6

Evidence, High Probability, and Belief

I propose to argue that potential and veridical evidence require probability greater than $\frac{1}{2}$. "Probability" here will be construed as objective epistemic probability. The requirement that the probability exceed $\frac{1}{2}$ (which will be taken to be a necessary but not a sufficient condition for evidence) allows us to avoid various counterexamples introduced in chapter 4 against the positive relevance (increase-in-probability) account. More to the point, I will argue that a probability greater than $\frac{1}{2}$ is required by the fundamental idea underlying both potential and veridical evidence, viz. that of providing a good reason for belief (as well as by the fundamental idea underlying ES-evidence, viz. providing a justification for belief). In the course of defending this claim, I will examine several theories of evidence and belief, including two that deny any relationship between evidence and belief (or high probability), one that denies the usefulness of the concept of belief, insisting on replacing it with degrees of belief, and one that defines belief as high subjective probability.

1. The Counterexamples and High Probability

On the positive relevance account, *e*'s increasing *h*'s probability is sufficient to make *e* evidence that *h*. In the first lottery counterexample of chapter 4 the positive relevance definition counts the fact that (*e*) all the tickets except for those of John and Bill were destroyed as evidence that (*h*) Bill won, despite the fact that Bill had only 1 of the remaining tickets and John had 100. If we insist that a hypothesis must have probability greater than $\frac{1}{2}$ on the putative evidence then this counterexample is avoided. In the present case the probability assigned to *h* on *e* is $\frac{1}{101}$.

In the swimming counterexample the positive relevance definition counts the fact that (*e*) Steve was doing training laps in the water on Wednesday as evidence that (*h*) he drowned, despite the fact that he is a member of the Olympic swimming team who was in fine shape Wednesday morning. If the hypothesis *h* must have probability greater than $\frac{1}{2}$ on the putative evidence *e*, then, in this example, as in the previous one, *e* will not be evidence that *h*, and the counterexample is avoided.

According to the positive relevance account, not only is *e*'s increasing *h*'s probability sufficient for evidence, it is also necessary. The second lottery and intervening cause counterexamples present cases in which, intuitively, *e* is evidence that *h* but *e* does not increase *h*'s probability. In the second lottery case, given the background *b* (that the lottery is a fair one in which one ticket drawn at random will win) and given the information e_1 (that *The New York Times* reports that Bill Clinton owns all but one of the 1000 lottery tickets sold), the information e_2 (that *The Washington Post* reports the same thing) is evidence that *h* (Bill Clinton will win)—or so I claimed. The positive relevance account disallows this claim, since e_2 does not increase *h*'s probability over what it is on e_1 and *b* alone. In the intervening cause counterexample involving the medicine's relieving the symptoms exactly the same thing occurs: something that intuitively looks like evidence that a hypothesis is true is precluded from that category since it fails to increase the probability of the hypothesis. However, these two counterexamples may be avoided if we relinquish the idea that evidence must increase a hypothesis' probability in favor of the idea that the hypothesis must have high probability.

2. A Good Reason for Belief Requires Probability Greater Than One-Half

My argument for this claim makes two assumptions:

1. For any hypothesis *h* and putative evidence *e* there is some number *k* greater than or equal to zero such that if *e* is a good reason to believe *h*, then $p(h/e) > k$.

The concept of "a good reason for belief" employed here is one involved in potential evidence (where *e* can be a good reason to believe *h* even though *h* is false). The concept is objective, in the sense that whether *e* is a good reason to believe *h* does not depend on whether anyone in fact knows or believes *e*, *h*, or that *e* is a good reason to believe *h*. Therefore, the idea of probability in this assumption is objective. Since it is, and since assumption 1 is concerned with a good reason for belief, the concept of probability employed in assumption 1 (and throughout this chapter, except where otherwise indicated) will be the objective epistemic one of chapter 5, viz. degree of reasonableness of belief.

The assumption above is weak in two respects. First, it does not require (or disallow) that *k* be the same for all hypotheses. Second, the assumption places no restriction on *k*. (As far as this assumption is concerned, *k* could be zero.) The strength of the assumption is to tie "good reason to believe" to probability. For *e* to be a good reason to believe *h*, it is required that there be some probability of *h* on *e*, and that this probability be greater than 0, the lowest probability possible.

(The present assumption, although weak, is not without its critics. I will consider one of the most challenging in sections 12 and 13.)

2. For any e and h, if e is a good reason to believe h, then e cannot be a good reason to believe the negation of h ($-h$).

This is the central assumption of the argument for the claim that e is a good reason to believe h only if $p(h/e) > \frac{1}{2}$. I will return to assumption 2 after presenting the argument.

Suppose that e is a good reason to believe h, so that, in accordance with assumption 1, $p(h/e) > k$, for some k greater than or equal to 0. Suppose $k < \frac{1}{2}$. Then it could be the case that $p(-h/e) > k$. If so, $-h$ and e would satisfy a necessary condition for e's being a good reason to believe $-h$. So unless other necessary conditions for "good reason to believe" are imposed that preclude the possibilities imagined, for some e and h if $k < \frac{1}{2}$ it could be the case that e is a good reason to believe h and also a good reason to believe $-h$. This violates assumption 2.

Now consider the "other necessary conditions" that might be imposed on "good reason to believe" to form a set of sufficient, as well as necessary, conditions. If this set requires that k be greater than or equal to $\frac{1}{2}$, then a good reason to believe requires a probability greater than $\frac{1}{2}$. By contrast, suppose this set of conditions allows something to be a good reason to believe even when k is less than $\frac{1}{2}$. Then for some e and h, where $p(h/e) > k$ and $p(-h/e) > k$, it could be that e is a good reason to believe h and to believe $-h$, in violation of assumption 2. If k is greater than or equal to $\frac{1}{2}$, then such a possibility is precluded. If k is less than $\frac{1}{2}$, it is not precluded.

Providing a good reason to believe is a concept underlying both potential and veridical evidence. Accordingly, by the argument of this section, each of the latter requires probability greater than $\frac{1}{2}$. What I have said can be applied *mutatis mutandis* to the concept of providing a justification for belief, which underlies ES-evidence.

In the sections that follow I will be considering objections to this argument, and, more generally, to the claim that e is a good reason to believe h only if $p(h/e) > \frac{1}{2}$. Before doing so, however, one must be very clear that the claim I am defending provides only a necessary and not a sufficient condition for being a good reason for belief. It is not part of the claim that any probability greater than $\frac{1}{2}$ suffices for such a reason. We may want to require that the probability be significantly greater than $\frac{1}{2}$. And we may want to say that whatever the probability of h on e is, this by itself is not sufficient for e's being a good reason to believe h. Both issues will be addressed in chapter 7.

3. First Objection

One might be tempted to reject the previous argument, and, in particular, assumption 2 on which it is based, by invoking the following example. Suppose medicine M relieves symptoms S in half the cases and in the other half makes them worse. Then, it might be said, the fact that (e) Sam is taking M to relieve symptoms S is a good reason to believe (h) that his symptoms will be relieved, since M relieves S in half the cases. And it is a good reason to believe ($-h$) that his symptoms won't be relieved (but made worse), since that is what happens in half the cases. If so, e is a good reason to believe h and a good reason to believe $-h$, in violation of assumption 2.

My response is to admit that in such a case e is *just as good* a reason for believing h as for believing $-h$. But the fact that two things are equally good does not by itself make either of them good. (Two equally good essays can both be quite bad.) Suppose that Sam does not take medicine M to relieve his symptoms S. And suppose that 90% of those with S who do not take M (or any other medicine) get worse, while 10% get better. Then, taking M is clearly the better course of action than not taking M. Taking M has a better chance of relieving S (50%) than does not taking M (10%). But this is not enough to make the fact that Sam is taking M a good reason for believing that his symptoms will be relieved.

The claim that some e is a good reason for believing h is usually an empirical claim. It can be defended or criticized on empirical grounds. For example, the claim that

R: The fact that Sam is taking medicine M to relieve symptoms S is a good reason to believe that Sam's symptoms S will be relieved

might be defended by appeal to the following empirical association (here I change the percentages from the ones originally given):

A1: Symptoms S are relieved by M in 90% of the cases, while without M there is relief of S in only 5% of the cases.

Similarly, the claim R can be criticized and indeed falsified by showing A1 to be false and the following to be true:

A2: Symptoms S are relieved by M in 1% of the cases, while without M there is relief in 5% of the cases.

More generally, the weaker the association between taking M and relief of S the less one can expect relief of S from M, and the less good is the reason for believing S will be relieved by M.

One way to criticize a good-reason-to-believe claim is to show that the association is such that it provides as good a reason to believe h as it does to believe $-h$. Suppose you claim that the fact that this fair coin will be tossed once is a good reason to believe it will land heads. I am *criticizing* your claim by saying

> Not really, since this fact provides an equally good reason for believing it will land tails.

I am saying that the fair coin's being tossed is not a good reason to believe it will land heads *because* it is an equally good reason to believe that it won't.

Similarly in the symptom case, suppose that A1 and A2 are both false, but the following is true:

A3: Symptoms S are relieved by M in 50% of the cases, and made worse in the other 50%, while without M there is relief only in 5% of the cases.

I am criticizing claim R above by saying that, in view of A3, the fact that Sam is taking M to relieve S provides an equally good reason to believe that Sam's symptoms will not be relieved, but indeed will be made worse. I am saying that because of this, the fact that Sam is taking *M* is not really a good reason to believe his symptoms will be relieved. This is true despite the fact that Sam's chances for relief are considerably improved by taking *M*, just as the coin's chances for landing heads are considerably improved by tossing it. But because the facts leading to such improvements are equally good reasons for believing that the outcomes in question will not occur, they do not provide a good reason for believing they will occur.

Now we need to consider some other fundamental objections.

