclusion that rules out Tom's mother's testifying a certain way before a believing audience but does not rule out (simply) her testifying in that way? Or that rules out the existence of letters of a particular sort in the mail on your table but not simply the existence of those letters? Or that rules out a widely heard announcement of a certain sort without simply ruling out the announcement?

Since I am unable to formulate criteria that would distinguish among these cases, I will simply *label* cases of the first kind "undermining evidence one does not possess." Then we can say this: one knows only if there is no undermining evidence one does not possess. If there is such evidence, one does not know. However, these remarks are completely trivial.

It is somewhat less trivial to use the same label to formulate a principle concerned with inference.

Q One may infer a conclusion only if one also infers that there is no undermining evidence one does not possess.

There is of course an obscurity in principle *Q*; but the principle is not as trivial as the remarks of the

last paragraph, since the label “undermining evidence one does not possess” has been explained in terms of knowledge, whereas this is a principle concerning inference.

If we take principle *Q*, concerning inference, to be basic, we can use principle *P* to account for the differences between the two versions of each of the three examples described above. In each case an inference involves essentially the claim that there is no undermining evidence one does not possess. Since this claim is false in the first version of each case and true in the second, principle *P* implies that there can be knowledge only in the second version of each case.

So there is, according to my strategy, some reason to think that there is a principle concerning inference like principle *Q*. That raises the question of whether there is any independent reason to accept such a principle; and reflection on good scientific practice suggests a positive answer. It is a commonplace that a scientist should base his conclusions on all the evidence. Furthermore, he should not rest content with the evidence he happens to have but should try to make sure he is not overlooking any relevant evidence. A good scientist will not accept a conclusion unless he has some reason to think that there is no as yet undiscovered evidence which would undermine his conclusion. Otherwise he would not be warranted in making his inference. So good scientific practice reflects the acceptance of something like principle *Q*, which is the independent confirmation we wanted for the existence of this principle.

Notice that the scientist must accept something like principle *Q*, with its reference to “undermining evidence one does not possess.” For example, he cannot accept the following principle,

One may infer a conclusion only if one also infers that there is no evidence at all such that if he knew that evidence he could not accept his conclusion.

There will always be a true proposition such that if he learned that the proposition was true (and learned nothing else) he would not be warranted in accepting his conclusion. If *h* is his conclusion, and if *k* is a true proposition saying what ticket will win the grand prize in the next New

Jersey State Lottery, then *either k or not h* is such a proposition. If he were to learn that it is true that *either k or not h* (and learned nothing else), *not h* would become probable since (given what he knows) *k* is antecedently very improbable. So he could no longer reasonably infer that *h* is true.

There must be a certain kind of evidence such that the scientist infers there is no as yet undiscovered evidence of that kind against *h*. Principle *Q* says that the relevant kind is what I have been labelling “undermining evidence one does not possess.” Principle *Q* is confirmed by the fact that good scientific practice involves some such principle and by the fact that principle *Q* together with principle *P* accounts for the three Gettier examples I have been discussing.

If this account in terms of principles *P* and *Q* is accepted, inductive conclusions must involve some self-reference. Otherwise there would be a regress. Before we could infer that *h*, we would have to infer that there is no undermining evidence to *h*. That prior inference could not be deductive, so it would have to be inference to the best explanatory statement. For example, we might infer that the fact that there is no sign of undermining evidence we do not possess is explained by there not being any such evidence. But, then, before we could accept that conclusion we would first have to infer that there is no undermining evidence to *it* which one does not possess. And, since that inference would have to be inference to the best explanation, it would require a previous inference that there is no undermining evidence for its conclusion; and so on *ad infinitum*.

Clearly, we do not *first* have to infer that there is no undermining evidence to *h* and only then infer *h*. For that would automatically yield the regress. Instead, we must at the same time infer both *h* and that there is no undermining evidence. Furthermore, we infer that there is not only no undermining evidence to *h* but also no undermining evidence to the whole conclusion. In other words, all legitimate inductive conclusions take the form of a self-referential conjunction whose first conjunct is *h* and whose second conjunct (usually left implicit) is the claim that there is no undermining evidence to the whole conjunction.

Notes

- 1 Henry Kyburg, *Probability and the Logic of Rational Belief* (Middletown: Wesleyan University Press, 1961).
- 2 E.g., by Robert Nozick.
- 3 Edmond Gettier, "Is Justified True Belief Knowledge?", this vol., ch. 15.
- 4 Keith Lehrer, "Knowledge, Truth, and Evidence," *Analysis* 25 (1965), pp. 168–75.
- 5 Alvin Goldman, "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967), pp. 357–72.
- 6 Brian Skyrms, "The Explication of 'X knows that *p*,'" *Journal of Philosophy* 64 (1967), pp. 373–89.
- 7 Keith Lehrer and Thomas Paxson, Jr., "Knowledge: Undefeated Justified True Belief," *Journal of Philosophy* 66 (1969).
- 8 Here and in what follows I am indebted to Saul Kripke, who is, however, not responsible for any faults in my presentation.