Detachment, Probability, and Maximum Likelihood

GILBERT HARMAN
PRINCETON UNIVERSITY

When can we detach probability qualifications from our inductive conclusions? The following rule may seem plausible:

> If the probability of h on the total evidence is at least N-1/N, infer h.

However, what has been called "the lottery paradox" shows that we cannot accept this rule if we also accept the usual principles of deductive logic. For it would allow us to infer of each ticket in a fair N ticket lottery that the ticket will lose; and standard principles of deduction would then allow us to infer that no ticket will win (although we know that one of them will win). The suggested rule, taken with standard deductive principles, would permit us to derive a contradiction from consistent and possible evidence.

Philosophers have suggested at least four ways to deal with the lottery paradox. One view, associated with Savage, Carnap, Jeffrey, and others, denies that inductive inference ever permits detachment and says that inductive inference is an application of the deductive theory of probability. Some proponents of this view take probabilistic inference to be a method of ensuring consistent assignments of subjective probability or degrees of belief.[1] A second theory, of Kyburg's, takes high probability to warrant detachment but rejects the usual principles of deduction: we can infer of each

[1] This view has recently been developed in detail by Richard Jeffrey, *The Logic of Decision* (McGraw-Hill, New York, 1965). When Suppes, who accepts this view, speaks of a rule of detachment, he does not mean a rule for detaching a probability qualification: Patrick Suppes, "Probabilistic Inference and the Concept of Total Evidence," eds. Hintikka and Suppes, *Aspects of Inductive Logic* (North-Holland, Amsterdam, 1966): 49–65.

ticket that it will lose but we cannot go on to infer that all tickets will lose.[2] This apparently radical theory is essentially the first view coupled with an additional terminological proposal: whenever subjective probability or degree of belief assigned to a proposition exceeds a certain amount, we are to say the proposition is "believed." A third proposal by Hintikka takes high probability to warrant detachment, but only within a very limited language and only when the evidence is such that the lottery paradox cannot arise.[3] It is hard to see why in practice Hintikka's proposal will not reduce to the first proposal, since it would appear never to warrant detachment in practice. A fourth suggestion by Isaac Levi takes inductive inference with detachment to be relative to a set of strongest relevant answers: if strongest relevant answers are "Ticket number three loses" and "Ticket number three wins," one may infer that it loses; if the strongest relevant answers are "Ticket number one wins," "Ticket number two wins," etc., no inference is permitted, so the lottery paradox cannot arise even though high probability warrants detachment.[4]

Each of these theories assumes that inductive detachment is possible only if in certain contexts high probability alone warrants detachment. I shall argue that this assumption is wrong. I shall describe other relevant considerations. Although these considerations rule out the lottery paradox as stated, they do not rule out more complicated versions. Nevertheless, they must be taken into account in any realistic theory of inductive and statistical inference.

The following principle provides a method for discovering when an argument warrants detachment:

> If by virtue of a certain argument a person could come to know its conclusion without any probability qualification, the argument warrants detachment of any such qualification.

There are three reasons for accepting this principle. (1) Knowledge is in part warranted belief and warranted belief must in some way be related to warranted inference.[5] (2) To know something one

[2] Henry E. Kyburg, "Probability, Nationality, and a Rule of Detachment," *Logic, Methodology, and Philosophy of Science: Proceedings of the 1964 International Congress*, ed. Bar-Hillel (North-Holland, Amsterdam, 1965): 301–10. Kyburg's more sophisticated version of this paradox is discussed below.

[3] Jaakko Hintikka and Risto Hilpinen, "Knowledge, Acceptance, and Inductive Logic," in Hintikka and Suppes *op. cit.*: 1–20.

[4] Isaac Levi, *Gambling with Truth* (Knopf, New York, 1967).

[5] This has been denied by Peter Unger, "Experience and Factual Knowledge," *The Journal of Philosophy*, LXIV (1967): 152–173. But I am not con-

must believe it fully and not just partially; and where partial belief represents acceptance of a qualified conclusion, full belief represents acceptance of a conclusion whose qualification has been deleted.[6] (3) Acceptance of the principle proves fruitful for an investigation of knowledge, inference, and explanation, as I shall indicate in what follows.

In using the principle one must at first rely on ordinary non-philosophical intuitive judgments about when people know. Judgments biased by standard philosophical prejudices fail to make distinctions that are required if one is to give an accurate account of inductive inference.

