

*Bayesian Induction Is Eliminative Induction*¹

James Hawthorne
University of Oklahoma

INTRODUCTION

Eliminative induction is a method for finding the truth by using evidence to eliminate false competitors. It is often characterized as “induction by means of deduction”; the accumulating evidence eliminates false hypotheses by *logically contradicting* them, while the true hypothesis logically entails the evidence, or at least remains logically consistent with it. If enough evidence is available to eliminate all but the most implausible competitors of a hypothesis, then (and only then) will the hypothesis become highly confirmed. I will argue that, with regard to the evaluation of hypotheses, Bayesian inductive inference is essentially a *probabilistic* form of induction by elimination. Bayesian induction is an extension of eliminativism to cases where, rather than *contradict* the evidence, false hypotheses imply that the evidence is very unlikely, much less likely than the evidence would be if some competing hypothesis were true. This is not, I think, how Bayesian induction is usually understood. The recent book by Howson and Urbach,² for example, provides an excellent, comprehensive explanation and defense of the Bayesian approach; but this book scarcely remarks on Bayesian induction’s eliminative nature. Nevertheless, the very essence of Bayesian induction is the refutation of false competitors of a true hypothesis, or so I will argue.

The eliminative nature of Bayesian inference shows up most prominently in Bayesian convergence theorems. These theorems show that, under certain conditions, wide ranging *initial disagreement* among Bayesian agents regarding the plausibility of various hypotheses (as represented by prior probabilities for hypotheses) will eventually be “washed out” by the evidence. Evidence will eventually bring about a *convergence to agreement* on the posterior probabilities of hypotheses, regardless of the values of the prior probabilities. Thus, the influence of the subjective element in Bayesian inference, the subjective assessments of prior probabilities, can be overcome.

These Bayesian convergence theorems almost always rely on an underlying eliminative process. The prior probabilities only become “washed out” as the evidence drives the posterior probabilities of hypotheses to 1 (confirmation) or 0 (refutation). In L. J. Savage’s theorem,³ for example, the eliminative element is obvious. Savage’s theorem shows that if the accumulating evidence is of the right character, and if the true hypothesis is not initially too improbable (i.e., has a non-0 prior probability, prior to the evidence), then false alternative hypotheses will almost certainly become *highly refuted* by the evidence (i.e., obtain posterior probabilities, based on the evidence, arbitrarily close to 0). Hence, the true hypothesis will become *highly confirmed* (i.e., achieve a posterior probability arbitrarily near 1). According to Savage’s theorem, a false hypothesis will become highly refuted when it says that the accumulating evidence is extremely unlikely, much less likely than the true hypothesis would have it. The point of the theorem is to assure us that such refutations are achievable when the evidence is of *sufficiently good quality*, but falls short of logically contradicting false hypotheses.

In an exceptionally penetrating analysis of Bayesian induction, John Earman⁴ devotes a chapter to the examination of a number of Bayesian convergence theorems. Earman argues that the assumptions under which a wide variety of these theorems operate are too strong to apply in many important cases of inductive scientific inference. Thus, he argues, the objective core of Bayesian induction, which the convergence theorems are supposed to underwrite, is not generally assured.

Regarding Savage’s convergence theorem, Earman endorses the important earlier critique of Hesse.⁵ Hesse points out that Savage’s theorem assumes that the evidence consists of a sequence of independent, identically distributed events; it assumes that the likelihoods of observational outcomes are “objectively determined” relative to each theory or hypothesis; and it puts no bounds on the rate at which convergence takes place. Hesse and Earman argue that evidence for scientific theories will not generally satisfy the first two conditions, and that the lack of bounds on the rate of convergence suggests that convergence may take almost forever.

Earman ultimately rejects Bayesian induction (as it is usually understood) and opts for a version of *eliminative induction with Bayesian-like considerations*. He suggests that the version of eliminativism he favors might

be dubbed *Bayesian eliminativism*. I will argue that Bayesian induction is itself a form of induction by elimination, *without* the addition of any extra eliminative element. Thus, I think, Earman is right about the form that scientific inference should take, but Bayesian induction already fills most of the bill.

Earman anticipates the possibility of a purely Bayesian form of eliminative induction. In the introduction to the seventh chapter of his book, his "Plea for Eliminative Induction," he says he will argue that⁶

... much of the bad press eliminative induction has received is unjustified, and that to succeed at the level of scientific theories, Bayesianism must incorporate elements of the eliminative view.

Later, on the same page, he adds:

Moreover, the ability to provide a Bayesian gloss does not mean that Bayesianism has any real explanatory power. Indeed, the eliminative inductivist will see the Bayesian apparatus merely as a tally device to keep track of a more fundamental process.

I will argue that the Bayesian apparatus is indeed a tally device for eliminative induction, a device that keeps a numerical tally of how hypotheses are faring in the eliminative process. But I think there is no shame in this for Bayesian induction; for, as we will see, any reasonable eliminative account of induction should rely on a similar sort of tally device.

I will explore the eliminative nature of Bayesian induction and its connection with Bayesian convergence. I will investigate very general conditions under which Bayesian convergence may be achieved. Some of these conditions are necessary for convergence (e.g., Theorem 4), and others are sufficient (e.g., Theorems 5 and 6).

In section 1, I give a formal characterization of *deductivist* eliminative induction. I show that in contexts where all of the competing hypotheses are deductively related to the evidence, Bayesian induction is equivalent to eliminative induction plus a simple scheme for representing and maintaining plausibility weightings among hypotheses not yet falsified.

In section 2, I investigate probabilistic eliminative induction, induction in cases where hypotheses may fail to logically entail evidence, but instead assign some probability to its occurrence. I will show how probabilistic eliminative induction operates in a Bayesian framework, and establish a fundamental connection between Bayesian convergence and eliminative induction (Theorem 4). Roughly, Bayesian convergence can occur *just in case* the evidence is strong enough to refute competing hypotheses. Differing probability functions, which represent alternative assessments of the initial plausibility to hypotheses, will not generally come into agreement on the posterior probability of a hypothesis *unless* the evidence is either sufficiently strong to refute that hypothesis or else is so powerful as to refute all of its competitors. Thus, Bayesian convergence is essentially an eliminative

process; for the most part, Bayesian convergence amounts to achieving agreement about which hypotheses are practically refuted by the evidence.

In section 3, I will develop a generalized form of Savage's Bayesian convergence theorem. It shows that if a hypothesis differs even slightly from the true hypothesis in the likelihoods that it assigns to possible outcomes of experiments and observations, then a large collection of such evidence will almost certainly suffice to bring about the refutation of the hypothesis. The generalized version of the theorem avoids the main objections raised against Savage's theorem by Earman and Hesse. In particular, this version doesn't depend on independent, identically distributed evidence; it doesn't require the likelihoods to be objective in any objectionable sense; and it puts bounds on the rate at which convergence takes place, bounds that explicitly depend on a quantitative information-theoretic measure of the *quality of the evidence*.

1. THE DEDUCTIVIST MODEL OF ELIMINATIVE INDUCTION

In this section I will develop a formal representation of the deductivist version of eliminative induction. I will show that the eliminative strategy for finding the true hypothesis naturally suggests a tally device, a way of measuring the impact of evidence on hypotheses, that is equivalent to Bayesian conditionalization relative to evidence. This will set the stage for the succeeding sections, where we will see that Bayesian induction provides a generalization of eliminativism for hypotheses that are not deductively related to the evidence.

1.1 THE LOGICAL STRUCTURE OF HYPOTHESIS TESTING

The central logical features of both eliminative and Bayesian induction may be captured in a common framework, in terms of formal structures that I will call *induction structures*. First I will specify the structure of a *pre-structure*. Then I will impose additional conditions that, intuitively, an *induction structure* should satisfy.

A *pre-structure* for an *induction structure* is a quadruple $S = \langle b, H, C, O \rangle$, where, for some formal language L that is adequate for the expression of scientific theories (i.e., at least a first-order language):

' b ' is a sentence of L (intuitively, b expresses relevant, reasonably uncontroversial background knowledge or assumptions);

$H = \langle h_1, h_2, \dots \rangle$ is a finite or denumerable list of sentences of L (each h_i represents a competing hypothesis or theory about some domain or subject matter);

$C = \langle c_1, c_2, \dots \rangle$ is a finite or denumerable list of sentences of L (each c_k is a description of initial conditions, or an experimental setting, or of how an observation is conducted);

$O = \langle O_1, O_2, \dots \rangle$ is a list of *lists of sentences* of L , where O has the same number of members as C ;

each $O_k = \langle o_{k,1}, o_{k,2}, \dots, o_{k,m} \rangle$ in O is a finite list of sentences of L (the number, m , of sentences may differ among the various O_k in O ; intuitively, for each observation condition c_k in C , O_k is a list of the alternative possible experimental or observational outcomes).

Let ' \models ' represent the logical entailment relation appropriate to the language L , and let ' \neg ', ' \cdot ', and ' \vee ' be the symbols in L for negation, conjunction, and disjunction, respectively. A *pre-structure* $S = \langle b, H, C, O \rangle$ is an *induction structure* provided that it satisfies the following conditions:

- 1) for $i \neq j$, $b \models \neg (h_i \cdot h_j)$, i.e., the hypotheses in H are mutually exclusive;
- 2) for any h_j and c_k , and $r \neq s$, $(b \cdot h_j \cdot c_k) \not\models \neg (o_{k,r} \cdot o_{k,s})$, i.e., no two outcomes of c_k can both occur;
- 3) for any h_j and c_k , $(b \cdot h_j \cdot c_k) \models (o_{k,1} \vee \dots \vee o_{k,m})$, i.e., one of the outcomes of each c_k must occur.

The roles that the various components of an *induction structure* (i.e., b , members of H and C , and members of the O_k in O) are intended to play in the inductive confirmation of hypotheses should be pretty clear, but a few words about the formal details are in order.

H is a possibly infinite list of hypotheses that express alternative theories about some common domain of phenomena. The members of H may be broad, powerful scientific theories, or they may be much more specific, limited claims. For some induction structures, the list of hypotheses H may be logically exhaustive relative to the common background knowledge b . That is, $\{b, \neg h_1, \neg h_2, \dots\}$ may be a logically inconsistent set of sentences. However, H need not be so all-inclusive. Hypotheses that are considered too implausible to be worth testing (including those constructed around grue-predicates) may be excluded from H . But the more exclusive H is, the greater the danger of excluding the true hypothesis. I will assume that, however exclusive, H does contain a true hypothesis. Of course, the *logic* of hypothesis confirmation will not depend on which member of H is true. Rather, the logic will specify how *any* given member of H would fare on evidence *if it were true*. If the truth is not one of the possibilities in H , then it is simply out of the running. If the true hypothesis is never considered, then there is no way for evidence to support it, and no way for eliminative induction, Bayesian induction, or any other method to confirm or support it with evidence.