4. Belief and Degrees of Belief

It might be held that there are no beliefs as such, only degrees of belief. This is the view of some Bayesians, who seek to define evidence (and related concepts) entirely by reference to probability. Since the concepts of evidence I have distinguished involve the idea of providing a good reason (or justification) for believing, they presuppose that there are beliefs, not simply degrees of belief. If there are only degrees of belief, not beliefs as such, the objective concepts of evidence, or at least my characterizations of them, seem flawed.[1]

On the Bayesian view, belief is like temperature. Both come in degrees. Anything that has a temperature has some degree of temperature. This view has two versions. On the stronger one, noted above, there are no beliefs at all, only degrees of belief. The assertion that one believes a proposition is either false or lacking in truth-value. The assertion that one believes it to some particular degree has a truth-value. On the weaker version there are beliefs but these are to be understood in terms of degrees of belief. One might take the state of affairs of *X*'s believing *h* to be identical with the state of affairs of *X*'s believing *h* to a degree greater than some specified number *k*.[2] Or one might take the state of affairs of *X*'s believing *h* to be identical with the state of affairs of *X*'s believing *h* to the particular degree *X* does, provided that this is greater than *k*.

Despite appearances, neither the stronger nor the weaker version of the degree-of-belief view is particularly damaging to the idea that evidence provides a good reason (or justification) for belief. Bayesians understand degrees of belief as probabilities. So, the view that evidence provides a good reason to believe can be understood as saying that evidence furnishes a high probability. The relationship between evidence and belief can be expressed in a way that will satisfy even the more militant Bayesians who deny the existence of beliefs in favor of degrees of belief.

1. These claims about belief refer not to the content of belief (the proposition believed) but to the state or act of believing. It is the latter, rather than the former, that is being claimed to be subject to degrees.
2. This version seems to be one offered by Daniel Hunter, "On the Relation between Categorical and Probabilistic Belief," *Nous*, 30 (1996), 75–98. On Hunter's view, "to believe a proposition is simply to assign that proposition a high subjective probability, where what counts as sufficiently high probability is vague" (p. 87).

One argument brought against both versions of the Bayesian view is empirical. It is that if degrees of belief exist, but not beliefs, or if beliefs are simply sufficiently high degrees of belief, we would be, and need to be, overwhelmed with information that we could not process. But in fact we are not overwhelmed. One defender of this argument is Richard Foley, who writes:

> If all of the information provided to us by others were finely qualified with respect to the provider's degree of confidence in it, we would soon be overwhelmed. It is no different with our private deliberations. We normally don't have finely qualified degrees of confidence in a wide variety of propositions . . . but even if we did, we would soon find ourselves overwhelmed if we tried to deliberate about complicated issues on the basis of them.[3]

I will not pursue this here. In this book I will assume that there are states of believing and that these are not identical with having degrees of belief. If they are not, how are they related? Believing is something we do. In this respect it is like speaking, hitting, or running (even though obviously in other respects it is different). These doings admit of degrees, indeed along one or more dimensions. One can speak with different degrees of speed and loudness. One can hit with different degrees of frequency and intensity. And one can believe with different degrees of firmness or strength. The fundamental idea is the doing, the dimensions represent ways it can be done, and the degrees represent how much is done along a given dimension. So the fundamental idea is speaking, which can be qualified along the dimension of loudness, which admits of varying degrees. With whatever degree of loudness one is speaking, one is still speaking. Similarly, believing is the fundamental idea; it can be qualified along the dimension of firmness or strength, which admits of degrees. With whatever degree of strength or firmness one believes, one is still believing.[4]

It might be objected that analogous claims can be made for temperature, which comes in degrees. Although "having a temperature" is not to "do" anything, one might say that it is the fundamental idea. Whatever degree of temperature something has, it has a temperature. Nevertheless, having a temperature is always having a temperature of some degree. Indeed, a standard dictionary offers this definition of temperature:

> The degree of hotness or coldness of anything, usually as measured on a thermometer.[5]

In the kinetic theory of gases temperature is understood as a quantity proportional to the mean translational kinetic energy of the molecules. Even if this is not taken as a definition, nevertheless temperature is by definition a quantity that admits of degrees (where this quantity is proportional to the mean molecular kinetic energy). Now, just as having a temperature is by definition the same

3. Richard Foley, "The Epistemology of Belief and the Epistemology of Degrees of Belief," *American Philosophical Quarterly*, 29 (1992), 111–124; quote on p. 122. See Gilbert Harman, *Change in View* (Cambridge, MA: MIT Press, 1986), ch. 3.
4. See Isaac Levi, *Gambling with Truth*, p.122.
5. *Webster's New World Dictionary*.

as having a temperature of some degree, so, it may be said, believing is by definition believing to some degree.

My response is to say that the analogy has not been drawn correctly. To be sure, since temperature is by definition a quantity admitting of degrees, having a temperature is by definition having a temperature of some degree. The analogy with belief should be this. By definition believing is believing something (a proposition, a person). Believing, by definition, requires an object of belief. Necessarily one cannot simply believe, just as necessarily an object cannot have a temperature without having some specific degree of temperature.

Proponents of the original analogy may now argue that believing is not *by definition* believing to some degree, but just that this is always so, even necessarily so. Even if it is not true by definition that one cannot believe without believing to some degree, it is nevertheless true (just as one cannot speak without speaking with some degree of loudness). So, suppose that someone who believes h does so to the degree r. Then, it might be claimed, the state of affairs of his believing h is identical to the state of affairs of his believing h to the degree r (just as the state of affairs of a gas's having a certain temperature is identical to the state of affairs of its having the corresponding mean molecular kinetic energy). This identity thesis suffices for a degree-of-belief view (so long perhaps as the degree of belief exceeds some threshold k).

This version of the degree-of-belief view I would reject by invoking the causal principle that if states of affairs are identical, they have the same causes and effects. Even if it is true that if you believe h you must do so to some degree or other, your believing h and your believing h to the degree you do (where this exceeds k) can have different causes and effects. Therefore, they are not identical. For example, suppose that you believe very strongly (say to the degree .99) that O. J. Simpson is guilty of murder, and that your believing it to this degree causes me to be annoyed. What causes me to be annoyed, however, is not your believing it (something that I too believe), but your believing it to the degree you do (how can you be so damn sure?). But if your believing that O. J. Simpson is guilty is the same state as your believing this to the degree you do, then, by the causal principle, they must have the same effects. Since they do not, these states are not identical.

Accordingly, a degree-of-belief theorist must either reject beliefs altogether in favor of degrees of belief or hold that the state of affairs of believing some h is identical with the state of affairs of believing h to a degree greater than some specified k. As I said earlier, however, neither of these views requires abandoning the idea that evidence provides a good reason for belief. The latter will simply be understood as a high probability.

5. Must Evidence Supply a Good Reason for Belief?

Suppose it is granted that a good reason for belief requires high objective epistemic probability (greater than $\frac{1}{2}$) on that reason. Let us confine our attention to potential and veridical evidence, both of which provide reasons for belief (rather than to ES-evidence and subjective evidence which do not). Why must (potential and veridical) evidence provide a *good* reason for belief? Won't it suffice for such evidence to provide just *some* reason to believe a hypothesis, but not necessarily a good one, or perhaps a *better* reason for belief than without it, but one that is not necessarily a good reason?

Suppose, according to the background information b, there is a coin that is precariously balanced on its edge. The new information e is that this coin is being subjected to a force. Let h be the hypothesis that this coin will soon be lying heads up. The fact that the precariously balanced coin is being subjected to a force is, I think, *some* reason to believe it will soon be lying heads up, and also *some* reason to believe it will soon be lying tails up and hence not heads up. It is not, however, a *good* reason for believing either, since (as I argued in sections 2 and 3) if something is a good reason to believe h it cannot also be a good reason to believe $-h$. One way of criticizing the claim that e is a good reason to believe h is to show that it is an equally good reason to believe $-h$.

However, suppose we were to drop the "good reason to believe" requirement. Can we do so in favor of a "some reason to believe" requirement? Can we say that since the fact that this precariously balanced coin is being subjected to a force is some reason to believe it will soon be lying heads up and some reason to believe it will soon be lying tails up, this fact is potential evidence that each claim is true?[6] I reject this idea because of what I will call the "canceling" effect to which evidence is subject.

Potential (as well as veridical) evidence claims of the form "e is evidence that h" are usually empirically incomplete. They can be falsified or at least have doubt cast upon them by additional empirical information. So if you claim that the fact that this precariously balanced coin is being subjected to a force is evidence that it will land heads, I can cast doubt on your claim by pointing out the empirical fact that this coin has two sides, one of which is not heads, and that the force is being applied so as to make the coin spin randomly, so that there is no more reason to expect it to land heads rather than tails. If this is so, then your evidence claim has been seriously weakened. Has it been canceled altogether? Has it been falsified?

I think it has. In this respect potential evidence is like net force in physics. A book sitting on the table is being subjected to two forces. There is the force of gravity directed toward the center of the earth; and there is the force exerted by the table on the book in the opposite direction. The net force in this case is zero, since the two independent forces are equal and opposite. The net force is determined by the balance of all the forces. Whether some particular force acting on a body is the net force depends upon whether there are other forces acting on it. The claim that this force is the net force is defeated by the existence of other forces. In this respect evidence is like net force. Some fact might pull in one direction, while another might pull equally in the opposite direction and prevent the first from being (potential or veridical) evidence. Or (unlike the case of force) some fact might pull equally in opposite directions, with a net pull in neither. (This is so with our precariously balanced coin.)

6. Obviously it cannot be veridical evidence, since this requires the hypothesis to be true; and both hypotheses cannot be true.

It might be objected that in many situations (including scientific ones) the evidence pulls equally in opposite directions, and it is still evidence. For example, suppose that the effectiveness of a new drug is being tested with regard to its capacity to reduce certain symptoms. Two trials are run. In the first there are 2000 patients with the symptoms, 1000 of which are given the drug and 1000 a placebo. Of those taking the drug, 950 have their symptoms relieved, while none in the control group does. Call this information e_1. The second trial is run just like the first, except that the result is quite different (call it e_2): of the 1000 patients taking the drug only 50 have their symptoms relieved, while none in the control group does. Consider the hypothesis h that the drug is more than 50% effective. The results of the two trials pull equally in opposite directions with respect to h. Yet we might say that the combined result $e_1 \& e_2$ constitutes some evidence that h and some evidence that $-h$.

My response is that what is being claimed is that part of the combined result, viz. e_1, is evidence that h, and a different part, viz. e_2, is evidence that $-h$. The concept of evidence here is ES-evidence. Where the epistemic situation does not contain knowledge of the result e_2, e_1 can be ES-evidence that h (analogously for e_2 with respect to $-h$). Someone in an epistemic situation in which result e_2 is unknown might be justified in believing h on the basis of e_1, just as someone in an epistemic situation in which result e_1 is unknown might be justified in believing $-h$. But where we are not separating and relativizing evidence in this way, the combined results pulling in opposite directions cancel. The combined results are not evidence that h is true, and they are not evidence that $-h$ is true. They are puzzling and inconclusive.