The lottery paradox arises from the principle that high probability alone warrants detachment. This principle allows us to infer of each ticket that it has not been selected; and the most plausible way to avoid this paradox is to avoid the inference. That we should avoid it follows from the knowability test, since one cannot come to know that a ticket will lose by arguing from the high probability of its losing.[7] On the other hand people often come to know things that their evidence does not make certain. What one comes to know by reading the morning paper has a smaller probability than the probability that a particular ticket will lose the lottery. It follows from this and the knowability test that the possibility of detachment is not solely a matter of the probability of one's conclusion but depends on other factors.

One such factor is that in the newspaper case but not in the lottery case a person's inference may be treated as an inference to the best of competing explanations. Suppose that the item in the newspaper is a signed column by a reporter who witnessed what he describes. A more complete description of what the reader needs to believe might be this:

> The newspaper says what it does because the printer has accurately set up in type the story as written by the reporter. The reporter has written as he has in part because he believes

---

vinced; see my "Unger on Knowledge," *The Journal of Philosophy* LXIV (1967): 390–395.

[6] There are uses of "know" for which this does not hold. See "Unger on Knowledge." Full belief is compatible with being more or less certain. See Levi, *op. cit.*

[7] Notice that Levi's proposal is incompatible with this. One cannot come to know a particular ticket will lose, even if the strongest relevant answers are that it wins and that it loses.

what he has written. He believes this because he was in fact a witness of the events he relates.

The reader's reasons are presented by three inductive steps of reasoning, each step to a conclusion that explains part of his previous evidence. No similar analysis can be given to the argument in the lottery example. We cannot attempt to explain our evidence, namely that there is to be a lottery of a certain kind, in any way that entails that a particular ticket will lose. The requirement that detachment be permitted only when one infers an explanation adequately distinguishes the lottery example from the newspaper example. (However other forms of the lottery paradox require more sophisticated treatment.)

I have elsewhere described inductive inference as the inference to the best of competing explanations.[8] The present paper attempts to clarify what it is for an explanation to be sufficiently better than competing explanations to permit the inference that the explanation is true.

In order to treat statistical inference as inference to the best explanation, one must have some notion of statistical explanation. This raises large issues; but my position can be sketched as follows: (a) inductive probability, the probability of a conclusion on certain evidence, is to be distinguished from statistical probability, the probability of a particular outcome in a chance set up; (b) a statistical explanation explains something as the outcome of a chance set up; (c) the statistical probability of getting such an outcome in that chance set up need not be greater than some specified amount, although the size of the statistical probability of the outcome is relevant in determining whether one explanation is better than competing alternatives. For example, we may explain a run of fifty reds on the roulette wheel as the result of chance, one of those times the unexpected has occurred. A competing explanation would be that the roulette wheel has been fixed so as to result in red every time. (Of course in the first case we explain what led to the run of reds, not why that run rather than some other.)

---

[8] "The Inference to the Best Explanation," *Philosophical Review*, LXX (1965): 88–95: "Knowledge, Inference, and Explanation," forthcoming in *American Philosophical Quarterly*. My view resembles that of Alvin Goldman, "A Causal Theory of Knowing," *The Journal of Philosophy*, LXIV (1967): 357–72. An important point of difference is that he limits the relevant sort of explanation to causal explanation, which rules out knowledge based on inference to the best statistical explanation.

This is to agree with Carnap's view of statistical explanation[9] as against Hempel's view.[10] The main reason to agree with Carnap is that it simplifies one's theory of knowledge, particularly one's account of statistical inference. Statistical sampling techniques are best seen as based on the assumption that statistical explanation cites a chance set up whose outcome is the thing to be explained. The statistical probability of that outcome in that chance set up is not relevant to the question whether an explanation has been given, it is only relevant to the issue whether the explanation given is sufficiently better than competing explanations to be inferable. Statistical inference often permits us to infer that the explanation of an observed outcome is a chance set up in which that outcome has a small statistical probability. This happens when the statistical probability of that outcome would be even smaller on competing alternatives. For example, suppose a particular coin has been randomly selected from a pair of coins, one of which is completely fair, the other biased towards heads with a statistical probability of 0.9. An investigator tosses the coin ten times, observing that it comes up heads three times and tails seven times. He can in this way come to know that he has the unbiased coin, for the cited statistical evidence permits the inference that this is the fair coin. That he can thus come to know can be accommodated on the present analysis only if it is granted that statistical explanation need not involve high statistical probability.