Usually the background knowledge or assumptions represented by 'b' will express relatively uncontroversial claims about the operation of instrumentation, and perhaps bridge principles that give empirical content to some of the theoretical terms that occur in hypotheses in H (e.g., that trails in properly functioning bubble chambers are the tracks of charged particles). In some cases 'b' may contain sophisticated physical theories (e.g., a theory about the geophysics of the impact of large bodies with the earth, and another about the physics of crystals and the force required to create shocked quartz); and the hypotheses in H may make relatively narrow claims that are to be evaluated relative to the presumed truth of these theories (e.g., the hypothesis that a large asteroid struck the earth at the end of the cretaceous period, and that the impact spread debris over a large part of the earth). I will assume that the relevant background knowledge may be expressed by some finite (perhaps very large) collection of sentences, and that all of these sentences are conjoined into a single sentence 'b'.

In *deductivist* versions of eliminative induction b must be of sufficient strength to aid each possible hypothesis in conjunction with the initial conditions c_k (for each k) in the deductive entailment of one of the possible outcomes $o_{k,r}$ in O_k . In the more general context of Bayesian induction, $(b \cdot h_j \cdot c_k)$ need not deductively entail outcomes. Rather, the background knowledge may only aid $(h_j \cdot c_k)$ in the assignment of a probability of occurrence to each possible outcome in O_k . If each hypothesis in H is detailed enough to provide precise deductive or probabilistic implications from the initial condition statements in C without any help from background assumptions, then 'b' can be taken to be some simple tautology. In practice, however, some contingent background knowledge is almost always presupposed in experimental or observational contexts.

One may think of the ordering on initial conditions provided by C as representing the order in which the evidence is gathered, the order in which observations are made or in which experiments are performed. It will be convenient to represent the conjunction of the first n members of C, ' $c_1 \cdot c_2 \cdot \dots \cdot c_n$ ', by ' c^n '.

Each possible outcome of the *sequence of initial conditions* c^n is a conjunction of the possible outcomes of n observations, one from each of O_1, O_2, \dots, O_n . That is, for each k, let ' e_k ' (from O_k) represent one of the possible evidential outcomes of c_k . Then the conjunction ' $(e_1 \cdot e_2 \cdot \dots \cdot e_n)$ ' represents one possible conjunction of outcomes of the first n observation conditions c^n . I will employ the expression ' e^n ' to represent conjunctions of sequences of n outcomes. The expression ' e^n ' is a meta-linguistic variable that ranges over possible sequences of outcomes of c^n for a given induction structure. More picturesquely, each possible value of e^n is a possible "path through the outcome space" of c^n . Let E^n be the set of all such "paths." The sentences in E^n represent each distinct possible sequence of outcomes for the induction structure.

A degree of idealization is necessary for a formal treatment of inductive inference, but such idealizations are common in the theoretical specification of a logic. For example, induction structures may possess an infinite list of “all reasonable” hypotheses (a list that is inclusive enough to contain the true hypothesis). The sentences of L are enumerable, so an enumeration of all hypotheses must *mathematically exist*. But human reasoners will seldom have a complete list of the reasonable alternatives before them. At any given time most of the reasonable alternatives of a scientific theory will not have been contemplated. However, once we see how inductive inference should work for logically ideal reasoners, the concluding section of this paper will address the implications for more limited beings.

In the next subsection I will focus on the deductivist version of induction structures, where each competing hypothesis logically entails one of the possible outcomes of each observation c_k . Here induction structures will provide a formal model of eliminative induction as it is commonly understood.

1.2 THE DEDUCTIVIST MODEL OF ELIMINATIVE INDUCTION

An induction structure S will be called a *deductive induction structure* (more simply, a *deductive structure*) just in case, in addition to conditions 1–3, above, it also satisfies the following condition:

- 4-d) for each sequence of observations c^n (for all values of n)
and for each h_j in H , there is a sequence of possible outcomes e^n in E^n such that $(b \cdot h_j \cdot c^n) \models e^n$.

Notice that for any given hypothesis h_j , the sequence of possible outcomes e^n that is entailed by $(b \cdot h_j \cdot c^n)$ must be a single member of E^n , by condition 2 for *induction structures*. Hence, condition 4-d (‘d’ for ‘deductive’) implies that each hypothesis in the structure logically entails exactly one sequence of possible outcomes, one path through the space of possible outcomes.

Induction by elimination is usually framed in the context of *deductive structures*. Suppose a *deductive structure* S contains the true hypothesis h_t . Will the evidence flush it out? The following theorem implies that it will.

THEOREM 1: *Deductivist Version of Savage’s Convergence Theorem.*

For a *deductive induction structure* S , suppose b and h_t and the c^n are true (for all n). Let h_f be any alternative hypothesis in H that differs with h_t regarding some possible outcome e_k of some observation c_k (i.e., $(b \cdot h_f \cdot c_k) \models e_k$, and for some alternative outcome o (not e_k), $(b \cdot h_f \cdot c_k) \models o$).

Then, for n large enough (so that c^n has c_k as a conjunct), the *actual* sequence of outcomes e^n will be whichever sequence of outcomes is logically entailed by the true hypothesis—the e^n such that $(b \cdot h_t \cdot c^n) \models e^n$. And e^n will falsify h_f , i.e., $(b \cdot c^n \cdot e^n) \models \neg h_f$.

According to this theorem the true hypothesis says that the evidence will *falsify* those competitors that disagree with it on the evidence, and hence eliminate them. The proof of the theorem is completely trivial—it follows from the defining conditions for *deductive induction structures*. I have elevated this result to the status of a theorem only because it is almost an exact analog of Savage's Bayesian convergence theorem. Savage's theorem⁷ extends Theorem 1 to cases in which hypotheses need not *deductively entail* the evidence. Later, in section 3, I will offer a generalized version of Savage's theorem.

In *deductive structures*, whenever any two hypotheses disagree on some possible outcome, the evidence must eliminate at least one of them. Each competitor of the true hypothesis will eventually be falsified, except those that agree with the truth on all possible evidential outcomes (for the deductive structure). If there are *evidentially indistinguishable alternatives* to the true hypothesis, only some sort of plausibility considerations (e.g., ontological economy, explanatory power, comprehensiveness) can offer any hope of ferreting out the truth among them.

Instrumentalists, of course, will be perfectly content if the evidence can pare away all but an empirically equivalent set of hypotheses, provided each hypothesis is *empirically adequate* to cover the intended range of phenomena. Only realists need be concerned with finding the *true hypothesis* among these. My concern in this paper is primarily with what evidence can do to distinguish among hypotheses. Although I will often speak as a realist, I do so because realism has the tougher epistemological challenge to meet. In almost all of what follows (including *all* formal results), whenever I speak of the *true hypothesis* any *empirically adequate hypothesis* can play the same role.

Suppose that (for some structure) the true hypothesis has no evidentially indistinguishable competitors, or that such competitors are laid low by plausibility considerations. If *H* contains only a finite number of hypotheses, then clearly the evidence will eventually (for some finite sequence of observations *cⁿ*) refute all of the false alternative hypotheses, and the truth will be whatever remains. But when the hypotheses are scientific theories, the number of alternative hypotheses is (potentially) infinite. Although a single bit of evidence may falsify an infinite number of the alternatives, there will generally remain an infinite number among the unrefuted, and each successive bit of evidence may leave an infinite number of remaining hypotheses. How can eliminative induction find the truth when there are always so many unrefuted alternatives? It cannot, as long as all unrefuted hypotheses are held on a par, i.e., are considered equally plausible. It seems that eliminative induction cannot uncover the truth, but only toss out an ever increasing number of the false alternatives. Thus eliminative *induction* appears to reduce to Popperian falsificationism.⁸

There is a simple way out of this quandary, a way that takes advantage of the ordering that *H* imposes on hypotheses. Each hypothesis, including the

true one, lies some finite distance from the beginning of the ordering. Suppose h_i is true. If all hypotheses before h_i in H are evidentially distinguishable from it, then the evidence will eventually refute them, and h_i will rise to the top of the list (as ordered by H) of unrefuted hypotheses. Once at the top, h_i will remain there. Indeed, each of h_i 's evidentially distinguishable rivals farther out in H will also be eliminated eventually, and h_i will become preminent. But, although this eliminative strategy must eventually yield the true hypothesis (if it is in H), one can never be sure that the hypothesis presently at the top of the unfalsified list is the truth. Still, the truth will surface eventually.

There is nothing sacred about the ordering H imposes; any ordering of the members of H will do. For *any* enumerative ordering of the hypotheses in H , the true h_i (and all of its evidentially indistinguishable competitors) will eventually rise to the top of the list, and remain there. But although any ordering will do, it seems appropriate to order hypotheses so that those judged more plausible are ranked above less plausible competitors. Of course, *plausibility* is not a logical notion, and any ordering of the hypotheses is a logically possible plausibility ordering. But, if the true hypothesis happens to be rated fairly highly by a plausibility ordering, then it will generally get to the top of that ordering's unrefuted list sooner than it would for orderings that rate it less favorably. So it would be perverse for a person to employ an ordering that places hypotheses that she considers the most plausible far below those she thinks highly implausible.

Total orderings on hypotheses fail to capture most of the characteristic features of the notion of plausibility. For example, if there are a number of equally plausible hypotheses, a total ordering must arbitrarily rank some above others. Eliminative induction, however, is not very sensitive to the subtleties of plausibility measures. *Some* ordering of hypotheses is required to assist the refutationist strategy, and it makes good sense to place those hypotheses one finds more plausible higher in the ordering. But, even if judgements of plausibility are way off the mark, and even though plausibility is poorly represented by a total ordering, the elimination by falsification of competing alternatives from an ordering for a deductive structure will eventually yield up the truth (along with its empirically equivalent competitors). Bayesian induction provides for a more sophisticated treatment of plausibility orderings, but in Bayesian induction eliminative processes operate in much the same way.

1.3 THE DEDUCTIVIST SIDE OF BAYESIAN INDUCTION

A plausibility ordering functions as a kind of tally device for eliminative induction, i.e., it keeps track of how well hypotheses are doing relative to the evidence, and it will bring the truth (and its empirically equivalent competitors) to the fore, given enough evidence. A mild extension of these devices may be generated by associating a weight with each hypothesis.

Intuitively, one might take these weights to represent the degree of plausibility of a hypothesis relative to alternatives.

Define P_α to be a *possible plausibility rating* for a deductive structure S just in case P_α is a function that assigns to each h_j in H some positive weight w_j (relative to the background b for S), where the sum of these weights is finite (i.e., $\sum_j w_j = r$, for some positive real number r). The value of the sum, r , may as well be taken as 1 since the weights can be normalized by replacing w_j with w_j/r . I'll use the expression ' $P_\alpha[h_j | b] = w_j$ ' to say that the degree of plausibility of h_j , given that b is true, has weight w_j according to plausibility weighting P_α . The sum of the weights is $\sum_j P_\alpha[h_j | b] = 1$. Normalization to 1 is only a convenience. The important feature of plausibility weights is not their numerical values, but the values of their ratios. The ratio $P_\alpha[h_j | b] / P_\alpha[h_i | b]$ represents the number of times more (or less) plausible h_j is than h_i , according to the plausibility weighting function P_α on S .