A somewhat similar analysis applies to the case of the precariously balanced coin. Where h is that the coin will soon be lying heads up, and e is that a force is being applied to the coin, e can be evidence that h and also evidence that $-h$, if we mean ES-evidence and if we have in mind different ES's for each claim. One ES includes knowledge of the fact that one side of the coin shows heads, but includes no knowledge about the other side. The other ES includes the analogous situation involving tails. Someone who knew that at least one side of the coin is marked heads, and who was not in a position to know that the two sides of the coin are different, might be justified in believing that the coin will soon be lying heads up; similarly, for someone else with this knowledge concerning tails. But where ES includes the fact that the two sides of the coin are marked differently, then, with respect to such an ES, one would not be justified in believing the coin will soon be lying heads up. And the fact that a force will be exerted on the precariously balanced coin is not ES-evidence that the coin will soon be heads up, nor is it ES-evidence that the coin will soon be tails up. It is evidence that the coin will soon be *either* heads up *or* tails up.

6. How Scientists Deal with Conflicting Studies

Multiple studies with conflicting results frequently occur in the sciences. Bemoaning this fact, here is how one pair of authors describe the situation in social science research:

> The sequence of events that led to this state of affairs has been much the same in one research area after another. First, there is initial optimism about using social science research to answer socially important questions that arise. . . . Next, several studies on the question are conducted, but the results are conflicting. There is some disappointment that the question has not been answered, but policy makers—and people in general—are still optimistic. They, along with the researchers, conclude that more research is needed to identify the interactions (moderators) that have caused the conflicting findings. . . .
>
> In the third phase, a large number of research studies are funded and conducted to test these moderator hypotheses. When they are completed, there is now a large body of studies, but instead of being resolved, the number of conflicts *increases*. The moderator hypotheses from the initial studies are not borne out. No one can make much sense out of the conflicting findings. Researchers conclude that the phenomenon that was selected for study in this particular case has turned out to be hopelessly complex, and turn to the investigation of another question, hoping that this time the question will turn out to be more tractable.[7]

What these authors advocate, and what their book expounds, is *meta-analysis*, a method or set of methods for combining the results of disparate and even conflicting studies to determine whether new statistical information can be gleaned from these studies that will support or disconfirm a hypothesis of interest. Even before the term "meta-analysis" was introduced in 1976 by Gene V Glass, other investigators had decried conflicting results of studies in various areas. In 1971 Light and Smith, who introduced a precursor to meta-analysis which they called the "cluster approach," quoted the following passage from a 1966 U.S. Department of Health, Education, and Welfare survey regarding the benefits of ability-grouping in schools:

> Many studies throughout the years have compared academic progress of children grouped according to ability with progress made when grouped heterogeneously. Conclusive and definite anwers to questions commonly asked are difficult to get. Some studies show gains favoring ability grouping, some favoring heterogeneous grouping. Others show little or no significant difference between procedures used. The evidence against or in favor of ability grouping remains vague in spite of a rather persistent belief that learning problems would be greatly alleviated if children on similar levels of ability or achievement could be grouped together for instructional purposes.[8]

In a recent book on meta-analysis for the more general reader, Morton Hunt documents the problem of conflicting studies in many different fields of science.[9] These include several examples of medical studies yielding conflicting conclusions. For instance,

7. John E. Hunter and Frank L. Schmidt, *Methods of Meta-Analysis* (Newbury Park, CA: Sage Publications, 1990), p. 36.
8. Quoted in Richard J. Light and Paul V. Smith, "Accumulating Evidence: Procedures for Resolving Contradictions among Different Research Studies," *Harvard Educational Review*, 41 (1971), 429–471. Quote on p. 438.
9. Morton Hunt, *How Science Takes Stock* (New York: Russell Sage Foundation, 1997); see pp. 2–4.

Twenty-one studies of the use of fluorouracil against advanced colon cancer all find it beneficial, but findings of its effectiveness vary so widely—from a high of 85% to a low of 8%—as to be meaningless and useless to clinicians. (p. 2)

When there are studies with conflicting results what do researchers in the field generally do besides try to conduct new studies or give up and tackle a new problem? Several strategies are noted by the authors cited above:

1. Write a review article that summarizes the different studies and results, without attempting to resolve the issue.
2. Choose a single, favorite study from the set and agree with its conclusions.
3. Compute overall averages for relevant statistics across the entire set of studies, independently of the sizes of the sample in each study or the conditions under which the samples were taken.
4. Take a vote. If a majority of the studies favor one conclusion, then that is the conclusion supported by the studies.
5. Employ meta-analysis, which its proponents regard as a much more sophisticated and reliable set of methods than 3 and 4 above.

When different studies yield conflicting results strategies 2–5 may generate the claim that, despite these conflicting results, the evidence, considered as a whole, does support (or disconfirm) a certain hypothesis. But it need not generate such a claim at all. Various conflicting studies may seem equally good, so no favorite emerges using strategy 2. Overall averages for the studies computed using strategy 3 may not favor a hypothesis over its negation. The studies may be evenly divided with respect to a given conclusion, so the vote in strategy 4 results in no selection. And even a sophisticated meta-analysis of strategy 5 may not yield a conclusion favoring one hypothesis over another. Indeed the use of meta-analyses is controversial.[10] Different meta-analyses by different analysts have scored the same studies differently, resulting in different conclusions. And the results of a meta-analysis of a fairly large set of small trials may conflict with the results of a very large controlled trial.[11]

Accordingly, even strategies 2–5, which can lead to some resolution, need not do so. If not, scientists are in the same boat as Hunter and Schmidt described earlier: "No one can make much sense out of the conflicting findings." They cancel. There is (ES-)evidence that h is true (the results of some of the studies). There is (ES-)evidence that h is false (results of other studies). But considered as a whole, the studies are not both evidence that h is true and evidence that $-h$ is true. They are not evidence that either hypothesis is true, despite the fact that they contain parts that are (ES-)evidence that each is true. That is why the government report, quoted by Light and Smith, concerning ability-grouping in school, claims that while "some studies show gains favoring ability grouping,

10. A summary of criticisms is contained in a review of Hunt, *How Science Takes Stock*, by John C. Bailar III, *Science*, 227 (1997), July 25, 1997, 528–529. See also Bailar, "The Practice of Meta-Analysis," *Journal of Clinical Epidemiology*, 48 (1995), 149–157.
11. See, for example, Jacques LeLorier et al., "Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials," *The New England Journal of Medicine*, August 21, 1997, 536–542.

some favoring heterogeneous grouping . . . the evidence against or in favor of ability grouping remains vague." That is also why Hunter and Schmidt say that in such cases of conflict "researchers conclude that the phenomenon that was selected for study . . . has turned out to be hopelessly complex, and turn to the investigation of another question, hoping that this time the question will turn out to be more tractable."

From the discussion in this and the previous section, I conclude that if e is potential or veridical evidence that h, then e provides a good reason to believe h, not simply some reason. And, I have argued in section 2, e provides a good reason to believe h only if h's (objective epistemic) probability on e is greater than $-h$'s probability on e, that is, only if h's probability on e is greater than $\frac{1}{2}$. So if e is potential or veridical evidence that h, then h's probability on e is greater than $\frac{1}{2}$. In the sections that follow I propose to consider two new objections to these conclusions. The first divorces evidence from belief; the second divorces belief from probability.

7. The Likelihood View of Evidence

On this view, evidence is unrelated to belief, or at least, it is not related in any of the ways I have outlined. The idea has been defended by Ian Hacking,[12] A. W. F. Edwards,[13] and most recently, Richard Royall, a biostatistician.[14] I will consider the latter in what follows.

Royall defines evidence by formulating what he calls the

> *Law of likelihood:* If hypothesis A implies that the probability that a random variable X takes the value x is $p_A(x)$, while hypothesis B implies that the probability is $p_B(x)$, then the observation $X = x$ is evidence supporting A over B if and only if $p_A(x) > p_B(x)$, and the likelihood ratio $p_A(x)/p_B(x)$ measures the strength of that evidence.[15]

An example will be helpful, and I will introduce one used by Royall himself. There is a diagnostic test for a disease D which is such that when D is present the test detects it 95% of the time and gives an erroneous reading 5% of the time. When D is absent the test correctly yields a negative result 98% of the time and incorrectly yields a false positive 2% of the time. The probabilities, then, are represented as follows:

$p(+/D) = .95$ (the probability that the test will yield a positive result given the presence of D is .95)

$p(-/D) = .05$

$p(-/not\text{-}D) = .98$

$p(+/not\text{-}D) = .02$

12. Ian Hacking, *Logic of Statistical Inference* (Cambridge: Cambridge University Press, 1965).
13. A. W. F. Edwards, *Likelihood* (Baltimore: Johns Hopkins Press, 1992).
14. Richard Royall, *Statistical Evidence* (London: Chapman and Hall, 1997).
15. Royall, p. 3. Although he does not provide an interpretation for probability, Royall is talking about some concept of objective probability.

Now we consider two hypotheses regarding a certain patient.

A: The patient has D
B: The patient does not have D

On Royall's view, hypothesis A implies that the probability of a positive test result for this patient is .95. We can write this as $p(+/A) = .95$, or $p_A(+) = .95$. Similarly, we can write $p(+/B) = .02$, or $p_B(+) = .02$, for the probability of a positive test result on hypothesis B. $p(+/A)$ is called the *likelihood* of hypothesis A on the test result +. The *random variable* X in this case is the test result, which can take two possible values: positive (+) and negative (−). Let us suppose that our patient is given the test, and that it yields a positive result (+). This result is what Royall denotes by the lower case variable x. So, in this example, "the observation $X = x$" is understood to mean that the test yields a positive result. I shall refer to this outcome simply as e.

The question of interest to Royall is whether e is evidence supporting hypothesis A (that the patient has disease D) over hypothesis B (that the patient does not have D). On his view, it is if and only if

(1) $p(e/A) > p(e/B)$,

that is, if and only if the likelihood of A on e is greater than that of B on e.[16] In the present case,

$p(e/A) = p(+/D) = .95$
$p(e/B) = p(+/not\text{-}D) = .02$.