I would describe his inference as follows: There are two possible competing explanations of his getting three heads and seven tails. This outcome may be the extremely improbable result of random tosses of the biased coin (the statistical probability of such a result with the biased coin being less than .001) or the mildly improbable result of random tosses of the fair coin (the statistical probability of such a result with the fair coin being slightly more than 0.1). Since the latter explanation is much better than its only competitor, one may infer that the coin is fair.

But this is too brief, in fact, several conditions must be satisfied before one can make an inference to the best statistical ex-

---

[9] Rudolf Carnap, *Philosophical Foundations of Physics*, ed. Martin Gardner (Basic Books, New York, 1966), p. 8. I take the notion of the outcome of a chance set up from Ian Hacking, *Logic of Statistical Inference* (Cambridge University Press, 1965).

[10] C. G. Hempel, "Aspects of Scientific Explanation," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, (Free Press, New York, 1965): 376–412.

planation. Some disputes about the theory of statistical inference, e.g. between Bayesians and proponents of a maximum likelihood principle, result from oversimplification. Suppose that we wish to infer an explanatory hypothesis h from observed evidence e_h. A Bayesian theorist requires that the hypothesis h be highly probable on the total evidence (which includes e_h as a part). A maximum likelihood theorist requires that the probability of our having observed e_h be greater on the hypothesis h than on any of the alternatives. These views conflict only if they are taken to offer sufficient conditions of warranted statistical inference. Both provide necessary conditions, as can be seen from a case in which Bayesian and maximum likelihood considerations conflict.

Notice that in the example in which a coin is tossed to determine whether it is fair, both the Bayesian and maximum likelihood conditions support the same hypothesis since we imagine the prior inductive probability that the coin is fair to be 0.5. (In any case in which the prior probabilities of various chance set ups are equal, the inductive probabilities of each chance set up on the total evidence, including the observed outcome, will be proportional to the statistical probabilities of getting that outcome on each chance set up.) We can obtain a conflict by supposing different prior probabilities. For example suppose that the prior probability of the coin's being biased is very large, say 0.99999. Then, in the situation described, even though what we have observed (three heads and seven tails) has a much greater statistical probability on the hypothesis that the coin is fair than on the alternative hypothesis, the inductive probability on the total evidence that the coin is fair is very small. (This probability is about 0.01.) The Bayesian condition is satisfied by the hypothesis that the coin is biased; the maximum likelihood condition by the hypothesis that the coin is fair. In such a case one could not come to know whether the coin is biased or fair, *i.e.* one can make no inference with detachment. A hypothesis must satisfy both conditions before it can be inferred as the conclusion of an inference to the best statistical explanation.

Maximum likelihood theorists are typically interested in the so-called problem of estimation. Such a problem arises when the distribution of chances in a chance set up is partially but not fully determined, *i.e.* when this distribution is known to be a function of some parameter, although the exact value of the parameter is not known. One wishes to estimate the actual value of the parameter, given an observed outcome of the chance set up. For example, one may know that the statistical probability of a particular coin's

coming up heads when tossed is some fixed number, although one does not know exactly what the number is. The problem is to estimate the statistical probability after observing the outcome of several tosses of the coin.

Notice that such estimation is not the inference that the statistical probability has exactly its estimated value but rather that it has approximately this value. If the inference is taken to warrant detachment, strictly speaking one must specify an acceptable range surrounding the estimated value. In the example one would like to find two numbers such that one can infer that the statistical probability of heads lies somewhere between these numbers. This inference would not be simply an inference to the truth of an explanation. It would be the inference that the explanation is one of several. Thus at first one knows that the observed sequence of heads and tails is correctly explained as the outcome of the binomial distribution of chances, which is a function of the statistical probability of this coin's coming up heads on a single toss; and one infers that this statistical probability lies within a particular specified range.

Although in estimation one does not infer the truth of an explanation, one's inference can be viewed as an instance of a more general form of the inference to the best explanation. Moreover, it must satisfy a generalized maximum likelihood condition. This generalized condition requires that, if one wishes to infer that the actual distribution of chances is one in which the parameter satisfies a particular condition, then the observed actual outcome must have a greater statistical probability in each distribution that satisfies this condition than in any that does not.