In technical jargon, P_α measures plausibility on a *ratio scale*. By introducing these weighting functions, and calling them *plausibility weightings*, I do not mean to suggest that any real person *has* a ratio scale plausibility rating for hypotheses. For one thing, real people are in no position to place relative plausibilities on infinitely many alternative hypotheses. And even if some computationally efficient method could be found to generate an ever increasing (but necessarily always finite) list of plausible alternative hypotheses, real people would probably be hard pressed to come up with precise relative weights for them. So, ratio scale plausibility functions on deductive structures are certainly an idealization with regard to practical inductive reasoning. But for my purposes it suffices to regard these ideal functions as merely a device that tracks the effects of *eliminative induction*. It is a better device than total orderings of hypotheses, for it permits the representations of relative plausibilities when they *are* available. Ambiguity or indeterminacy in relative plausibilities may be modeled as classes of these functions that represent all ways that the ambiguity could be resolved into ratio scales. Under the influence of eliminative induction the precise value of plausibility weights (assigned by a function or by a class of functions) will be of little consequence, just as the precise order of hypotheses in a plausibility ordering made little real difference in the ability of eliminative induction to bring the truth toward the top of the list. But, in general, the higher the weight of the true hypothesis initially, the sooner it will attain a high plausibility rating from evidence, since it will have to await the defeat of fewer competitors in order to become the most plausible remaining hypothesis.

With weights in place, eliminative induction on deductive structures proceeds as always. But now, whenever a hypothesis in H is falsified by one of the first n bits of evidence, its plausibility weight should be set to 0, relative to the evidence, i.e., $P_\alpha[h_j | c^n \cdot e^n \cdot b] = 0$. Once a hypothesis receives a weight of 0, its weight never rises; in effect it is dropped from the plausibility list.

The weights of the remaining, unfalsified hypotheses are not to change relative to one another, but the weights of *these* hypotheses will not add to 1. Some positive weight will have been lost with the removal from the list of those that are falsified. It is convenient to renormalize the weights on the unfalsified hypotheses, so that they again sum to 1. Thus, define the updated weight of each unfalsified hypothesis on the evidence as follows:

$$P_{\alpha}[h_j | c^n \cdot e^n \cdot b] = P_{\alpha}[h_j | b] \div r = w_j/r, \text{ where } r = \sum_{[U]} P_{\alpha}[h_j | b],$$

for U the set of hypotheses in H that remain unfalsified on
(b · cⁿ · eⁿ).

The new degree of plausibility for an unrefuted hypothesis is its original plausibility divided by the sum of the original weights of all of the remaining unfalsified hypotheses. Notice that the ratios of their original weights remains unchanged. The evidence makes no difference to the relative plausibilities of hypotheses it doesn't refute.

Every plausibility weighting on a deductive induction structure eventually succumbs to eliminative induction. All false alternatives with weights that initially rivalled that of the true hypothesis eventually obtain weights of 0. If the true hypothesis has no empirically equivalent competitors, then its weight must converge to 1 as evidence increases. If there are empirically equivalent alternatives to the truth, then the *sum of the weights* of these hypotheses must converge to 1 as evidence increases.

Any P_{α} that updates weights in this way (i.e., by the renormalization of weights for unrefuted hypotheses) is essentially a Bayesian probability function; indeed, all that Bayesian induction adds to eliminative induction on deductive structures are *constant relative plausibility weightings* for unrefuted hypotheses and *renormalization of the absolute weights* to sum to 1. To see this, assume that P_{β} is a probability function on the language of a deductive structure S. Then, provided that the truth of any given hypothesis is irrelevant to the probability that the observational conditions cⁿ hold (i.e., $P_{\beta}[c^n | h_j \cdot b] = P_{\beta}[c^n | b]$, which Bayesians implicitly assume), and provided that the initial probabilities of all members of H add to 1 (i.e., $\sum_j P_{\beta}[h_j | b] = 1$), Bayes' theorem is as follows:

$$P_{\beta}[h_i | e^n \cdot c^n \cdot b] = \frac{P_{\beta}[e^n | c^n \cdot h_i \cdot b] \times P_{\beta}[h_i | b]}{\sum_j P_{\beta}[e^n | c^n \cdot h_j \cdot b] \times P_{\beta}[h_j | b]}.$$

The axioms of probability theory imply that the values of $P_{\beta}[e^n | c^n \cdot h_j \cdot b]$ (the probability that the evidence would have occurred if (cⁿ · h_j · b) were true) will be 0 when the evidence contradicts h_j; and $P_{\beta}[e^n | c^n \cdot h_j \cdot b] = 1$ when (b · h_j · cⁿ) ⊢ eⁿ. Thus, Bayes' theorem says that hypotheses are to have probability 0 on falsifying evidence. Hypotheses that entail the evidence are updated by dividing their initial weights by the sum of the initial weights of all as yet unfalsified hypotheses. It follows that the ratios of the updated

probabilities of unfalsified hypotheses remain equal to the ratios of their initial probabilities. Hence, for deductive structures Bayes' theorem gives precisely the updating scheme described earlier for plausibility weightings.

Although plausibility assignments may differ widely on the weights they assign hypotheses, if there are no empirically equivalent competitors of the truth in H , then all plausibility weightings (or Bayesian probability functions) on a deductive structure S will *converge to agreement* on the plausibilities for hypotheses as evidence increases. Convergence is *guaranteed* to occur, because, for any given plausibility weighting, all the most plausible competitors of the true hypothesis (and an ever increasing number of less plausible competitors) will eventually be falsified. The plausibility weightings of all falsified hypotheses become 0, and the renormalized weights of the true hypothesis converge to 1. In short, the evidence brings about *convergence of opinion* about the plausibilities of hypotheses; convergence is to 0 or 1.

If the true hypothesis is empirically equivalent to some set of alternatives in H , then the only *convergence to agreement* that can be brought about by the evidence is the agreement that the falsified hypotheses have a plausibility weight of 0. Each plausibility function will eventually assign almost all of the weight jointly to the true hypothesis *and* its empirical equivalent alternatives. The various *possible* plausibility functions may differ widely on the relative plausibilities of empirically equivalent hypotheses, and the evidence cannot overcome such differences—nor should it.

2. PROBABILISTIC INDUCTION

Deductivist models of eliminative induction cannot represent the effect of *probabilistic evidence*, evidence that the hypotheses imply only statistically or probabilistically. Probabilistic evidence is of little consequence for hypotheses when falsifying evidence can be found. But hypotheses that are empirically equivalent to the true hypothesis for *deductively* related evidence may differ on the *likelihoods* they assign to a body of probabilistic evidence. Such hypotheses are not truly empirically equivalent to the true hypothesis. Indeed, the probabilistic evidence may come to *highly refute* false alternatives (i.e., to make them highly unlikely) when falsifying evidence is not available. Bayesian induction is a simple formalization of this idea. In this section I will examine the eliminative nature of Bayesian induction. Roughly, what I will show is that in Bayesian induction the influence of the values of the subjective prior probabilities for hypotheses can become “washed out” by the evidence (so that the values of posterior probabilities will come to agreement) *just in case* the hypotheses become highly refuted or highly confirmed by the evidence. Thus, Bayesian agents who disagree on the initial plausibilities of hypotheses (as represented by prior probabilities) can be

brought into agreement on the posterior probabilities by the evidence *only* through eliminative means. Bayesian induction is a simple extension of the deductivist model of eliminative induction to accommodate probabilistic evidence. The essence of Bayesian induction is the *relative elimination* of competing hypotheses.

2.1 THE STRUCTURE OF THEORY TESTING BY PROBABILISTIC EVIDENCE

In section 1.1, I defined the notion of an *induction structure* for hypothesis evaluation, $S = \langle b, H, C, O \rangle$. Recall that these structures are subject to three conditions:

- 1) hypotheses in H are pairwise logically incompatible relative to b ;
- 2) no two alternative outcomes of an observational condition c_k can occur, given $(h_j \cdot b)$;
- 3) at least one of the possible outcomes in O_k (the list of possible outcomes of condition c_k) must occur, if $(c_k \cdot h_j \cdot b)$ holds.

Induction structures were called *deductive induction structures* if they also satisfied:

- 4-d) for any n observations, c^n , each hypothesis (together with background knowledge b) logically entails some conjunction e^n of outcomes.

When hypotheses only assign probabilities to outcomes, this fourth condition fails.

Bayesian induction emerges from the assignment of probabilities to the sentences of *induction structures*. *Probabilistic induction structures* will capture the essential features of Bayesian induction. Define a *probabilistic induction structure* (more briefly, a *probabilistic structure*) as a pair $\langle S, P \rangle$, where S is an *induction structure* (i.e., $S = \langle b, H, C, O \rangle$ satisfies conditions 1–3 on *induction structures*) and P is some set of probability functions on the language of the sentences in S , such that:

- 4) each probability function P_α in P satisfies the axioms of classical probability theory on the language of S ;
- 5) for each P_α in P , $\sum_j P_\alpha[h_j | b] = 1$;
- 6) for each P_α in P there is some upper bound $K > 0$ such that, for all pairs of hypotheses h_i and h_j in H , and for each n , $0 < P_\alpha[c^n | h_i \cdot b] / P_\alpha[c^n | h_j \cdot b] \leq K$ (more simply, but with less generality, it would be reasonable to assume that $P_\alpha[c^n | h_j \cdot b] = P_\alpha[c^n | h_i \cdot b]$);
- 7) the true hypothesis is in H (although we don't know which member of H it is), and for all P_α in P , if h is the true hypothesis, then $P_\alpha[h | b] > 0$;

- 8) for any pair of probability functions P_α and P_β in \mathbf{P} and any h_j in H , and for all possible sequences of n bits of evidence e^n (for each n), $P_\alpha[e^n \mid c^n \cdot h_j \cdot b] = P_\beta[e^n \mid c^n \cdot h_j \cdot b] = P[e^n \mid c^n \cdot h_j \cdot b]$.

I will briefly discuss Conditions 4 through 6, and then devote the next subsection to Conditions 7 and 8.

In section 1 we saw that for *deductive structures* the introduction of a ratio scale measure on the relative plausibilities of hypotheses together with a reasonable scheme for updating plausibilities on falsifying evidence is tantamount to the application of a Bayesian probability function to sentences of the structure. In that context the precise values of the plausibility weights were of little consequence. *Probabilistic structures* will, of course, capture the same ratio scale measures on relative plausibilities of hypotheses, and, as it turns out, the precise values of the plausibility weights will be of little consequence when sufficient probabilistic evidence is available. Condition 4 merely formalizes the idea that a set \mathbf{P} of possible plausibility assignments will be associated with a *probabilistic structure*, and that the plausibility assignments are to be consistent with the classical axioms of probability theory, as Bayesian induction requires.⁹

From Conditions 1 and 4 it follows that the hypotheses are mutually exclusive; for each pair of hypotheses in H , every probability function in \mathbf{P} will yield $P_\alpha[h_i \cdot h_j \mid b] = 0$. Condition 5 adds that the hypotheses in H are all of the plausible alternative hypotheses for a probabilistic structure—they have all of the weight (normalized to 1). Also notice that Conditions 2 and 3 together with 4 imply that for each hypothesis h_j and observational condition c_k , the probabilities for possible outcomes (i.e., members of O_k) must sum to 1—i.e., $\sum_s P[\alpha_{k,s} \mid c_k \cdot h_j \cdot b] = 1$. These conditions also imply that for any sequence of n observations, the sum of all the possible sequences of outcomes (members of E^n) is 1—i.e., for each n , $\sum_{[E^n]} P[e^n \mid c^n \cdot h_j \cdot b] = 1$.