So (1) is satisfied, and e is evidence supporting A over B. The strength of the evidence is given by the ratio $p(e/A)/p(e/B) = .95/.02 = 47.5$.

Royall makes a number of claims about this view, which help to clarify it and highlight its advantages.

1. One advantage he claims is that it is a plausible extension of reasoning in deterministic cases to probabilistic ones. He writes:

> One favorable point is that it seems to be the natural extension, to probabilistic phenomena, of scientists' established forms of reasoning in deterministic situations. If A implies that under specified conditions x will be observed, while B implies that under the same conditions something else, *not x*, will be observed, and if those conditions are created and x is seen, then this observation is evidence supporting A versus B. This is the law of likelihood in the extreme case of $p_A(x) = 1$ and $p_B(x) = 0$. The law simply extends this way of reasoning to say that if x is more probable under hypothesis A than under B, then the occurrence of x is evidence supporting A over B, and the strength of that evidence is determined by how much greater the probability is under A. This seems both objective and fair—the hypothesis that assigned the greater probability to the observation did the better job of predicting what actually happened, so it is better supported by that observation. (p. 5)

16. As mentioned in footnote 15, Royall is speaking of some type of objective probability, although he does not specify any particular one. In what follows, in order to contrast his view of evidence with the one I am developing for potential evidence, I will construe his probabilities as objective epistemic ones. However, the criticisms I develop will be applicable no matter what objective interpretation is given for his probabilities.

2. A second advantage claimed by Royall is that, unlike the positive relevance definition of evidence, the likelihood definition does not require the determination of a prior probability for a hypothesis h, which Royall regards as "generally unknown and/or personal." Obviously, then, Royall is regarding the likelihood probability $p(e/h)$, and therefore evidence itself, as objective rather than subjective. Immediately following his rejection of prior probabilities as generally unknown and/or personal, he writes:

> Although you and I agree (on the basis of the law of likelihood) that given evidence supports A over B, and C over both A and B, we might disagree about whether it is evidence supporting A (on the basis of the law of changing probability) purely on the basis of our different judgements of the a priori [prior] probabilities of A, B, and C. (p. 11)

3. Royall emphasizes that the definition is comparative. It is intended to explicate the idea that some set of observations is evidence for one hypothesis over another. It

> represents a concept of evidence that is essentially relative, one that does not apply to a single hypothesis, taken alone. Thus it explains how observations should be interpreted as evidence for A vis-a-vis B, but it makes no mention of how those observations should be interpreted as evidence in relation to A alone. (p. 8)

So from the fact that e is evidence for h_1 over h_2, or that e supports h_1 more than h_2, we cannot conclude that e is evidence for h_1, or that e supports h_1.

4. Finally, let me mention a claim made by Edwards, another defender of the likelihood account. He writes:

> It is sometimes objected that the [likelihood] measure of support does not have any "meaning," by which is usually meant any "probability interpretation." It is indeed true that a statement of support, though derived from probabilities, does not make any assertion about the probability of a hypothesis being correct. And for good reason: the [likelihood] method of support has been developed by people who explicitly deny that any such statement is generally meaningful in the context of a statistical hypothesis. . . .
>
> There is, however, a perfectly simple "operational interpretation" of a likelihood ratio for two hypotheses on some data. It is, of course, the ratio of the frequencies with which, in the long run, the two hypotheses will deliver the observed data.[17]

What Edwards seems to have in mind by the last statement can be illustrated by means of the following example. We have two weighted coins. One is weighted so that in the long run 60% of the tosses yield heads. The other is weighted so as to yield 40% heads. We cannot tell which coin is which simply by looking, only by tossing. We choose one of the two coins at random and consider these hypotheses:

h_1: This coin has the 40% weight.
h_2: This coin has the 60% weight.

17. Edwards, *Likelihood*, p. 33.

Now we toss the coin 10 times, resulting in the following outcome:

e: The coin landed heads 3 times.

The likelihoods are these:

$p(e/h_1) = .2150$
$p(e/h_2) = .0425$

The likelihood ratio $p(e/h_1)/p(e/h_2) = .2150/.0425$, which is approximately 5. So on the likelihood account, *e* supports h_1 over h_2, and the strength of the evidence, as measured by the likelihood ratio, is approximately 5.

Now what Edwards is claiming is that the "operational interpretation" of this likelihood ratio can be given as a long-run relative frequency: if we continue to toss this coin indefinitely, then on the hypothesis h_1 (that the coin is the 40% coin) the relative frequency of sequences of 10 tosses in which 3 yield heads is 5 times greater than it would be on the hypothesis h_2 (that the coin is the 60% coin). More generally, if the likelihood ratio $p(e/h_1)/p(e/h_2) = r > 1$, then if we continue to run the test as we have been doing, we will obtain results like *e*, the results we have actually obtained, *r* times more if hypothesis h_1 is true than if h_2 is true.

8. Likelihood and Belief

Is there a likelihood concept of evidence? Do scientists and others actually employ such a concept? If not, should they?

To begin with, Royall explicitly acknowledges that the concept of evidence he introduces is to be distinguished from any concept that concerns whether, or to what degree, one should believe the hypothesis on the basis of the evidence. Suppose one has made an observation. Royall distinguishes what he regards as three separate questions:

1. What do I believe, now that I have this observation?[18]
2. What should I do, now that I have this observation?
3. What does this observation tell me about *A* versus *B*? (How should I interpret this observation as evidence regarding *A* versus *B*?) (p. 4)

An answer to the third question, he makes clear, does not provide an answer to the first two.

Indeed, assuming, as I have been doing, that if *e* supplies a good reason to believe *h* then *p*(*h*/*e*) must be high ($> \frac{1}{2}$), we can readily see why the likelihood definition does not guarantee that evidence supplies a good reason for belief.[19] On the likelihood definition, *e* can be evidence for hypothesis *A* over hypothesis *B* even if *p*(*A*/*e*) is extremely low. Suppose, for example, that the presence of a certain gene *G*1 reduces the chance of getting a certain disease *D* from .05 to .02, while the presence of gene *G*2 reduces it from .05 to .01. That is, *p*(*D*) = .05, *p*(*D*/*G*1) = .02, and *p*(*D*/*G*2) = .01. And suppose that in the general population gene *G*1 is on the average present in one out of 1000 people, while *G*2 is present in one out of 100 people. That is, *p*(*G*1) = .001 and *p*(*G*2) = .01. Now, using Bayes' theorem, we have

$$p(G1/D) = \frac{p(G1) \times p(D/G1)}{p(D)} = \frac{.001 \times .02}{.05} = .0004$$

$$p(G2/D) = \frac{p(G2) \times p(D/G2)}{p(D)} = \frac{.01 \times .01}{.05} = .002$$

Since *p*(*D*/*G*1) > *p*(*D*/*G*2), on the likelihood definition, the fact that a certain patient has disease *D* is evidence for the hypothesis that the patient has gene *G*1 over the hypothesis that he has *G*2. (The likelihood ratio is 2.) Yet in this case, the probability that a patient has gene *G*1 given that he has disease *D*, *p*(*G*1/*D*), is very low (.0004). If *p*(*G*1/*D*) must be greater than $\frac{1}{2}$ for the presence of disease *D* to be a good reason to believe the patient has gene *G*1, the fact that the patient has *D* in this case is not a good reason for such a belief. Yet on the likelihood view, this fact is supporting evidence that the patient has *G*1 over the hypothesis that the patient has *G*2. As noted, likelihood evidence for one hypothesis over another need not provide a good reason to believe the better supported hypothesis.

If likelihood evidence does not necessarily provide a *good* reason to believe a hypothesis, does it provide at least *some* reason? Since Royall distinguishes his concept of evidence for *h* from a concept of whether, or to what degree, to believe *h*, the answer would seem to be no. This can be confirmed by reference once again to probability. If *e* is at least *some* reason to believe *h* (though not necessarily a good reason), then, we might suppose, there is some threshold of probability (say, less than $\frac{1}{2}$) that *h*'s probability on *e* must exceed. For example, in section 5, in the case of the coin precariously balanced on its edge, I claimed that the fact that it is being subjected to a force is not a good reason to believe it will soon be lying heads up. Nevertheless, I claimed it was *some* reason to believe this (and equally some reason to believe it will soon be lying tails up). The probability we assign to each of these outcomes is $\frac{1}{2}$. Without assuming that for *e* to provide some reason to believe *h*, *p*(*h*/*e*) must be this high, let us suppose that it must exceed some threshold value $k > 0$. Whatever that value we can devise a case in which it is not exceeded, but *e* is likelihood evidence for h_1 over h_2.

For example, suppose the probability threshold value for "some reason" is .01. That is, for *e* to be at least some reason to believe *h*, the probability of *h* on *e* must exceed .01. Yet in our gene example, *p*(*G*1/*D*) = .0004, which is only 4% of this threshold value. If .01 is the threshold value, then the fact that the patient has disease *D* is not a reason to believe he has gene *G*1, even though, on the likelihood definition, the fact that the patient has *D* is evidence for his having *G*1 over his having *G*2. Could we say at least that if *e* is likelihood evidence for h_1 over h_2, then *e* is a better reason for believing h_1 than h_2, or a reason for believing h_1 more than h_2? This also seems precluded. If *e* is a better reason for believing h_1

18. In the context of his discussion, it seems pretty clear that Royall intends this question to be interpreted as "What should I believe?"
19. This is perfectly consistent with construing Royall's probabilities as objective epistemic ones measuring degrees of reasonableness of belief. Royall's concept of evidence, understood in terms of such probabilities, requires the degree of reasonableness of believing *e* given *h* to be greater than the degree of reasonableness of believing *e* given *h*′. It does not require the degree of reasonableness of believing *h* given *e* to be greater than $\frac{1}{2}$, or than *h*′ given *e*.

than h_2, or a reason to believe h_1 more than h_2, then h_1 should be more probable on e than is h_2. Yet in the gene example, $p(G1/D) < p(G2/D)$. But since $p(D/G1) > p(D/G2)$, D is likelihood evidence for $G1$ over $G2$.

Indeed the presence of disease D lowers the probability that a person has gene $G1$ (from .001 to .0004). Yet it is likelihood evidence that a patient has gene $G1$ over gene $G2$. (The presence of D also lowers the probability that a person has gene $G2$, from .01 to .002.)