Generalizations of all the preceding remarks on the inference to the best explanation are as follows: Let $c(h,e)$ be the inductive probability (degree of confirmation) of h on the evidence e. Let $p(x,y)$ be the statistical probability of getting the outcome x in the chance set up y. Let e be the total evidence. Let $D(x)$ be a statistical distribution of chances as a function of a parameter x. Suppose it is known that some $D(x)$ is actual, *i.e.* suppose that e logically entails $(Ex)D(x)$. Let H be the hypothesis that the correct explanation is a $D(x)$ that satisfies a certain condition $f(x)$: *i.e.* let H be $(Ex) \cdot D(x)\&f(x)$. Then we may directly[11] infer H only if (but not "if and only if") there is an e_H such that

[11] In inductive inference one may directly infer by one step of inference with detachment only a conclusion that a particular explanation is true or that the true explanation is one of a specified set. Of course other kinds of conclusions may be reached indirectly by combining induction and deduction.

(1) e logically entails e_H.
(2) Any D(x) if true would explain e_H.
(3) $c(H,e) \gg c(\text{not-}H,e)$. (The Bayesian condition.)[12]
(4) (x)(y): If f(x)&not-f(y), then $p(e_H, D(x)) > p(e_H, D(y))$.

The last of these conditions represents the generalized maximum likelihood condition.

Let us see how this works in the example. Our observed outcome, e_H, is a sequence of heads and tails. We know this is to be explained as the outcome of a distribution D(x), which is the binomial distribution as a function of x, the statistical probability of heads. We infer that x lies between two numbers y and z, i.e. f(x) is $y < x < z$. The generalized maximum likelihood condition (4) tells us that the likelihood of our getting the observed outcome must be greater if the statistical probability of heads lies at any point in the selected range than if it lies at any point outside that range.

Defense of the generalized maximum likelihood condition rests on two important considerations. First is the basic idea behind any maximum likelihood condition, namely that evidence cannot count toward the truth of an explanatory hypothesis directly inferred from the evidence unless the truth of that hypothesis would have made a relevant difference to our having obtained the evidence we in fact obtained. I shall return to this idea in a moment. A second and possibly more important consideration favoring the generalized maximum likelihood condition (4) is that it allows the present theory to escape a difficulty raised by Kyburg.[13]

Kyburg's problem is a version of the lottery paradox. It is illustrated by the example in which one wants to discover the statistical probability x of a given coin's coming up heads on a single toss. Kyburg shows that conditions (1)–(3) are not sufficient for detachment in this case, *i.e.* that high probability is not sufficient for detachment even when an inference is to a set of explanatory hypothesis.[14] For suppose that y and z are such that it is extremely probable that $y < x < z$. Imagine that we have been generous in

[12] The symbol "$\gg$" means "is much greater than." Let us suppose that $a \gg b$ only if $a/b > 20$. To some extent it is arbitrary how much greater a must be than b. In the present paper I am concerned with quite different issues.

[13] *Op. cit.*

[14] And given that one accepts the usual principles of deduction. As noted above, Kyburg's response to this difficulty is to surrender the deductive principle that warrants the inference from two statements to their conjunction.

selecting y and z so that the interval from y to z is more than enough to satisfy the Bayesian condition (3). Then, if (1)–(3) were sufficient, we could infer that x is in this interval. Furthermore for some N we may divide the interval from y to z into N segments such that for any one of these, the hypothesis that x is in the other N-1 segments will satisfy (1)–(3). We could then infer for each segment of the interval that x is not in it, although we have also inferred that x is in one of these segments. Therefore, if we were to assume that (1)–(3) are sufficient for detachment, we would be led into contradiction.

Condition (4) permits us to meet Kyburg's difficulty since it prevents the inference for an arbitrary segment of the interval between y and z that x is not in that segment. The hypothesis that x is not in some arbitrary segment will fail to satisfy the generalized maximum likelihood condition; and this provides additional support for the condition.

I have said that the idea behind the maximum likelihood condition is that evidence ought to make a relevant difference to the possibility of making one's inference. The same idea supports the following quasi maximum likelihood condition: one can infer h only if our having gotten the evidence explained by h was much more likely if h is true than if it is false. Suppose we want to know whether Sam's father is back in town. Our total evidence makes it extremely likely that he is, although there is a slight chance that he has not yet returned from a business trip. This alone cannot give us knowledge that Sam's father is definitely back in town and is therefore not enough for inference with detachment. But suppose we also know that Sam has just received enough money to pay his rent, that we know that if Sam's father should be back in town he would have given Sam this money, and that it is extremely unlikely that Sam could have received this money if his father were not in town. On the basis of this additional evidence we can come to know that Sam's father is back in town by inferring that he gave Sam the money. Such evidence does support an inference to the best explanation.