Condition 6 on probabilistic structures says that the initial conditions do not themselves function as evidence. The members of C describe initial experimental or observational conditions that bear on the likelihoods hypotheses assign to possible outcomes. The hypotheses should not substantially differ among themselves on the likelihood that the sequence of initial conditions c^n will occur. If hypotheses differ greatly on the likelihood for a condition c_k , then c_k should be classified among the evidential outcomes in O rather than as initial conditions in C . It would be natural to assume that the likelihoods of initial conditions are independent of particular hypotheses—i.e., $P_\alpha[c^n \mid h_j \cdot b] = P_\alpha[c^n \mid h_i \cdot b]$ for each pair of hypotheses. In that case, all occurrences of such likelihoods in the remainder of this paper may be safely ignored, since $P_\alpha[c^n \mid h_j \cdot b] / P_\alpha[c^n \mid h_i \cdot b]$ will equal 1. However, I will only assume the weaker condition that these ratios are all bounded above by some large number K , as stated in Condition 6.

2.2 THE TRUE HYPOTHESIS, AND THE OBJECTIVITY OF LIKELIHOODS

Conditions 7 and 8 for probabilistic structures require more extensive explanations than the first six conditions. This subsection will discuss them in some detail.

A set of probability functions \mathbf{P} for a probabilistic structure may contain all possible probability functions that assign non-zero probability to the hypotheses in H , or \mathbf{P} may possess some more limited subset of probability functions. Thus, various sets of probability functions can play the role of \mathbf{P} on the same induction structure S . Intuitively, a probability function may represent an ideal person's *rational belief function* on sentences.¹⁰ A set of probability functions \mathbf{P} might, then, model the belief functions of a number of ideal persons (i.e., Bayesian agents) who differ in their assessments of the initial plausibilities of hypotheses. Alternatively, \mathbf{P} may be used to represent imprecision or uncertainty regarding the plausibilities for hypotheses for a single Bayesian agent; the members of \mathbf{P} may consist of just those probability functions that are consistent extensions of some rather imprecise intuitions about the range within which the relative plausibilities of the hypotheses lie.

Regardless of how \mathbf{P} is understood, if a probability function in \mathbf{P} assigns a prior probability of 0 to a hypothesis, then no evidence can raise its posterior probability above 0. So, if the true hypothesis is in H , and if each probability function in \mathbf{P} is to have the ability to yield a non-zero posterior probability for it, then each member of \mathbf{P} *must* assign a non-zero prior probability to the true hypothesis. If, on the other hand, the true hypothesis does not occur in H —i.e., if the true hypothesis is never seriously considered by the Bayesian agents represented in induction structure S —then no mode of reasoning can ever come to support it. Condition 7 on probabilistic structures accommodates these concerns. It assumes only that the true hypothesis does come under consideration and has a non-zero initial plausibility for all Bayesian agents represented in probabilistic induction structure $\langle S, \mathbf{P} \rangle$.

Bayesians usually assume that hypotheses assign *objective* probability values to each of the possible ways the evidence might turn out. Condition 8 captures this idea. It says that, for each hypothesis h_j and sequence of possible outcomes e^n on a probabilistic structure, the probability functions in \mathbf{P} (for the structure) must agree on the *likelihood* of the occurrence of e^n if $(c^n \cdot h_j \cdot b)$ is true. To signify this “objectivity of likelihoods” relative to a given set \mathbf{P} , the subscript will be dropped from the ‘P’ in expressions for likelihoods like ‘ $P[e^n \mid c^n \cdot h_j \cdot b]$ ’. However, Condition 8 does not require likelihoods to be “absolutely objective”; it does *not* say that likelihoods must agree across all probabilistic structures. Various probabilistic structures on the same induction structure S may well disagree on the numerical value of $P[e^n \mid c^n \cdot h_j \cdot b]$. If there is imprecision or uncertainty regarding the numerical values of likelihoods, the ambiguity can be modeled as *classes* of probabilistic structures on a common induction structure S . I won’t pursue the details here.

Bayesians think of the likelihoods as objective with good reason. The likelihoods represent the empirical content of hypotheses and theories. They are the probabilistic counterparts of the logical entailments of outcomes by hypotheses in *deductive induction structures*. Indeed, *deductive structures* supplemented with weights on hypotheses are just special cases of *probabilistic structures*, cases for which all the likelihoods have value 1 or 0.

If the hypotheses in H are explicitly statistical, or if they are stochastically tied to outcomes by a statistical model of measurement error in the background b , then the objectivity of the likelihoods should be unproblematic. They go by a *logical* version of the rule Lewis¹¹ calls the *Principal principle*. This principle may be formulated roughly as follows:

1. Suppose ' $(h_j \cdot b)$ ' logically entails that systems of type Ψ on which a measurement of type Q is performed *have a propensity or statistical probability equal to r* of producing a measurement outcome of type ψ . Let ' c_k ' say that a particular system of type Ψ is measured appropriately (a Q measurement is performed on it); and let ' o ' say that the measurement outcome is of type ψ .
2. In addition, suppose ' $(c_k \cdot h_j \cdot b)$ ' *does not* logically entail that the measured system belongs to any other statistical reference class Φ *unless* either ' $(c_k \cdot h_j \cdot b)$ ' logically entails that Φ assigns a statistical measure equal to r to outcomes of type ψ , or else ' $(c_k \cdot h_j \cdot b)$ ' logically entails that every Ψ system is a Φ system (i.e., that Ψ is a sub-reference class of Φ).

Then the *likelihood* assigned to ' o ' by ' $(c_k \cdot h_j \cdot b)$ ' is r , i.e.,
 $P[o \mid c_k \cdot h_j \cdot b] = r$.

Likelihoods arising from this principle are often called *direct inferences*, after Carnap's¹² term for a similar notion. In a logical theory of probability, likelihoods are treated as meta-linguistic logical relationships between pairs of object language sentences, on a par with *logical entailments*. The idea is that just as 'all A s are B s, and c is an A ' logically entails ' c is a B ', so 'the *statistical measure* of B s among A s is r , and c is an A ' *probabilistically entails to degree r* that ' c is a B '. Kyburg¹³ has developed the most rigorous version of a logical formulation of direct inference probabilities, a version in which the probabilities for likelihoods depend only on the logical form of the object language sentences. Although his system is incompatible with Bayesian induction in some cases, Harper¹⁴ has adapted Kyburg's approach to Bayesian probability functions.¹⁵ I think some such approach will provide an adequate formalization of direct inference probabilities, although the details are still not completely settled. But, however this logical program pans out, when hypotheses are precise enough for the *Principal principle* to operate, it seems reasonable to require the probability functions in P to agree on the probabilities for likelihoods.

When hypotheses are not so precise as to furnish likelihoods via the *Principal principle*, a certain degree of objectivity should still be expected of them. To argue for this claim I must first briefly sketch how likelihoods operate in Bayesian induction. Then we will see just how objective the likelihoods *need* to be.

Generally, as the amount of evidence increases, the likelihoods for the accumulated evidence will decrease toward 0, even with respect to a true hypothesis. For example, the likelihood of obtaining any specific sequence of outcomes consisting of exactly 50 heads on 100 tosses of a coin is $(1/2)^{100}$ on the hypothesis that the coin is "fair." This sequence of outcomes supports the "fairness hypothesis" so strongly *not* because it is likely, but rather because the likelihood of this outcome is much smaller relative to alternative hypotheses, e.g., that the coin is "bent" or biased to a certain degree. In Bayesian induction the *relative likelihoods* of the evidence with regard to alternative hypotheses does all of the evaluative work. When a hypothesis that has a non-zero prior probability says that the actual sequence of outcomes is *much* more likely than does a competing hypothesis, I will say that the competitor becomes *highly refuted* by the evidence relative to that hypothesis. *Highly refuted* hypotheses are usually not falsified; additional evidence might revive them. However, in section 3 we will see that under certain minimal assumptions about the quality of the accumulating evidence, false hypotheses will almost certainly become highly refuted by enough evidence, and they will remain highly refuted on additional evidence. This result does not depend essentially on the objectivity of likelihoods. Indeed, in Bayesian induction the precise values of likelihoods are almost never important (except when they are 0). Evidence only has an influence in so far as the likelihood ratios, $P[e^n \mid c^n \cdot h_j \cdot b] / P[e^n \mid c^n \cdot h_i \cdot b]$, rise or fall as the amount of evidence increases. If this ratio goes toward 0, then h_j becomes highly refuted relative to h_i on e^n for the probabilistic structure in question.

When Bayesian refutation occurs, the precise values of likelihood ratios are not really important. If two probability functions, P_α and P_β , from different probabilistic structures do not represent radically different assessments of the likelihoods, then any hypothesis that becomes increasingly *refuted* for P_α will do so for P_β . For example, it would suffice if the likelihood ratios given by P_α and P_β are always within some very large constant multiple of each other. When P_α and P_β are so related, the likelihood ratios (for a pair of hypotheses) go to 0 for P_α *just in case* they go to 0 for P_β . On the other hand, if two probability functions assign *radically* different values to likelihood ratios, then their values for likelihoods have to be *very* different, too. Any two such probability functions must disagree radically on the empirical import of some hypothesis or theory; these probability functions widely disagree about what the theory says the observable part of the world is like. Effectively the two probability functions employ the same syntax to express quite different theories, and this difference shows up in their radically

different empirical import. Thus, even when hypotheses are not so precise as to furnish objective likelihoods via the *Principal principle*, the likelihoods should be objective to the degree that all probability functions that represent approximately the same understanding of the empirical import of each hypothesis will come to highly refute the same hypotheses.

As it happens, practically none of the analysis of Bayesian induction in the remainder of this paper relies on the objectivity of the likelihoods. In most of what follows the reader may safely relativize likelihoods and their ratios to a particular probability function, and write in a subscript ' α ' where I've left it out. In contexts where objective likelihoods are required I will say so explicitly. However, the idea that likelihoods should be relatively objective is a central theme in Bayesian induction. The whole point of using Bayes' theorem to evaluate posterior probabilities of hypotheses (relative to evidence) is that the likelihoods used in Bayes' theorem are supposed to provide a fairly objective way of assessing the impact of evidence on the plausibilities of hypotheses.