Accordingly, in this example, the presence of disease D is likelihood evidence for the hypothesis that the patient has $G1$ over the hypothesis that the patient has gene $G2$, even though

1. $p(G1/D) < p(G2/D)$, that is, even though the probability that the patient has $G1$ is less (not more) on D than the probability that the patient has $G2$.
2. $p(G1/D) = .0004$, that is, the probability that the patient has $G1$, given that he has D, is *very* low.
3. $p(G1/D) < p(G1)$, that is, the probability that the patient has $G1$ is lowered by his having D.

It is for these reasons that Royall wants to divorce support from belief.

If likelihood evidence does not provide a good reason, or indeed some reason, to believe a hypothesis, or a better reason to believe one hypothesis than another, what does it provide? What of value do we get when we get likelihood evidence? Let us grant that we get something that does not require a determination of prior probabilities, and that we get something for which objective probabilities can be used. Of what value is the product? What can we do with it, if it does not provide a basis for believing a proposition or (on Royall's view) a basis for action? Why should we be interested in such a concept? Why is this a concept of *support* or *evidence?*

In its defense, we recall, Royall claims that the likelihood concept of evidence is a natural extension of what he calls "deterministic" reasoning. He construes the latter as involving two hypotheses h_1 and h_2, where h_1 implies that e is the case and h_2 implies that e is not the case, and e turns out to be the case. If so, he claims, e is evidence supporting h_1 over h_2 (p. 5). There is, however, an important difference between the deterministic case and the nondeterministic ones of interest to Royall. In the former, if h_1 implies e, and h_2 implies $-e$, then $p(e/h_1) = 1$ and $p(e/h_2) = 0$. In the deterministic cases, unlike the nondeterministic ones, it is not possible that $p(h_1/e) < p(h_2/e)$, since the latter is equal to zero. Unless $p(h_1/e) = 0$, in the deterministic cases we are guaranteed that $p(h_1/e)$ will always be greater than $p(h_2/e)$. Let us assume that if $p(h_1/e) > p(h_2/e)$, then, given e, it is more reasonable to believe h_1 than h_2, or at least it is less unreasonable to believe h_1 than h_2. In the deterministic cases, then, that is, in cases in which $p(e/h_1) = 1$ and $p(e/h_2) = 0$, there will be at least some relationship between Royall's likelihood evidence and belief. But in nondeterministic cases there is no such guarantee, since even though $p(e/h_1) > p(e/h_2)$, in such cases it is possible that $p(h_1/e) < p(h_2/e)$ (and that $p(h_1/e)$ is very small, and that $p(h_1/e) < p(h_1)$). In short, even if the deterministic cases have some (minimal) connection with belief, no such connection is guaranteed for the nondeterministic ones. The question remains, then, why in the nondeterministic cases (which are, after all, the ones Royall is

concerned with) we should be interested in his likelihood concept of evidence. What makes it a concept of evidence?

Another claim made by Royall is that since on the likelihood definition "the hypothesis that assigned the greater probability to the observation did the better job of predicting what actually happened, so it is better supported by that observation" (p. 5). (Edwards makes a related claim about the continued predictive sucesses of the hypotheses.) This reply, I believe, begs the question. What justification is there for saying that of two hypotheses, the one that assigned the greater probability to the event that actually occurred is the *better supported* by that event? Such a justification seems necessary since we cannot conclude that the event provides a good, or any, reason for believing the preferred hypothesis, or that it provides a better reason for believing the preferred hypothesis than for believing its competitor.

Ian Hacking makes a claim that might provide some justification for taking likelihood to be a measure of support. After formulating a principle of likelihood which is essentially the one propounded by Royall,[20] Hacking writes:

> This I venture as the explication of the thesis, "if p implies that something happens which happens rarely, while q implies that something happens which happens less rarely, then, lacking other information, q is better supported than p."[21]

I want to focus here on Hacking's phrase "lacking other information." The idea is that if the likelihood of h_1 on e is greater than that of h_2 on e, then, *in the absence of any other information*, e supports h_1 better than h_2. Perhaps the thought is that if one is in a situation where one can determine only whether the likelihood of h_1 on e, that is, $p(e/h_1)$, is greater than that of h_2 on e, that is, $p(e/h_2)$, and where one cannot compare prior probabilities, $p(h_1)$ and $p(h_2)$, and where one cannot compare posterior probabilities, $p(h_1/e)$ and $p(h_2/e)$, then at least one can judge support based on likelihoods.

If this is the idea, then again the question is begged, since it needs to be shown that in the absence of other information about these probabilities, judgments about support can be made at all. By analogy, suppose we want to determine whether one ship on the ocean is closer to the Baltimore harbor than another. Suppose that latitudes can readily be determined, and we can say that one ship is on a latitude closer to that of Baltimore than the other ship. But suppose that (as was historically the case until the invention of a gyroscopic clock) longitudes are very difficult to determine. It could be disasterous to say that, in the absence of longitudinal information, one ship is closer to Baltimore than the other because its latitude is closer to that of Baltimore than the other. Both latitude and longitude are needed to determine which is closer. What Hacking needs to demonstrate is that there is some intuitive sense of support for which information regarding prior and posterior probabilities is not needed. He needs to show that in that sense of support, in the absence of such information, the hypothesis with the higher likelihood is the better supported.

20. Indeed, Royall cites Hacking when he formulates his principle.
21. Hacking, *Logic of Statistical Inference*, p. 65.

9. Does Evidence Require Prior or Posterior Probabilities? Mayo's Error-Statistical Account

Deborah Mayo develops a novel probabilistic approach to evidence that forcefully rejects the positive relevance and high probability accounts (as well as a likelihood view of the sort just discussed).[22] She rejects positive relevance and high probability because both require a determination of the posterior probability $p(h/e)$, and, in addition, the former requires a determination of a prior probability $p(h)$. In general, she believes, it is not possible or necessary to assign probabilities to hypotheses, whether these are prior or posterior probabilities. I will not discuss Mayo's rejection of these standard views, but only her own positive view, which she calls an error-statistical approach.

On her view, to have evidence for a hypothesis, in a sense that is important to science, is to have something that passes a good test for that hypothesis:

> Following a practice common to testing approaches, I identify "having good evidence (or just having evidence) for *H*" and "having a good test of *H*." That is, to ask whether *e* counts as good evidence for *H*, in the present account is to ask whether *H* has passed a good test with *e*. (p. 179)

She divides this idea into two parts: (a) passing a test, and (b) the test's being a good one. The first idea she describes by saying that the test result *e* must "fit" the hypothesis. Her criterion of "fit" is rather open-ended and contextual. She does not require that the putative evidence *e* be entailed by the hypothesis *h*, or even that *e* be highly probable on *h*. Minimally, she says, *e* must not be improbable on *h* (p. 179).

It is the second idea, that of a good test, that is the most important in her account. For this purpose she proposes a

> *Severity requirement:* Passing a test *T* (with [result] *e*) counts as a good test of or good evidence for *H* just to the extent that *H* fits *e* and *T* is a *severe test* of *H* (p. 180).

The notion of a severe test in this requirement is to be understood by reference to what she calls a severity criterion, for which she offers two formulations:

> *Severity criterion 1:* There is a very high probability that test procedure *T* would not yield such a passing result [*e*] if *H* is false. (Alternatively, there is a very low probability that the test procedure *T* would yield such a passing result, if *H* is false.)

> *Severity criterion 2:* There is a very high probability that the test procedure *T* would yield a worse fit if *H* is false. (Alternatively, there is a very low probability that test procedure *T* would yield so good a fit if *H* is false.)

22. Deborah Mayo, *Error and the Growth of Experimental Knowledge* (Chicago: University of Chicago Press, 1996). I am very much indebted to her for her patience and impatience in explaining her very interesting but complex views to me. Her patience allowed me to correct mistakes and improve my formulations; her impatience motivated me to do so. In any case, induction tells me that she will not completely sanction what I say here.

Mayo's criteria for "evidence" (or a "good test") invoke a concept of probability. However, unlike the two standard probability accounts, which she rejects, her account does not appeal to a posterior probability for hypothesis *h*, that is, to $p(h/e)$, or to a prior probability for *h*, that is, to $p(h)$. The only probabilities she invokes are these:

(a) the probability that the test *T* will yield the putative evidence *e*, given the hypothesis *h*. I will write this as $p(e(T)/h)$. This probability is not supposed to be low, thus satisfying her minimal condition for *e* "fitting" *h*.

(b) the probability that the test *T* will yield the putative evidence *e*, given that the hypothesis is false, that is, $p(e(T)/-h)$. This probability is supposed to be low, thus satisfying her severity requirement.

There is no requirement, however, that the following probabilities be determined—$p(h/e(T))$ or $p(h)$—or that the former be high or that it be higher than the latter, in order that $e(T)$ be evidence that (or a good test of) *h*. Finally, the probabilities that Mayo invokes in (a) and (b) are objective, not subjective. Her examples employ probabilities construed in the relative frequency sense, and it is clear that these are the kinds of probabilities she has in mind for her concept of evidence.

Let's consider an example. There is a very large set of bags, all of which contain many balls. Each bag in the set is one of two types. In the first type, 60% of the balls are white and 40% are red. In the second type, 40% are white and 60% are red. We will proceed with the following experiment. We will choose one bag at random and select a sample of balls from it, with replacement, determining how many are white. (We will assume that the probability of selecting a white ball is the same each time.) We consider two hypotheses. The first, *h*, is that the bag selected contains 40% white balls. The second, *h'*, is that it contains 60% white balls.[23]

The particular test *T* we run involves selecting 100 balls, with replacement, from the chosen bag. We need a test rule for deciding whether a result passes the test *T* with respect to the hypotheses *h* and *h'*. Here is one:

> *Test Rule:* In 100 selections of balls from the chosen bag, if the observed proportion of white balls is less than .5 (that is, if less than 50 out of 100 are white), then the resulting selection passes the test *T* for hypothesis *h* as against hypothesis *h'*.