On the other hand, suppose there had been another way Sam might have received the money. Suppose that we had known that Sam's mother would definitely have given Sam the money if his father had not. Then the fact that Sam has the money would not have made a relevant difference and we could not have legitimately inferred that Sam's father is back in town. In the actual case we

can infer this, in the imagined case we could not. Both cases involve an inference to the most probable explanation. The difference between them is that only the former satisfies the quasi maximum likelihood condition. In the former case it is much more probable that Sam would have the rent money if his father gave it to him than if his father did not give it to him. This would not be true in the latter case where Sam would have the rent money in either event. I say "much more probable" here because it is not enough to say "more probable." Even if it is only highly probable that Sam's mother would have given him the money if his father had not, we still cannot legitimately use the fact that Sam has the money to infer that his father gave it to him, even though his getting the money would be slightly more probable if his father had given it to him than if his father had not.

The following is not an adequate way to specify the quasi maximum likelihood condition: "$c(e_h, h) \gg c(e_h, \text{not-}h)$." It is inadequate because it would not rule out the latter case. The probability is high that Sam would have received the money even if his father should not have given it to him, but only because of what we know (i.e. what our evidence tells us) about his mother. Nor can we say, *e.g.*, "$c(e_h, e\&h) \gg c(e_h, e\&\text{not-}h)$," which takes into account our total evidence. Since our total evidence e logically entails e_h, the suggested condition would never be satisfied. On the other hand we cannot say "$p(e_h, h) \gg p(e_h, \text{not-}h)$," because it is unclear that such statistical probabilities are defined here or, if they are defined, that we could know what they are in the former of the two cases (in which detachment is definitely possible).

Apparently the requirement must be stated with a subjunctive or independent of fact conditional. Let "r, ifsub q" represent "if q should be true, then r would be true." As before let e be the total evidence. Then one can infer a single explanatory hypothesis h only if there is an e_h such that

(1′) e entails e_h.
(2′) h if true would explain e_h.
(3′) $c(h,e) \gg c(\text{not-}h,e)$.

These conditions resemble conditions (1)–(3) of the generalized inference to the best explanation. Corresponding to the generalized maximum likelihood condition:

(5′) $c((e_h, \text{if sub } h), e) \gg c((e_h, \text{ifsub not-}h), e)$.

The same condition applies to generalized inference to the best explanation as follows:

$$(5)\ c((e_H, \text{ifsub } H), e) \gg c((e_H, \text{ifsub not-}H), e).$$

To see why (5) is needed consider the following case. Suppose that we know that the statistical probability x of a given coin's coming up heads on a single toss is greater than 0.5; and suppose we want to determine more precisely what x is. I submit that if we toss the coin and it comes up heads, we can make no inference for we have as yet learned nothing that would definitely rule out one of the prior possibilities. (5) automatically prevents any inference in this case, since the right hand side of (5) must be at least 0.5 and the left hand side can be no greater than 1.0.[15] Without (5), conditions (1)–(4) would sometimes permit an inference in a case like the present one. Suppose that prior inductive probabilities are as follows: The probability that x is exactly 1.0 (i.e. that we have a two headed coin) is 0.999. For arbitrary y and z, where $0.5 < y < z < 1.0$, the probability that $y < x < z$ is $.002(z - y)$. If we toss the coin once and it comes up heads, conditions (1)–(4) are satisfied by the hypothesis that x is arbitrarily close to 1.0. Without condition (5) we should be able to infer the truth of any such hypothesis, which is absurd. So (5) is a necessary condition on generalized inference to the best explanation.

In this paper I have discussed some of the conditions on detachment in inductive and statistical inference. I have not presented a set of sufficient conditions, indeed a version of the lottery paradox is not met by the set of conditions I have mentioned. For it is possible that there are N different explanations, each accounting for a different fraction of the total evidence, each satisfying the conditions so far mentioned, even though the evidence also ensures that one of these explanations cannot be correct. I do not know what further conditions must be added to give a complete account of inductive inference. My claim in this paper has been that, although any successful account of inductive inference must go beyond the considerations I have advanced, it must include them.

[15] Cf. note 12 above.