One further point about the defining conditions of induction structures should be noted. Bayesians often assume that the individual bits of evidence are *probabilistically independent* of one another *relative to a hypothesis*. Independence would mean that the likelihood of the accumulated evidence relative to h_j is the multiplicative product of the likelihoods of each part of the accumulation, i.e., $P[e^n | c^n \cdot h_j \cdot b] = \prod_{k=1}^n P[e_k | c_k \cdot h_j \cdot b]$. For explicitly stochastic hypotheses, where the *Principal principle* applies, the evidence can usually be chunked into independent pieces. Indeed, it seems that any reasonable scientific theory should provide a way to carve the evidence into separate, stochastically independent events. If a theory fails to do so, then each time the theory is used to predict an observable event it has to employ *all of the past* observational and experimental data as initial conditions. The following analysis of Bayesian induction will, however, in no way depend on the independence of bits of evidence. Independence will not be *assumed* for the possible outcomes in induction structures, but occasionally I will point out the effect it would have, as a special case.

The next subsection will describe Bayesian induction in a manner that makes the central role of likelihood ratios apparent. Then, the remainder of the paper will investigate conditions under which the likelihood ratios bring about the relative refutation or confirmation of hypotheses.

2.3 LIKELIHOOD RATIOS AND BAYES' THEOREM

For any sequence of possible outcomes, e^n , the *ratio* of the likelihood that e^n will occur if hypothesis h_j is true *over* the likelihood that e^n will occur if h_i is true is called *the likelihood ratio for h_j over h_i relative to a possible sequence of outcomes e^n* . Let ' $LR[e^n | j/i]$ ' represent this ratio, i.e., relative to a given probabilistic structure $\langle S, P \rangle$, define $LR[e^n | j/i]$ as follows:

$$LR[e^n | j/i] = P[e^n | c^n \cdot h_j \cdot b] / P[e^n | c^n \cdot h_i \cdot b] .$$

The numerical value of a likelihood ratio represents *the number of times more (or less) likely* the occurrence of outcome e^n would be if h_j were true than it would be if h_i were true. In Bayesian induction, likelihood ratios provide *the* mechanism through which evidence has its impact on the relative plausibilities of hypotheses. But, the likelihood ratios do not show up explicitly in Bayes' theorem as it's usually written. In this subsection I will describe the posterior probabilities of hypotheses relative to the evidence (i.e., the Bayesian updating of their plausibility ratings) in a form that makes the role of the likelihood ratios conspicuous.

When the sequence of outcomes e^n consists of independent conjuncts e_k , where each is independent of the rest relative to any given hypothesis h in H , then a likelihood ratio can be expressed as a *product* of the likelihood ratios of the conjuncts:

$$LR[e^n | j/i] = \frac{P[e^n | c^n \cdot h_j \cdot b]}{P[e^n | c^n \cdot h_i \cdot b]} = \prod_{k=1}^n \frac{P[e_k | c_k \cdot h_j \cdot b]}{P[e_k | c_k \cdot h_i \cdot b]}.$$

Though desirable, independence is not essential to the workings of likelihood ratios, and I will not assume that it holds. But later I will discuss independent outcomes as a special case.

For each P_α in a probabilistic induction structure, the relative plausibility of one hypothesis over another relative to the evidence is measured by the product of their likelihood ratio with their relative initial plausibilities:

$$(I) \quad \frac{P_\alpha[h_j | e^n \cdot c^n \cdot b]}{P_\alpha[h_i | e^n \cdot c^n \cdot b]} = LR[e^n | j/i] \times \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} \times \frac{P_\alpha[c^n | h_j \cdot b]}{P_\alpha[c^n | h_i \cdot b]}.$$

This equality is a theorem of probability theory (i.e., for each hypothesis h ,

$$P_\alpha[h | e^n \cdot c^n \cdot b] = P_\alpha[e^n \cdot c^n \cdot h | b] \div P_\alpha[e^n \cdot c^n | b] = \\ P[e^n | c^n \cdot h \cdot b] \times P_\alpha[c^n | h \cdot b] \times P_\alpha[h | b] \div P_\alpha[e^n \cdot c^n | b].$$

The last ratio on the right of (I) represents the relative plausibility of the occurrence of the initial conditions, c^n , on each hypothesis. It seems reasonable to assume that the occurrence of the initial conditions is not a lot more likely on one hypothesis than on another. Thus, Condition 6 on probabilistic induction structures supposes that plausibility ratios for initial conditions are bounded above (for all n). It would not be unreasonable to take the value of this ratio to be 1, and then to ignore it throughout the rest of the paper. I will continue to display these ratios, but only to *show* that if they are bounded, then nothing of substance would be lost by ignoring them.

Equation (I) suggests a way to extend eliminative induction to probabilistic evidence. Although a hypothesis, h_j , might avoid falsification, it may become *increasingly refuted* relative to an alternative, h_i , provided that h_i 's initial plausibility is above 0 (for P_α) and the likelihood ratios $LR[e^n | j/i]$ go to 0. In other words, if the evidence becomes increasingly less likely on h_j

than on h_i , and if h_i has some plausibility, then the plausibility of h_j must fall toward 0. Indeed, the theorems in the next subsection will show that virtually the only worthwhile role evidence can play in Bayesian induction is this eliminative role via likelihood ratios.

The relationship between relative plausibilities of hypotheses (on the evidence) and the likelihood ratios, as expressed in equation (I), is really all there is to Bayesian induction. The *absolute* probability of a hypothesis (for a probability function P_α) comes directly from the sum of the relative plausibilities. To see this, first consider the odds, Ω_α , against a hypothesis h_i relative to evidence, defined as follows:

$$(II) \quad \Omega_\alpha[\neg h_i | e^n, c^n, b] = \frac{P_\alpha[\neg h_i | e^n, c^n, b]}{P_\alpha[h_i | e^n, c^n, b]} = \sum_{j \neq i} \frac{P_\alpha[h_j | e^n, c^n, b]}{P_\alpha[h_i | e^n, c^n, b]}.$$

The odds against a hypothesis is the sum of the relative plausibilities for its competitors, the sum of instances of equation (I). The relationship between the odds against a hypothesis and the likelihood ratios follows directly from equations (I) and (II):

$$(III) \quad \Omega_\alpha[\neg h_i | e^n, c^n, b] = \sum_{j \neq i} L[e^n | j/i] \times \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} \times \frac{P_\alpha[c^n | h_j, b]}{P_\alpha[c^n | h_i, b]}.$$

The absolute probability of a hypothesis on evidence is related to the odds against it by:

$$(IV) \quad P_\alpha[h_i | e^n, c^n, b] = 1 / (1 + \Omega_\alpha[\neg h_i | e^n, c^n, b]).$$

Equation (IV) follows easily from the definition of odds given in equation (II). Taken together, equations (III) and (IV) are a form of Bayes' theorem. The expression of Bayes' theorem in terms of the odds against a hypothesis makes the role of the likelihood ratios more perspicuous than does the more usual form of the theorem:

$$(V) \quad P_\alpha[h_i | e^n, c^n, b] = \frac{P[e^n | c^n, h_i, b] \times P_\alpha[h_i | b] \times P_\alpha[c^n | h_i, b]}{\sum_j P[e^n | c^n, h_j, b] \times P_\alpha[h_j | b] \times P_\alpha[c^n | h_j, b]}.$$

To see the role of likelihood ratios in Bayesian inference, consider equation (III). The odds against a hypothesis are at least 0, and can be arbitrarily large. Equation (III) implies that if h_i becomes increasingly refuted relative to at least one alternative h_j , i.e., if $L[e^n | i/j]$ converges to 0, then its inverse, $L[e^n | j/i]$, blows up toward ∞ and the odds against h_i go to ∞ with it. Hence, by equation (IV), the probability of h_i goes to 0. On the other hand, if *every* alternative to h_i becomes increasingly refuted relative to it, i.e., if for every alternative h_j , $L[e^n | j/i]$ converges to 0, then by equation (III) the odds

against h_i converge to 0. When this happens, equation (IV) says that the probability of h_i converges to 1.

Now, suppose the accumulating evidence does not drive the probability of h_i to either 0 or 1. Then for some alternative hypothesis h_j , the likelihood ratios $L[e^n \mid j/i]$ will neither blow up nor converge to 0. If these likelihood ratios do not go to extremes, but instead remain fairly close to 1, then equation (I) implies that the relative probabilities of h_j over h_i *on the evidence* will remain close to their initial relative probabilities. Indeed, if the likelihood of evidence on h_i agrees with the likelihood on h_j , then $L[e^n \mid j/i] = 1$, and the evidence yields no change in the relative plausibilities of h_j over h_i . When the evidence fails to take the likelihood ratios to extremes, the initial plausibility assessments, $P_\alpha[h_i \mid b]$, will continue to significantly influence the revised plausibility assessments $P_\alpha[h_i \mid c^n \cdot e^n \cdot b]$.

Clearly, evidence that yields extreme likelihood ratios will bring all probability functions in a probabilistic structure into agreement on the posterior probabilities of hypotheses, agreement converging on probabilities of 0 or 1. Such extreme evidence completely washes out the influence of the prior probabilities by highly refuting all but one of the competitors. When the evidence is not so extreme (i.e., when it fails to refute some competitors), one might hope that evidence can at least overcome the differences in prior probabilities for hypotheses and bring *convergence to agreement* on their posterior probabilities. The various probability functions in a probabilistic structure represent alternative assessments of the initial plausibilities of hypotheses, and one might hope that evidence can overcome such initial differences. In the next subsection we will see that this non-eliminative form of convergence generally cannot happen.

2.4 THE ELIMINATIVE NATURE OF BAYESIAN INDUCTION

For a given probabilistic structure $\langle S, P \rangle$, the probability functions in P can disagree widely in the prior probabilities they assign to hypotheses. Bayesians take P to represent a diversity of views on the initial plausibilities of the alternative hypotheses among ideal Bayesian agents. Theorems 2 through 4 will specify stringent conditions that the evidence must satisfy in order to bring *convergence to agreement* among functions in P on the posterior probabilities of hypotheses. Thus, these conditions will characterize what it takes for evidence to bring agreement among Bayesian agents regarding the degree to which the hypotheses are plausible. Roughly, Theorem 2 will show that posterior probabilities for hypotheses converge to 0 just when they become highly refuted via likelihood ratios that converge to 0. Theorem 3 will show, roughly, that posterior probabilities go to 1 just in case all alternative hypotheses are refuted by 0-converging likelihood ratios. Finally, Theorem 4 will show that if any two modestly disagreeing probability functions in P *converge to agreement* on the posterior probabilities for a

hypothesis, then the agreement will only be reached as the posterior probabilities converge to 0 or 1. Thus, we will see that for Bayesian induction, the evidence overcomes the initial disagreement or uncertainty among ideal Bayesian agents regarding the relative plausibilities of hypotheses *just in case* it is powerful enough to highly refute competing hypotheses through diminishing likelihood ratios.

Theorems 2 and 3 formally express intuitively obvious relationships between extreme likelihood ratios and posterior probabilities. I call them the *Zero-Converging* and *One-Converging Likelihood Ratio Theorems*. The Zero-Converging theorem states some necessary and some sufficient conditions for likelihood ratios to bring posterior probabilities to converge to 0. The One-Converging theorem states conditions under which likelihood ratios raise posterior probabilities to 1. I will omit the proofs of these theorems, since they follow easily from the connections between likelihood ratios and ratios of probabilities (equation (I)), between likelihood ratios and odds (equation (III)), and between odds and posterior probabilities (equations (II) and (IV)).