Now we run test *T*. We select 100 balls, with replacement, and obtain the result $e(T)$ that 40 are white. Since 40 of the 100 are white, the hypothesis *h* (that the bag selected is 40% white) "fits" this result (as precisely as you like). The question is how severe the test is. According to Mayo's severity criterion 1, there must be a very high probability that our test procedure would not yield this pass-

23. Mayo prefers to state hypotheses of these sorts to be making an assertion about the value of a parameter *p* in a binomial distribution. So if we consider drawing balls from the chosen bag, *p* represents the percentage of white balls drawn in *n* trials. The hypotheses of concern to her would then be *h*: $p = .4$, *h'*: $p = .6$, in *n* trials (for sufficiently large *n*). In what follows I simplify by letting *h* and *h'* be the hypotheses given in the text.

ing result $e(T)$ if hypothesis h were not true but h' were. The probability that our test procedure would yield this passing result if h were false and h' were true,

p (we obtain this passing result for h (40 white out of 100)/h' is true),

is .00002. So the probability that our test procedure would *not* yield this passing result if h were not true but h' were is 1 − .00002 = .99998, which is indeed a very high probability. Accordingly, Mayo's severity requirement is satisfied. On her view, then, the result $e(T)$ of the test (40 of the 100 balls selected are white) is good evidence for the hypothesis h.

Finally, we note that determining whether this test result $e(T)$ is evidence that h does not involve determining $p(h)$, the prior probability of h, or $p(h/e(T))$, the posterior probability of h given the test result $e(T)$.

10. Does Mayo's Account Give Us What We Want?

Is the result $e(T)$ described in the previous section really a good test of, and hence evidence for, hypothesis h, as Mayo's account implies? I suggest that it depends on whether certain additional conditions obtain. In the example we stipulated that the bag in question is chosen at random from a very large set of bags. Suppose that half of these bags are ones in which 40% of the balls are white and 60% are red, while the other half are ones in which 60% are white and 40% are red. If this is so, then, where a bag is selected at random, the probability of selecting a bag in which 40% of the balls are white is $\frac{1}{2}$, that is,

(1) $p(h) = \frac{1}{2}$.

Now, the probability of obtaining 40 white balls in a sample of 100, $e(T)$, on the assumption that h is true, is

(2) $p(e(T)/h) = .08$.

And the probability of obtaining the result $e(T)$, assuming that h is false (and that h' is true) is

(3) $p(e(T)/-h) = .00002$.[24]

According to Bayes' theorem,

$$p(h/e(T)) = \frac{p(h) x p(e(T)/h)}{p(h) x p(e(T)/h) + p(-h) x p(e(T)/-h)}.$$

From (1), (2), and (3) above, using Bayes' theorem, the posterior probability

$$p(h/e(T)) = \frac{.5\ x\ .08}{(.5\ x\ .08) + (.5\ x\ .00002)}$$

24. Following Mayo, probabilities in (1)–(3) can be construed in a relative frequency sense. To simplify notation, however, I use h both for the type of event "selecting a bag in which 40% of the balls are white" (e.g., in (1)) and for the hypothesis "the bag selected has 40% white balls" (e.g., in (2)). Similar interpretations of h can be given for Mayo's "parameter" idea in note 23.

is approximately 1, and $p(-h/e(T))$, which is equal to $p(h'/e(T))$, is approximately 0. Note that in this case, since $p(h/e(T)) > .5$, obtaining the result $e(T)$ comports with my high-probability condition on evidence. So far so good.

By contrast, let us consider a very different scenario. Suppose that instead of half of the bags being 40% white, only 1 in 100,000 are, so that the probability of selecting a bag in which 40% of the balls are white is .00001, that is,

(4) $p(h) = .00001$.

Using this value in Bayes' theorem, and keeping the other values the same, we obtain

(5) $p(h/e(T)) = .004$ and $p(h'/e(T)) = .996$.

Accordingly, the probability that the bag selected is 40% white, *even assuming that in the sample of 100 examined 40 are white*, is extremely low, considerably less than 1%.

In what sense, if any, does this count as *passing a good test of, or evidence for, h?* This raises a fundamental question of what Mayo means by these italicized terms. I find the following two different claims about evidence (or "passing a good test") in her work.

(a) Sometimes she claims that "evidence indicates the correctness of hypothesis H, when H passes a severe test" (p. 64). This seems to imply that evidence that H provides a good reason to believe that H is *true*. But if this is what evidence is supposed to do, then, as I argued earlier in this chapter, it will do so only if the probability of the hypothesis, on the putative evidence, is greater than $\frac{1}{2}$. Clearly if (5) is the case, obtaining the experimental result $e(T)$ from test T is not evidence that h, in the sense required.

(b) Sometimes she claims that evidence that H indicates that H is *reliable*, which for her means that "what H says about certain experimental results will often be close to the results actually produced—that H will or would often succeed in specified experimental applications" (p. 10).

There are hypotheses that are extremely reliable, with reference to certain types of experimental results, yet very improbable, and indeed false. Consider the following hypothesis about a certain coin:

H = The devil made this coin fair, that is, he made it so that it will land heads approximately half the time.

If we consider experimental results pertaining to how the coin actually lands when tossed, this hypothesis may be very reliable. It may accurately predict what happens in the long run when the coin is tossed. Yet the hypothesis itself is extremely improbable. So, Mayo may say, tossing the coin 1000 times and obtaining heads approximately half the time is evidence for the reliability of the devil hypothesis H in experiments in which the coin is tossed and the side it lands on is noted. It is not evidence for the truth of the devil hypothesis H.

Mayo seems to allow evidential claims of both types (a) and (b). Which one someone is justified in making will depend on the hypothesis and the putative evidence. I do not at all reject claims of type (b) since they simply amount to saying that e is evidence (in a "good reason to believe" sense) not for the truth of H, but for the truth of the claim "*H* is experimentally reliable (in a specified range of experiments)." I would reject the claim that this is the only, or the main, question about evidence of concern to scientists and others. With respect to a hypothesis such as the devil hypothesis *H*, one may seek evidence not only of its "reliability" but of its truth as well. And, if $p(H/e)$ is extremely low, and hence *e* is not a good reason to believe *H*, then *e* is not evidence that *H* is true, even if it is evidence that *H* is "reliable" in experiments in which the coin is tossed and the side it lands on is noted.[25] Moreover, as I will argue in chapter 8, any plausible attempt to define instrumental evidence—evidence that *H* is experimentally reliable—will need to appeal to the probability that *H* is true.

What about my example involving bags of balls? Here Mayo may claim that the evidence obtained (40 of the 100 selected are white) is evidence that hypothesis *h* (the bag selected contains 40% white) is "experimentally reliable" (with respect to experiments that involve selecting balls from the bag). Or she may claim that it is evidence for the truth of *h*. Or she may claim that the truth of *h* in such a case amounts to its experimental reliability. Any of these alternatives results in problems with her account, I believe. The evidence obtained (that 40 of the 100 are white) is not evidence that *h* is true or "experimentally reliable" (with respect to experiments involving selecting balls from the bag). If only 1 in 100,000 of the bags contains 40% white balls, that is, if $p(h) = .00001$, then, as we said, $p(h/e(T)) = .004$. More generally, in the type of case envisaged, the probability is extremely low that the hypothesis *h* is true or "experimentally reliable," no matter what the test results $e(T)$ show. Yet on her view, the specific test result is a good test of, and hence evidence for, hypothesis *h*. On the contrary, I would claim, it is not.

11. Mayo Responses

Mayo considers an example that is analogous in certain respects to my bags-of-balls example. Her response, tailored to my own example, is this.[26] Selecting a sample from the chosen bag, is, or at least can be, a severe test of the hypothesis *h*, that 40% of the balls in the bag are white, the larger the sample the more severe the test. If the bag contains such a large number of balls that counting them all is infeasible, then sampling is surely the best means of detecting error. So, in a suitable sample, if approximately 40% of the balls are white, then obviously we cannot infer that 60% of all the balls in the bag are white. Our sample

25. Complicating the picture is the fact that for the purposes of testing, Mayo wants to divide a hypothesis such as the devil hypothesis *H* into two separate ones, each of which answers a different question. H_1: the devil made this coin; H_2: the coin is fair. These hypotheses would be tested in different ways. Perhaps, then, the way in which each would be tested (for example, by tossing the coin in the case of the second hypothesis but not the first) would be such that passing a severe test would indicate not only the "reliability" of the hypothesis but its "correctness" as well. So it might be the case, on her view, that when testable hypotheses are suitably isolated, then passing a severe test indicates both "reliability" and "correctness."

26. Deborah G. Mayo, "Response to Howson and Laudan," *Philosophy of Science*, 64 (1997), 323–333.

does not indicate that, despite the fact that in the example the probability that the bag has 60% white balls, given that 40% in the sample are white, is .996.[27]

My response to this contains two points.

1. The fact that 40% of the balls in the sample are white ($e(T)$) does not "indicate," in the sense of "is evidence," that 60% of the balls in the bag are white (h'). Although $p(h'/e)$ is .996 (so that my high probability condition is satisfied), my second (explanatory) condition of evidence (to be developed in chapter 7) is not satisfied. Given the truth of h' and *e*, the probability is not high that there is an explanatory connection between the two. For example, it is not highly probable that in the sample 40% of the balls are white *because* in the bag approximately 60% are white.

2. Perhaps what tempts Mayo to conclude that finding 40% white balls in the sample is evidence for the hypothesis that 40% in the bag are white is that there is *conflicting* evidence in this case. The fact that we have a sample from the bag in which 40% are white, *considered by itself*, is, or seems to be, evidence that 40% of the balls in the bag are white. On the other hand, the fact that the bag has been randomly chosen from a large set of bags only 1 in 100,000 of which contain 40% white balls, the remaining being 60% white, is, or seems to be, evidence that 60% of the balls in the bag are white. Earlier we spoke of cases of this sort, some of which, at least, can be understood in terms of ES-evidence. If one is in an epistemic situation in which all one knows is that 40% of a suitable sample of balls in this bag are white, but one has no idea that this bag was randomly selected from a large set of bags only 1 in 100,000 of which contains 40% white balls, then one may be justified in believing that 40% of all balls in the bag are white. And if one is in a situation in which all one knows is that the bag has been randomly selected from a large set of bags, .99999 of which contains 60% white balls, then one may be justified in believing that 60% of the balls in the bag are white.