THEOREM 2: Zero-Converging Likelihood Ratio Theorem.

Let P_α be any probability function in a *probabilistic induction structure* $\langle S, P \rangle$, and let h_i and h_j (in H for S) be any distinct pair of hypotheses such that $P_\alpha[h_i | b] > 0$ and $P_\alpha[h_j | b] > 0$. Then:

- 1) $\lim_n (P_\alpha[h_j | e^n \cdot c^n \cdot b] / P_\alpha[h_i | e^n \cdot c^n \cdot b]) = 0$ iff $\lim_n LR[e^n | j/i] = 0$;
- 2) if $\lim_n LR[e^n | j/i] = 0$, then $\lim_n P_\alpha[h_j | e^n \cdot c^n \cdot b] = 0$;
- 3) if $\lim_n P_\alpha[h_j | e^n \cdot c^n \cdot b] = 0$ and there is a real number r such that, for all n , $P_\alpha[h_i | e^n \cdot c^n \cdot b] \geq r > 0$, then $\lim_n LR[e^n | j/i] = 0$.

Theorem 2 says that for any sequence of outcomes, the likelihood ratios for h_j over h_i converge to 0 *just when* the ratios of posterior probabilities for h_j over h_i converge to 0. And such convergence requires the posterior probabilities of h_j to go to 0 as well. On the other hand, the convergence to 0 of the posterior probabilities for h_j *does not* imply that the likelihood ratios of h_j over h_i will go to 0, for the posterior probabilities for h_i may go to 0 even faster than those for h_j . But, if the evidence is such that, for some h_i , the posterior probability of h_i does not get arbitrarily close to 0, then the posterior probability of any h_j will converge to 0 if and only if its likelihood ratios (over h_i) go to 0.

All probability functions within a probabilistic structure agree on the likelihoods. So, Theorem 2 implies that a hypothesis will become increasingly refuted by decreasing likelihood ratios (relative to an alternative) *just in case*, for *every* probability function in the probabilistic structure, the posterior probability of the hypothesis converges to 0. The theorem affirms

that refutation via likelihood ratios is related to posterior probabilities near 0, just as it should be, and that all plausibility measures in a structure come to agree on the near-zero probabilities of highly refuted hypotheses.

THEOREM 3: One-Converging Likelihood Ratio Theorem.

Let P_α be any probability function in a *probabilistic induction structure* $\langle S, P \rangle$, and let h_i (in H for S) be any hypothesis such that $P_\alpha[h_i | b] > 0$. Then:

- 1) if $\lim_n P_\alpha[h_i | e^n \cdot c^n \cdot b] = 1$, then for every h_j distinct from h_i , such that $P_\alpha[h_j | b] > 0$, $\lim_n LR[e^n | j/i] = 0$;
- 2) if for every h_j distinct from h_i such that $P_\alpha[h_j | b] > 0$, $\lim_n LR[e^n | j/i] = 0$, and there is some real number $K > 1$ such that, for every such h_j , for all n , $LR[e^n | j/i] \leq K$ (i.e., the likelihood ratios are bounded above), and for all n , $P_\alpha[c^n | h_j \cdot b] / P_\alpha[c^n | h_i \cdot b] \leq K$ (the ratios of likelihoods for initial conditions are bounded above), then $\lim_n P_\alpha[h_i | e^n \cdot c^n \cdot b] = 1$.

Theorem 3 says that the connection between posterior probabilities that approach 1 and the refutation of all alternative hypotheses is just as one might expect. If the posterior probability of h_i goes to 1, then the *likelihood ratios* of *all* competitors relative to h_i must go to 0, and the alternatives become increasingly refuted. If h_i has empirically equivalent competitors, then only the sum of their posterior probabilities can approach 1. If this sum goes to 1, then all other hypotheses must be refuted by 0-converging likelihood ratios.

The “eventual” increasing refutation of *all* competitors is not quite sufficient to guarantee that the posterior probability of a hypothesis goes to 1. At each place in the sequence of evidence there may be an alternative hypothesis for which the likelihood ratio relative to h_i has not yet started toward 0. And as this hypothesis becomes highly refuted, there may still be another *as yet unrefuted* hypothesis with an even larger likelihood ratio than the last. So, even if each alternative is *eventually* refuted, there may always be another likely contender, another alternative that has been made highly likely by the evidence thus far. However, the second part of Theorem 3 implies that if the likelihood ratios for false alternatives relative to a true hypothesis, h_i , satisfy a very plausible condition—they never become larger than some perhaps very large finite number (and the probability ratios for initial conditions are also bounded)—then the *eventual* increasing refutation of all alternatives relative to h_i *will suffice* to bring the posterior probability of h_i arbitrarily close to 1.

When h_i has evidentially indistinguishable competitors, its posterior probability cannot converge to 1, and the present version of Theorem 3 does not apply. But, it is easy to generalize Theorem 3, as follows:

THEOREM 3*: Extended One-Converging Likelihood Ratio Theorem.

Let W be any subset of the hypotheses in H such that all hypotheses in W have non-zero initial probabilities:

- 1) if $\lim_n \sum_{[W]} P_\alpha[h_i | e^n \cdot c^n \cdot b] = 1$, and for all h_i in W , $\lim_n P_\alpha[h_i | e^n \cdot c^n \cdot b] \neq 0$, then for every h_j not in W such that $P_\alpha[h_j | b] > 0$, $\lim_n LR[e^n | j/i] = 0$ for all h_i in W ;
- 2) if for every h_j not in W such that $P_\alpha[h_j | b] > 0$, $\lim_n LR[e^n | j/i] = 0$ for each h_i in W , and there is some real $K > 1$ such that, for every such h_j not in W and all h_i in W , for all n , $LR[e^n | j/i] \leq K$ and, for all n , $P_\alpha[c^n | h_j \cdot b] / P_\alpha[c^n | h_i \cdot b] \leq K$, then $\lim_n \sum_{[W]} P_\alpha[h_i | e^n \cdot c^n \cdot b] = 1$.

Roughly, the sum of the posterior probabilities of a set of hypotheses approaches 1 *just in case* all *other* alternative hypotheses are refuted via 0-converging likelihood ratios. This generalization applies both to sets, W , of hypotheses that are evidentially indistinguishable from the truth, and also to broader sets of hypotheses that do not differ *significantly enough* from the true hypothesis on the likelihoods of the accumulating evidence.

All probability functions within a probabilistic structure agree on the values for likelihoods. So, as a hypothesis undergoes increasing refutation (via likelihood ratios that approach 0), the probability functions in the induction structure must converge to agreement at 0 for the posterior probabilities of the hypothesis. And if all alternatives of a hypothesis become increasingly refuted, then all probability functions in P converge to agree at 1 for the posterior probabilities of the hypothesis. But, we might hope that there is another way for the evidence to force diverse probability functions into agreement on posterior probabilities, agreement for hypotheses that are neither highly refuted nor highly confirmed. Can *convergence to agreement* occur without convergence to 0 or 1? Theorem 4 will show that if a pair of probability functions *differ only modestly* (in a sense that I am about to explain) on initial probability assignments, then convergence to agreement can occur *only if* the evidence increasingly refutes all but one alternative, and its posterior probability goes to 1.

DEFINITION: Modestly Different Probability Functions.

A pair of probability functions P_α and P_β on a common *induction structure* will be said to *modestly differ* with each other *relative to* h_i in H *if and only if* one of them, say P_α , is related to the other as follows:

- i) $P_\beta[h_i | b] > 0$;
- ii) there is a real number $M > 1$ such that, for each alternative h_j to h_i , $P_\beta[h_j | b] \geq M \times P_\alpha[h_j | b] > 0$;

- iii) for every hypothesis h in H , $P_\beta[c^n \mid h \cdot b] = P_\alpha[c^n \mid h \cdot b]$
(for all n);
- iv) P_α and P_β agree on the likelihoods, i.e., for every hypothesis h in H and for each possible outcome sequence e^n ,
 $P_\beta[e^n \mid c^n \cdot h \cdot b] = P_\alpha[e^n \mid c^n \cdot h \cdot b]$ (for all n).

Suppose, for example, that $P_\alpha[h_1 \mid b] = 3/4$, and for $j \geq 2$, $P_\alpha[h_j \mid b] = 1/2^{j+1}$. One of the probability functions that *differs modestly* from P_α with respect to h_1 is a function P_β such that (for $M = 2$) $P_\beta[h_j \mid b] = 2 \times P_\alpha[h_j \mid b] = 1/2^j$, for $j \geq 2$; and $P_\beta[h_1 \mid b] = 1/2$. Indeed, for any given probability function, every other probability function that can be generated from it by increasing the initial probabilities of *all alternatives* to h_i by at least some constant multiple $M > 1$ will *differ modestly* from the original probability function with regard to h_i . Similarly, given any probability function P_β , all probability functions P_α whose prior probabilities for the alternatives to hypothesis h_i are *somewhere below* (e.g., less than 99.99% of) the respective prior probabilities that P_β assigns will *modestly differ* from P_β (where $1/M = .9999$ in this case). Then, the only hypothesis to which P_β assigns a higher prior probability than P_α is h_i .

The next theorem says that if any two modestly different probability functions (which agree on likelihoods) converge to agreement on the posterior probability of a hypothesis, then either the posterior probabilities for the hypothesis converge to 0 or else they converge to 1. In light of the previous two theorems this means that convergence to agreement is possible *only when* eliminative induction operates through zero-converging likelihood ratios.

THEOREM 4: Non-Zero Convergence is Convergence to One.

Consider any *probabilistic induction structure* $\langle S, P \rangle$. Let h_i be some hypothesis in H , and suppose there is a P_α in P that satisfies the following conditions:

- i) there is a real number r such that, for all n , $P_\alpha[h_i \mid e^n \cdot c^n \cdot b] \geq r > 0$, and
- ii) there is a P_β that *modestly differs* with P_α relative to h_i ,
and
- iii) $\lim_n |P_\alpha[h_i \mid e^n \cdot c^n \cdot b] - P_\beta[h_i \mid e^n \cdot c^n \cdot b]| = 0$.

Then $\lim_n P_\alpha[h_i \mid e^n \cdot c^n \cdot b] = 1$.

Thus, if *all* probability functions in P come arbitrarily close to agreement on the posterior probabilities for h_i , then (by Theorem 3) they do so only as the evidence (via likelihood ratios) increasingly refutes all alternatives to h_i .
(Proof is in the appendix.)

The upshot is that *generally* Bayesian convergence can occur in only two ways. The first is for a hypothesis to become increasingly refuted, via likelihood ratios, relative to an alternative. In that case, all probability functions assign posterior probabilities for the hypothesis that converge toward 0 as evidence accumulates. The other way that Bayesian convergence can occur (for probability functions that modestly differ) is when *all* probability functions assign posterior probabilities to a hypothesis that converge to 1, and assign posterior probabilities that converge to 0 to all alternatives. For this to happen, all competitors of the hypothesis must become increasingly refuted by way of likelihood ratios.

There are, of course, *special* classes of probability functions for which convergence short of 0 and 1 may occur. Suppose that all probability functions in **P** agree on the prior probabilities for most hypotheses, but disagree on the prior probabilities for some finite subset of hypotheses. If each of the finitely many hypotheses on which there is initial disagreement becomes increasingly refuted by the evidence, then all probability functions in **P** will converge on values for posterior probabilities for the unrefuted hypotheses, posterior probabilities that need not be 0 or 1. But, if **P** contains even one probability function that *modestly differs* from another on the initial plausibilities of hypotheses, then convergence to agreement occurs for these two probability functions (and so, for *all* members of **P**) *only* at 0 or 1.

If the true hypothesis has empirically equivalent alternatives that are not laid low by plausibility considerations, then the influence of their initial probabilities cannot be washed out. Equation (I) says that their relative posterior probabilities remain at the relative weights of their initial probabilities. The influence of initial probabilities is overwhelmed by evidence only for those hypotheses that the evidence increasingly refutes.

Is there any reason to think that empirically distinct alternatives to the true hypothesis *will be increasingly refuted* by the accumulating evidence? Theorem 1 says that if hypotheses differ on *deductively entailed* evidence, false competitors will eventually be falsified. The two theorems in the next section extend Theorem 1. They show that a sufficient amount of "low quality" probabilistic evidence will almost certainly suffice to increasingly refute each empirically distinct competitor of the true hypothesis.

3. BAYESIAN REFUTATION AND THE QUALITY OF THE POTENTIAL EVIDENCE

Under what circumstances is it likely that a given alternative to the true hypothesis will be refuted? Bayesian refutation of a false hypothesis is *logically assured* when the true hypothesis logically entails the occurrence of an event that is logically inconsistent with the false hypothesis (Theorem 1).

Although we do not know which hypothesis *is* true, we know that all of the empirically distinct alternatives to the true hypothesis on *deductively* related evidence can be falsified by evidence, and the posterior probability of the true hypothesis will rise. But suppose hypothesis h_j is empirically distinct from the true hypothesis only with respect to probabilistic evidence. The central question for Bayesian induction then becomes:

If a hypothesis h_i is true, under what circumstances is it likely that an alternative, h_j , *will become* highly refuted relative to h_i by probabilistically related evidence—i.e., under what conditions is it likely that a sequence of outcomes, e^n , *will occur* for which the likelihood ratios, $LR[e^n | j/i]$, go to 0 as n increases?

I will offer an answer to this question in the form of a generalized version of Savage's theorem.¹⁶ I will describe very general conditions on probabilistic evidence that, when satisfied, make it highly likely that empirically distinct alternatives to the true hypothesis will become highly refuted via likelihood ratios.

I will rely on two theorems to make my case, Theorems 5 and 6. They show that if a hypothesis empirically differs from the true hypothesis on the probabilistic likelihoods of some possible outcomes, and if the *expected quality* of the accumulating evidence is, *on average*, not extremely poor, then it is highly probable that some sequence of outcomes will occur that will drive the likelihood ratios toward 0. This claim will be made precise with the aid of an information-theoretic measure of the *expected quality of the information* from observations, a measure of the potential power of an experiment or measurement (or other *observation*) to discriminate between hypotheses. The ability of observations to discriminate between hypotheses depends entirely on the degree to which two hypotheses disagree on the likelihoods for the various *possible outcomes* that might result from experiments or measurements. If on average these likelihoods differ (even minutely), then the acquisition of a large enough quantity of this evidence will very probably (as near to 1 as you like) yield outcomes that collectively bring the *likelihood ratios* as close to 0 as you please.

Throughout this section I will restrict attention to a single probability function P_α that satisfies the rules for *probabilistic induction structures* on an induction structure S . The objectivity of likelihoods is not essential in what follows, but I will continue to drop the subscript ' α ' from P_α in expressions containing only likelihoods. This will emphasize the fact that none of the following considerations depend on the values P_α assigns as initial probabilities to hypotheses. I will also restrict attention to a pair of hypotheses h_j and h_i ; h_i will often play the role of the true hypothesis. The following *sufficient conditions* for the likely Bayesian refutation of h_j relative to h_i apply equally to any pair of hypotheses from an induction structure, and apply to every probability function, P_α , that satisfies the conditions for probabilistic induction structures.

3.1 THE REFUTATION OF HYPOTHESES THAT DENY REAL POSSIBILITIES

It will be convenient to divide the labor between two cases. First I will treat the special case in which some *really possible* outcomes—outcomes that have non-zero likelihoods on the true hypothesis—are assigned likelihoods of 0 by alternative h_j . The next subsection treats the more difficult case where h_j assigns *some* positive likelihood to all of the *real possibilities*.

Suppose that, for each n , every *possible* sequence of events e^n to which h_j assigns a non-zero likelihood *can be extended* by at least one of the next possible outcomes, $o_{n+1,r}$, (for the next observation c_{n+1}) to a sequence e^{n+1} to which h_j assigns likelihood 0. If h_i is the true hypothesis, and if it assigns non-zero likelihoods to some of *these* outcome sequences, then an event that refutes h_j is almost sure to occur eventually. The true hypothesis probabilistically asserts that such refutation is likely, approaching certainty as the evidence increases.

THEOREM 5: *The Special Refutation Theorem.*

If for all e^k such that $P[e^k \mid c^k \cdot b \cdot h_j] > 0$, the following two conditions hold:

- 1) $P[e^k \mid c^{k+1} \cdot b \cdot h_j] = P[e^k \mid c^k \cdot b \cdot h_j]$, and similarly for h_i ;
- 2) $P[\vee_r \{o_{k,r} \mid P[o_{k,r} \mid c^k \cdot e^{k-1} \cdot b \cdot h_j] = 0\} \mid c^k \cdot e^{k-1} \cdot h_i \cdot b] \geq \delta > 0$,

then $P[\vee\{e^n \mid LR[e^n \mid j/i] = 0\} \mid c^n \cdot h_i \cdot b] \geq 1 - (1-\delta)^n$,

and thus, $\lim_n P[\vee\{e^n \mid LR[e^n \mid j/i] = 0\} \mid c^n \cdot h_i \cdot b] = 1$.

(See the appendix for proof.)

The first condition in the antecedent of Theorem 5 merely asserts that the likelihoods assigned by h_i (and by h_j) to the possible sequences of outcomes e^k are independent of the initial condition c_{k+1} for the next observation or experiment. The expression in the second condition ' $\vee_r \{o_{k,r} \mid P[o_{k,r} \mid c^k \cdot e^{k-1} \cdot b \cdot h_j] = 0\}$ ' represents the sentence formed by taking the disjunction of all possible outcomes of the k^{th} observation, c_k , that have 0 probability relative to h_j (conjoined with $c^k \cdot e^{k-1} \cdot b$). Similarly, ' $\vee\{e^n \mid LR[e^n \mid j/i] = 0\}$ ' describes the disjunction of all possible outcome sequences for the first n observations that would yield a likelihood ratio of 0 for h_j over h_i .

Theorem 5 says that if hypothesis h_i assigns at least some minuscule probabilities, δ , to disjunctions of possibilities that h_j says are impossible, then h_i says (in terms of likelihoods) that increasing evidence will almost surely yield one of the outcomes that h_j says is impossible. When this occurs, the likelihood ratio of h_j over h_i will be 0, and h_j will be refuted. The posterior probability of h_j will be 0.

Notice that the theorem does not assume that future outcomes are independent of past outcomes. If, for likelihoods relative to h_i and to h_j , all outcomes in the sequences e^k are independent of one another, then reference to

evidence on the *right sides* of conditional probabilities may be dropped from Condition 2 of Theorem 5—i.e., both occurrences of ' $c^k.e^{k-1}$ ' may be replaced by ' c_k ' in Condition 2.

The claim that the theorem makes is really quite obvious. For example, let hypothesis h_i be a quantum theory of protons that implies that a proton has some perhaps quite low probability of decay in any given year; and consider an alternative theory, h_j , that says protons never decay (or that the probability of a proton decay is 0). If h_i is true, then eventually a proton decay will almost surely be detected, and h_j will be proved false (provided proper detectors can be built and billions of protons are kept under observation for long enough).

In the most general case Theorem 5 applies only to some subsequence of the total sequence of observations. Only some experiments or observations will have possible outcomes that have non-0 likelihoods relative to h_i and likelihoods of 0 for h_j . Theorem 5 puts a lower bound on the likelihood that any such subsequence of experiments or observations will produce outcomes that *absolutely refute* h_j , a lower bound that depends on the size of the subsequence. If there is a large enough finite (or an infinite) subsequence of such observations available, then h_j will almost surely receive a posterior probability of 0 on this part of the evidence, and thus will receive a posterior probability of 0 on the full sequence of evidence. But if there is only a rather small number of such *possibly refuting observations* available, and if the actual outcomes fail to refute h_j , one may still hope that the rest of the evidence, the sequence of all other experiments or observations, will bring the likelihood ratios for h_j over h_i near to 0. The next subsection describes conditions under which the rest of the evidence can be expected to highly refute h_j relative to h_i , evidence arising from possible outcomes that do not have likelihoods of 0 according to h_j unless their likelihoods are 0 for h_i , too.

3.2 THE RELATIVE REFUTATION OF HYPOTHESES BY PROBABILISTIC EVIDENCE

Theorem 6 will show that if the *average expected quality of information* (from a sequence of observations or experiments) for distinguishing between a true hypothesis and a competitor does not diminish to 0, and if the *average variance in the quality of information* is bounded above, then it is highly probable (as near 1 as you please) that a sequence of outcomes will occur that will drive the likelihood ratio of the competitor compared to the true hypothesis toward 0. Thus, probabilistic evidence will almost surely come to refute empirically distinct competitors to the true hypothesis to any desired degree, provided that a large enough number of observations can be made. In order to state Theorem 6 precisely I will first introduce an information-theoretic measure of the *expected quality of the information* for a possible experiment or observation, a measure of the potential power of an

experiment or measurement (or other *evidence gathering observation*) to discriminate between hypotheses. I will also introduce a measure of the *variance in the quality of information* for a sequence of experiments or observations. For a sequence of observations, the *variance* indicates how closely the *quality of information* for possible outcomes approximates the *expected quality of information*. I will briefly discuss these measures, and then state Theorem 6 and discuss its implications.

We will consider only sequences of observations for which each possible outcome that is assigned a non-zero likelihood by h_i is also assigned a non-zero likelihood by h_j ; Theorem 5 applies to any observations that violate this condition. The likelihood ratios $LR[e^n | i/j]$ will all be finite (and positive), since their denominators will always remain above 0 and their numerators are non-negative and less than or equal to 1.