With both pieces of information, what can we conclude? Consider just potential evidence. Can we conclude that the fact that 40% in the sample are white ($e(T)$) is evidence that the bag is 60% white (h')? No, despite the fact that

27. In Mayo's example, which is derived from one of Colin Howson, a randomly chosen student passes a college admissions test (*e*). There are two hypotheses. H = the student is ready for college (the student is not deficient); H' = the student is not ready for college. Now we suppose that college-ready students pass the test with probability 1 ($p(e/H) = 1$), while college-deficient students pass only 5% of the time ($p(e/H') = .05$). In the general population .999 are college-deficient, $p(H') = .999$, while $p(e) = .051$. Using Bayes' theorem, we calculate that $p(H'/e) = .98$, that is, the probability that the randomly chosen student is college-deficient, given that the student has *passed* the test, is .98, which, according to Mayo, Howson would take to indicate that the student is college-deficient, given that he has passed the test! Mayo responds that "scoring the passing grade *e* hardly shows *a lack* of readiness—and on these grounds, the error statistician denies that *e* indicates *H'*" (p. 328).

If by "indicates" Mayo means "is evidence that," then Howson would reject Mayo's claim. For Howson (who holds a positive relevance view of evidence), *e* is not evidence that H' (the student is college-deficient), but that *H* (the student is college-ready), since $p(H'/e) = .98$ and $p(H') = .999$, so $p(H'/e) < p(H')$. On the other hand, $p(H/e) = 1 - p(H'/e) = .02$, $p(H) = .001$, so that $p(H/e) > p(H)$, from which it follows, according to Howson, that *e* is evidence that ("confirms") *H*, while *e* "disconfirms" *H'*. So, contrary to Mayo, for Howson "scoring the passing grade" indicates "readiness," not a lack thereof.

$p(h'/e\&b)$ is approximately .996 (where b indicates that only 1 in 100,000 of the bags is 40% white). The reason is the one given in point 1 above. The explanation condition is violated. Can we conclude that the fact that 40% in the sample are white ($e(T)$) is evidence that 40% in the bag are white (h)? No, because, as we saw, $p(h/e(T)\&b)$ is .004. So, on my view, in the light of background information, $e(T)$ is not evidence that h or that h'.

I turn now to a second possible Mayo response.[28] The suggestion is to reformulate the test rule to reflect the fact that the sample of balls selected occurred in two stages: first, a bag was selected randomly from a set of bags with respect to which the probability of selecting a 40% white bag is .00001; second, 100 balls were randomly selected from the chosen bag, of which 40 are white. In view of this two-stage procedure, the test rule, given in section 9, should be reformulated as follows:

> *New test rule:* In 100 selections of balls from a bag chosen at random from a set of bags with respect to which the probability of selecting a 40% white bag is .00001, if the observed proportion of white balls is less than .5, then the resulting selection passes the test T' for hypothesis h as against hypothesis h'.

Hypothesis h is that the bag selected in this way contains 40% white balls; h' is that it contains 60% white balls. Now we run the new test T' by selecting a bag at random from the set in question and selecting 100 balls from the chosen bag. We obtain the result $e(T')$ that 40 of the 100 are white.

Does the new test rule make a difference? Does it prevent the result $e(T')$ from counting as good evidence for the hypothesis h, on Mayo's view? No, it does not. First, the result $e(T')$ passes the new test rule for h as opposed to h', since the observed proportion of white balls (.4) is less than .5. Second, Mayo's severity criterion is satisfied, since the probability of obtaining the result $e(T')$, assuming that h is false, is very low:

$$p(e(T')/-h) = p(e(T)/-h) = .0002.$$

Using Bayes' theorem, and the given assumption that $p(h) = .00001$, we get

$$p(h/e(T')) = .004, \text{ and } p(h'/e(T')) = .996.$$

In short, we get the same results using the new test rule as we do using the original one. Moreover, the new test rule has the prior probability $p(h)$ built into its formulation. This is something Mayo seeks to avoid.

Here is a third possible Mayo response. She may reply in the same manner she does when a hypothesis is constructed to fit the data, where the procedure for constructing it would probably yield a good fit even if the hypothesis is false. She calls this Gellerization. For example, let an experimental result consist in outcomes of tossing a coin, where a heads outcome is a "success" and tails a "failure." We are interested in hypotheses giving the probability of "success" (heads) on each coin toss. Suppose the experimental result e of tossing the coin 4 times is: s,f,f,s (s = success, f = failure). Let the Gellerized hypothesis $G(e)$ be that the

28. This possibility was raised for me by Kent Staley.

probability of success is 1 on trials 1 and 4, and 0 on trials 2 and 3.[29] Then $G(e)$ maximally fits e: $p(e/G(e)) = 1$.

The test T Mayo imagines is this: observe the series of coin tossing outcomes and "find a hypothesis $G(e)$ that makes the result e maximally probable, and then pass that hypothesis" (p. 202). Now, claims Mayo, in determining whether test T is a severe test for some such $G(e)$, you cannot just determine that $p(e/G(e))$ is maximal. You also need to consider "as part of the testing procedure, the particular rule that is used to determine which hypothesis to test" (p. 202). The following severity condition must be satisfied:

> There is a very high probability that the test procedure T would *not* pass the hypothesis it tests, given that the hypothesis is false. (p. 202)

With the test procedure T just noted for a Gellerized hypothesis $G(e)$, the probability that the test procedure would not pass the hypothesis it tests, given that the hypothesis is false, is not high but zero.

The question now is whether the hypothesis h in the previous section ("the bag selected is 40% white") is Gellerized in Mayo's sense, so that the procedure for constructing it would probably yield a good fit with observed data even if the hypothesis is false. It does not seem to be Gellerized. There is no procedure for constructing h analogous to that for the Gellerized coin tossing hypothesis $G(e)$. We are simply told in advance that any bag in the set is either 40% or 60% white. So we are considering each hypothesis in turn. Moreover, if the hypothesis h were false, so that the bag selected was 60% white not 40%, then the probability would not be high that the test T consisting of selecting 100 balls and noting colors would yield the result it did, viz. 40 white ($p(e(T)/-h) = .00002$). So the severity condition just noted is satisfied.

Let me briefly mention a final possible response by Mayo. It is to restrict her theory of evidence to cases in which prior and posterior probabilities of hypotheses are unobtainable.[30] Mayo admits that there are cases in which prior and posterior probabilities, in a frequency sense, can be assigned. She writes:

> *Except* for such contexts, however, the prior probabilities of the hypotheses are problematic. Given that logical probabilities will not do, the only thing left is subjective probabilities. For many [including Mayo], these are unwelcome in scientific inquiry. (p. 82)

On this proposal, Mayo's concept of evidence is applicable when and only when hypothesis h is such that a frequency interpretation of $p(h)$ and $p(h/e)$ cannot reasonably be applied, yet where frequency interpretations can be given to $p(e/h)$ and $p(e/-h)$.

What concept of evidence, then, are we to use when (as in my bags-of-balls example) a frequency interpretation can readily be assigned to $p(h)$ and $p(h/e)$? If we choose one according to which $p(h/e)$ should be high, we obtain a very different concept of evidence from the one proposed by Mayo. In any case, why

29. See Mayo, p. 201, footnote 17.
30. I thank Thomas Hood for this idea.

should our concept of evidence be different when frequency probabilities can be assigned to hypotheses? Why should *e*'s passing a severe test with respect to *h* (in Mayo's sense) count as evidence that *h* when $p(h)$ and $p(h/e)$ cannot be assigned by frequency means but not when they can?

12. Is Belief Related to Probability? Mark Kaplan's View

I turn now to the final objection to the view that for *e* to be potential or veridical evidence that *h*, and hence for *e* to provide a good reason to believe *h*, *h*'s probability on *e* must be greater than $\frac{1}{2}$. The objection derives from a theory that completely divorces belief from probability. I will examine an interesting version due to Mark Kaplan.[31]

On Kaplan's theory, there is no relationship between whether you believe something and your degree of confidence in it. For Kaplan, your degree of confidence is something that satisfies, or ought to satisfy, the probability rules. So on his view you can believe something even though you have very little confidence in it, that is, even though you assign very low probability to it. And you can fail to believe it, even though you have a great deal of confidence in it, that is, even though you assign very high probability to it. If this is right, then if you believe *h* and do so for a good reason, that reason need not be one that you take to confer a high probability on *h*. Kaplan is concerned here with subjective probability. But one who wants to invoke objective probability might extend Kaplan's argument and say that a good reason to believe something, and therefore potential and veridical evidence, need not confer a high degree of (objective) probability.

Let us consider Kaplan's reasons for the first part of this argument, namely, the claim that there is no relationship between your believing something and having some degree of confidence in it. On what he calls the "confidence threshold view" (p. 94), you believe *h* if your degree of confidence in *h* exceeds some threshold value *k*.[32] The confidence threshold view, Kaplan claims, violates a fundamental principle of rationality of beliefs, viz. "deductive cogency," which states that the set of hypotheses you believe includes the consequences of that set but no contradictions. (Among the consequences of a set of propositions is the conjunction of members of the set.) That this principle is violated can be shown by appeal to the lottery paradox (discussed in chapter 4). Suppose there is a fair lottery consisting of 1000 tickets, one of which will win. Suppose your confidence level for belief is .9, so that you believe a proposition *p* if your confidence in *p* exceeds .9. With respect to any particular ticket in the lottery your confidence level that it will not win is, let us suppose, greater than .9. Accordingly, you believe it won't win. Conjoining all these beliefs results in the belief that no ticket will win. But you are extremely confident (>.9) that some ticket will win. Therefore, your beliefs are contradictory.

31. Mark Kaplan, *Decision Theory as Philosophy* (Cambridge: Cambridge University Press, 1996).
32. This is a simplification of the version he considers.

A defender of the confidence-threshold view will not be persuaded, but will reject "deductive cogency." If you believe h_1 and you believe h_2, it does not follow that you believe the conjunction $h_1\&h_2$, since the probability you assign to the latter may be less than the threshold required for beliefs, even if the probabilities you assign to h_1 and h_2 separately exceed this threshold.

Kaplan is aware of this response by confidence-threshold theorists, and he offers a reply. It is that if you reject "deductive cogency" then you must reject *reductio ad absurdum* arguments generally (p. 96). In a *reductio* argument, Kaplan writes, "a critic derives a contradiction from the conjunction of a set of hypotheses which an investigator purports to believe," thereby demonstrating a defect in this set.

Is Kaplan right? Must a confidence-threshold theorist reject *reductio* arguments? In the lottery paradox the confidence-threshold theorist claims to believe each proposition in the set consisting of propositions of the form "ticket *i* will not win." But he does *not* believe the conjunction of propositions in this set. So we have the following situation. The *set* of beliefs which the confidence-threshold theorist believes is inconsistent: the beliefs can't all be true. Does the confidence-threshold theorist believe they are all true? Yes and No. He believes that *each* of them is true, but he does not believe that they are all true together; he does not believe that the *conjunction* of these beliefs is true, even just the conjunction of beliefs of the form "ticket *i* will not win." So even though the set of his beliefs can't all be true, you cannot derive a contradiction of the form "*p* and $-p$" from this set, because the conjunction of beliefs is not a member of the set of his beliefs. You cannot derive the proposition "no ticket will win, and some ticket will win," or even just the first conjunct, from his set of beliefs, unless, that is, you accept the principle of "deductive cogency," which he does not.[33]

This does not commit the confidence-threshold theorist to abandoning *reductio* arguments. In a *reductio* argument the critic takes the investigator to believe not just each proposition in the set of premises but the conjunction of these propositions as well. The critic then derives a contradiction of the form "*p* and $-p$" from that conjunction. In the lottery situation the confidence-threshold theorist does not believe the conjunction of propositions of the form "ticket *i* will not win."

Kaplan has a second reply for confidence-threshold theorists (p. 98). It is that even if we do not require "deductive cogency," it is a *possible* course of action. That is, it is possible to believe in such a way that the set of hypotheses you believe includes the consequences of that set but no contradictions. But, says Kaplan, this is not even possible on the confidence-threshold view. That view necessarily leads to the lottery paradox and the situation in which the set of your beliefs includes a contradiction.

To this a confidence-threshold theorist will respond that if the confidence-threshold view is correct, then Kaplan is just mistaken in his claim that "deduc-

33. Here is an analogy. Consider situations that I am willing to have be the case. These might include my seeing a movie tonight and my not seeing a movie tonight. But it doesn't follow from this that I am willing to have it be the case that I see a movie tonight and that I don't see a movie tonight. Although I am willing to have each of the two states of affairs obtain that cannot obtain together, I am not willing to have both of them obtain together.

tive cogency" is a possible course of action. It is impossible. If confidence beyond a certain threshold entails belief, then necessarily in a lottery situation of the sort in question one will believe each member of a set of propositions not all of which can be true, which is a violation of "deductive cogency." Kaplan must argue independently that confidence beyond a certain threshold does not entail belief.

This leads to Kaplan's own view about belief, to which I now turn.

13. Kaplan's "Assertion View" of Belief

According to Kaplan,

> You count as believing *P* just if, were your sole aim to assert the truth (as it pertains to *P*), and your only options were to assert that *P*, assert that $-P$, or make neither assertion, you would prefer to assert that *P*. (p. 109)

By having the "sole aim to assert the truth" Kaplan makes it clear that he means the truth and only the truth, so that avoiding error is part of this aim. "The truth," he writes

> is just an error-free, comprehensive story of the world: for every hypothesis *h*, it either entails *h* or entails $-h$, and it entails nothing false. (p. 111)

In pursuing the aim of truth,

> the desire for comprehensiveness and the desire to avoid error are bound to conflict. . . . We would expect that different individuals would differ as to how much risk of error they were willing to tolerate in order to satisfy the desire for comprehensiveness. . . . It is compatible with the Assertion View that you tolerate very great risk of error, in which case you might well believe hypotheses in whose truth you have very little confidence. (p. 111)

So this view divorces belief from confidence.

An immediate objection is that this leads to a bizarre consequence. You can believe something and be very confident that it is false. Kaplan admits that this is a consequence of this view.[34] His response is this. The reason it sounds bizarre is that we are operating with the "ordinary notion of belief," which "construes belief as a state of confidence short of certainty" (p. 142). The problem, he says, is that this ordinary notion of belief also "takes consistency of belief to be something that is at least possible, and perhaps even desirable." But he claims to have shown that "no coherent notion of belief can do both" (p. 142). So he is offering the assertion concept of belief as an alternative to the ordinary concept.

Kaplan's new concept of belief will also lead to a bizarre consequence for evidence, if the latter is construed as something that provides a good reason, or a justification, for believing.[35] For, on Kaplan's assertion concept of belief, it is possible that *e* is evidence that *h*, and hence that *e* provides a good reason (or justification) to believe *h*, even if it offers no basis for any confidence in *h*, and indeed even if it offers a basis for confidence that *h* is false. This, I think, departs considerably from any ordinary notions of evidence.

The reason that Kaplan offers his assertion concept of belief is that the ordinary one, which, he claims, is tied to confidence, is "incoherent," since it violates deductive cogency (p. 142). But, I suggest, this charge of "incoherence" does not stick. Rejection of "deductive cogency" by confidence-threshold theorists does not lead to believing contradictions of the form *p and* $-p$. It leads only to believing each member of a set of beliefs not all of which can be true.

Indeed, an inconsistent set of beliefs is possible on Kaplan's own "assertion view." Consider two propositions, *P* and *Q*. Suppose my sole aim is to assert the truth as it pertains to each, and my only options (in each case) are to assert that *P* (*Q*), assert that $-P$ ($-Q$), or make neither assertion. And suppose I prefer to assert that *P* (*Q*). Then, on the assertion view, I believe *P* and I believe *Q*. But this could happen even if *P* and *Q* form an inconsistent set.

Kaplan may reply that the lottery paradox shows that the confidence-threshold view of belief *necessarily* leads to an inconsistent set of beliefs, whereas his own assertion view does not entail that an inconsistent set of beliefs is necessary, only that it is *possible*. Since both views allow for the possibility of such inconsistency, and since Kaplan does not take this as a defect for his assertion view, I'm not sure why the necessary consequence of certain inconsistent sets of belief is particularly damaging for the confidence-threshold view (which Kaplan later calls the ordinary view of belief).

14. Conclusions

In this chapter I have defended the following claims.

1. A high-probability condition on evidence enables us to avoid counterexamples to the positive relevance view.

2. If *e* provides a good reason to believe *h*, then *h*'s objective epistemic probability on *e* must be greater than $\frac{1}{2}$. Moreover, if *e* is potential or veridical evidence that *h*, then *e* must provide a good reason to believe *h*. It does not suffice for *e* to provide just some reason. Conflicting studies may furnish some reason to believe *h* and some reason to believe $-h$, but not a good reason to believe both. Taken separately, they may provide ES-evidence for each hypothesis. But taken together, they cancel and do not provide potential or veridical evidence for both.

3. Neither the strong Bayesian view that there are no beliefs only degrees of belief, nor the weaker one that beliefs are degrees of belief greater than some

34. He writes: "Suppose we ask someone whether hypothesis *h* is true. It is easy to understand how she might reply 'Yes, *h* is true,' or 'No, *h* is not true,' or how she might be unwilling to commit herself either to *h*'s truth or to its falsehood. But it is hard to see what can be made of the responses, 'Yes, *h* is true, but I am extremely confident it is false,' and 'I believe *h* is true, but I'd give you 20 to 1 odds that I'm mistaken.' Yet, not only does the Assertion View countenance these apparently bizarre responses. . . , it deems them entirely reasonable ones to make in the context of inquiry." (p. 142)

35. Kaplan does not construe evidence in this way, but suggests a view according to which *e*, if true, is evidence that *h* for you if you have more confidence in *h* given *e* than you do in *h* without *e*, where you are not certain of the truth-values of *e* or *h* (p. 48). Since confidence for him is a probabilistic notion, Kaplan is suggesting a subjective positive relevance concept of evidence.

number, is incompatible with the idea that evidence provides a good reason to believe a hypothesis, and hence that it confers a probability greater than $\frac{1}{2}$.

4. The view that evidence provides a good reason to believe, and hence furnishes a probability greater than $\frac{1}{2}$, is defended against three views. The first, the likelihood view, completely divorces evidence that *h* from whether, or to what extent, one should believe *h*. So, in effect, does the second, Mayo's error-statistical account, since even "experimental reliability" does not guarantee a good reason to believe the hypothesis, or that its probability is greater than $\frac{1}{2}$. The third, due to Mark Kaplan, completely divorces belief from probability. I argue that the likelihood and error-statistical views fail to provide useful concepts of evidence, notwithstanding arguments to the contrary by their defenders. Also I argue that Kaplan is mistaken when he claims that only a divorce between belief and probability will save *reductio* arguments. Finally, I discuss unintuitive consequences of Kaplan's own assertion view of belief, which allows you to believe something to which you assign very low probability.

7

THE EXPLANATORY CONNECTION

High probability of *h* given *e*, by which is meant that *h*'s probability is greater than that of its negation, that is, greater than $\frac{1}{2}$, is necessary for *e* to be potential (and hence veridical) evidence that *h*. That is what I have argued in earlier chapters. But it is not sufficient. In this chapter I will formulate an additional necessary condition that invokes the idea of a probable explanatory connection between hypothesis and evidence. A concept of explanation that can be used to understand this condition will then be explicated. (In what follows, unless otherwise indicated, when I speak of evidence I mean potential evidence; as in chapter 6, probability is being construed as objective epistemic probability.)

1. High Probability Is Not Sufficient for Evidence

To see why, recall the "irrelevant information counterexample" of chapter 4:

e: Michael Jordan eats Wheaties
b: Michael Jordan is a male basketball star
h: Michael Jordan will not become pregnant

The probability of *h* given *b* is close to 1, and this probability is unchanged if we add information *e* to *b*. That is, $p(h/e\&b)$ is approximately 1. Yet it seems absurd to claim that, given that Michael Jordan is a male basketball star, the fact that he eats Wheaties is evidence that he will not become pregnant. In general, I have been claiming, for any *e*, *b*, and *h*, if *e* is evidence that *h*, given *b*, then, given *b*, *e* provides a good reason to believe *h*. The fact that Michael Jordan is male provides a good reason to believe he will not become pregnant. But given that he is a male basketball star, the fact that Michael Jordan eats Wheaties fails to provide such a reason.

More generally, suppose that the background information *b* itself is, or contains, evidence that *h*. (In our example, *b* contains the information that Michael Jordan is male.) And suppose we append to the background information *b* some item of information *e* that is, evidentially speaking, irrelevant for the hypothesis