Consider some particular sequence of outcomes e^n . The likelihood ratio $LR[e^n | i/j]$ effectively measures the extent to which the information content of e^n distinguishes between h_i and h_j . Likelihood ratios measure information content on a lopsided scale, a scale that ranges from 0 to infinity with the “midpoint” of the measure at 1. That is, $LR[e^n | i/j] = 1$ just when e^n does not distinguish at all between h_i and h_j , when h_i and h_j assign the same likelihoods to e^n . Likelihood ratios below 1 indicate that the evidence favors h_j over h_i . A likelihood ratio of .01, for example, indicates that h_j confers a likelihood on outcome e^n that is 100 times higher than that conferred by h_i , whereas a likelihood ratio of 100 would indicate the converse (i.e., that h_i makes the evidence 100 times more likely than does h_j). It will be convenient to employ a measure of the ability of e^n to empirically distinguish between hypotheses that is more symmetric than the raw likelihood ratios. The logarithm of the likelihood ratios provides just such a measure.

I will define $QI[e^n | i/j]$, the *quality of information supplied by e^n with regard to h_i over h_j (given $(c^n \cdot b)$)*, as the *base-2 logarithm* of the likelihood ratio $LR[e^n | i/j]$. Whereas the *quantity* of information for e^n is represented by ‘ n ’, QI measures its *quality*. The following relationships hold for QI :

$$\begin{aligned} QI[e^n | i/j] &= \log LR[e^n | i/j] = \log P[e^n | c^n \cdot h_i \cdot b] - \log P[e^n | c^n \cdot h_j \cdot b] \\ &= -\log LR[e^n | j/i] = -QI[e^n | j/i]. \end{aligned}$$

QI measures *information* on a scale that is symmetric about the natural midpoint 0. When h_i and h_j make the evidence equally likely, $QI[e^n | i/j] = 0$. That the logarithm is *base-2* simply means that if a likelihood ratio $LR[e^n | i/j]$ has a value equal to 2^r ($r \geq 0$), then $QI[e^n | i/j] = r$; and if $LR[e^n | i/j] = 1/2^r$, then $QI[e^n | i/j] = -r$. *Base-2* logarithms are commonly used in information-theoretic measures of information, in connection with measures of *binary bits* of information, but for my purposes nothing of substance will hang on the *base* of the log. QI is merely an alternative way to represent likelihood ratios. It measures the degree to which evidence plays a part in the support

or refutation of one hypothesis over another, but on a more symmetric scale.

Suppose we are contemplating whether to make some sequence of n observations, c^n . What value can we expect QI to take for these observations? Statisticians measure the *expected value* of a quantity by first multiplying each of its *possible values* by their respective probabilities of occurrence, and then summing these products. By this measure, the *expected value* of QI for h_i over h_j depends on the probabilities of the various *possible* outcome sequences e^n (in E^n). And the probability that any given sequence e^n occurs depends on which hypothesis is true. Thus, we will measure the *expected value of the quality of information* for observations c^n in support of h_i over h_j *relative to the truth of hypothesis h_i* . Define this function, $EQI^n[i/j \mid i]$ as follows:

$$EQI^n[i/j \mid i] = \sum_{[E^n]} QI[e^n \mid i/j] \times P[e^n \mid c^n \cdot h_i \cdot b].$$

Recall that in this subsection we are assuming that whenever h_j assigns a likelihood of 0 to a sequence e^n , so does h_i . By convention I will take the value of $QI[e^n \mid i/j] \times P[e^n \mid c^n \cdot h_i \cdot b]$ to be 0 when $P[e^n \mid c^n \cdot h_i \cdot b] = 0$, although technically in such cases $LR[e^n \mid i/j]$ is undefined since its denominator is 0. Assigning these products a value of 0 is just a way of excluding them from any influence on the value of $EQI^n[i/j \mid i]$. This is reasonable because if h_i is true, then such outcome sequences have 0 probability of occurring.

Whereas QI measures the ability of a particular *sequence of outcomes* to empirically distinguish between a pair of hypotheses, EQI^n is a measure of the ability of a *sequence of experiments or observations* to produce distinguishing outcomes. For a given sequence of observations c^n , the most likely values of QI are those that would result from the most likely possible outcome sequences e^n . Indeed, if h_i is true, then the most likely outcome sequences are those that will produce values of QI near EQI^n , as n increases (this claim can be proved by the same technique employed to prove Theorem 6).

The values of EQI^n *must* be positive if h_i is empirically distinct from h_j . That is, if for at least one possible sequence e^n , $P[e^n \mid c^n \cdot h_i \cdot b] \neq P[e^n \mid c^n \cdot h_j \cdot b] > 0$, then the values of $EQI^n[i/j \mid i]$ *must* be positive; otherwise, values of $EQI^n[i/j \mid i]$ must be 0, and the two hypotheses agree on the likelihoods of all possible outcomes of the first n observations. Savage proves this claim from a more general theorem about the expected values of logarithms of random variables.¹⁷

By averaging the values of $EQI^n[i/j \mid i]$ over the number of observations n , we obtain a measure of the *average expected quality of the information to be obtained from the n observations c^n* . I will denote this average by underlining 'EQI', thus:

$$\underline{EQI}^n[i/j \mid i] = EQI^n[i/j \mid i] \div n.$$

This function measures the degree to which *on average* each of the observations may be expected to produce outcomes that distinguish between h_i and

h_j , if h_i is true. Theorem 6 will merely assume that the value of $\underline{EQI}^n[i/j \mid i]$ does not become arbitrarily close to 0 as n increases, i.e., that the two hypotheses disagree by some minimal amount (on average) on the likelihoods of possible outcomes of observations. I will discuss this condition further after stating the theorem.

For any particular sequence of outcomes e^n , the *quality* of its distinguishing information (for distinguishing between a pair of hypotheses) will be some distance above or below the *expected quality of information* for the observations c^n . A measure of this distance is given by the square of the difference in values, $(QI[e^n \mid i/j] - \underline{EQI}^n[i/j \mid i])^2$. The *variance* of the quality of information (due to c^n , for distinguishing between h_i and h_j , given h_i) will be represented by ' $\underline{VQI}^n[i/j \mid i]$ '; it measures the *expected value* of these squared distances:

$$\underline{VQI}^n[i/j \mid i] = \sum_{[e^n]} (QI[e^n \mid i/j] - \underline{EQI}^n[i/j \mid i])^2 \times P[e^n \mid c^n, h_i, b].$$

Again the convention is that if $P[e^n \mid c^n, h_i, b] = 0$ for a given e^n , then the term $(QI[e^n \mid i/j] - \underline{EQI}^n[i/j \mid i])^2 \times P[e^n \mid c^n, h_i, b]$ equals 0. *Variance* is a common measure in statistics of the degree to which a quantity is spread out around its *expected value*. It is easy to see that $\underline{VQI}^n[i/j \mid i]$ will be positive unless h_i and h_j agree on the likelihoods of all possible sequences of outcomes in E^n , in which case both $\underline{EQI}^n[i/j \mid i]$ and $\underline{VQI}^n[i/j \mid i]$ equal 0.

The *average variance*, averaged with respect to the number of distinct observations, is defined as follows: $\underline{VQI}^n[i/j \mid i] = \underline{VQI}^n[i/j \mid i] \div n$. This function plays an important role in Theorem 6. I will discuss it in more detail after stating the theorem.

THEOREM 6: *The General Refutation Theorem.*

Suppose that there is a positive (lower bound) δ such that, for some (large enough) positive integer N , for every $n \geq N$, $\underline{EQI}^n[i/j \mid i] \geq \delta$.

Then, for any $m > 1$ (as large as you please), there is a positive integer M such that, for all $n \geq M$,

$$P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n, h_i, b] \geq 1 - \frac{1}{n} \times \frac{\underline{VQI}^n[i/j \mid i]}{(\underline{EQI}^n[i/j \mid i] - (m/n))^2}.$$

Suppose, in addition, that there is some constant $K > 1$ (K as large as you like) and some positive integer L such that, for all $n \geq L$, $\underline{VQI}^n[i/j \mid i] \leq (K-1) \times \underline{EQI}^n[i/j \mid i]^2$.

Then, for any $m > 1$ (as large as you please), there is a positive integer M such that, for all $n \geq M$,

$$P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n, h_i, b] > 1 - K/n, \text{ and}$$

$$\lim_n P[\forall \{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n \cdot h_i \cdot b] = 1.$$

(Proof is in the appendix.)

Theorem 6 implies that if h_i is true, and if, relative to h_i , the *average expected quality of the information* (for h_i over h_j) for a sequence of observations does not become arbitrarily small, and if the *average variance of the information* is bounded above (or at least does not outrun the square of the *average expected quality* by an unbounded amount), then a sufficient number of these observations will, with a probability as near to 1 as you like, produce a sequence of outcomes that will refute h_j relative to h_i to whatever degree is desired. Notice that the theorem places *explicit lower bounds* on the likelihood, given h_i , that the n observations will yield some sequence of outcomes that makes the likelihood ratio for h_j over h_i smaller than a *specific* fraction. And these lower bounds depend explicitly on the *average expected quality* and on the *average variance in quality* of the information for the n observations or experiments.

Theorem 6 does not depend on the assumption that likelihoods are objective; it holds for each probability function that satisfies the conditions on probabilistic structures. Of course, probability functions that disagree about likelihoods *may* end up disagreeing about *which* hypotheses become refuted by a particular sequence of outcomes, due to their disagreement about the empirical import of hypotheses. But, provided that probability functions agree on empirical import to the extent that their respective likelihood ratios for pairs of hypotheses lie within some large fixed multiple of one another, they must eventually agree among themselves about which hypotheses become highly refuted.

Theorem 6 does not assume that the outcomes of different observations are stochastically independent of one another. It only supposes that $EQI^n[i/j \mid i]$ never gets smaller than some positive constant δ , where δ may be chosen as near to 0 as you wish. However, stochastic independence *relative to a hypothesis* is not an implausible assumption (as I argued earlier). So, consider the special case in which evidence consists of independent outcomes relative to the hypotheses. The *expected quality of information* may be defined for each individual observation or experiment c_k as follows:

$$EQI_k[i/j \mid i] = \sum_{[o_k]} QI[o_{k,s} \mid i/j] \times P[o_{k,s} \mid c_k \cdot h_i \cdot b].$$

From the definition of EQI^n and the *independence of outcomes*, it follows that $EQI^n[i/j \mid i] = \sum_{k=1}^n EQI_k[i/j \mid i]$. So, for sequences of independent events, $EQI^n[i/j \mid i] = \sum_{k=1}^n EQI_k[i/j \mid i] \div n$. That this value is bounded below by some small $\delta > 0$ simply means that *on average* the values of the $EQI_k[i/j \mid i]$ are greater than δ (although $EQI_k[i/j \mid i]$ may be 0 for *some* values of k). This condition is easily satisfied if periodically some observation is made for which h_i and h_j disagree on the likelihoods of at least one of the outcomes by

some minimal amount, and if such outcomes have some minimal likelihood of occurring according to h_i . That is, there will be some $\delta > 0$ such that (for all n) $\underline{EQI}^n[i/j \mid i] > \delta$ *provided that*, for some lower bounds $\delta_1 > 0$ and $\delta_2 > 0$, periodically some observations are made (e.g., c_k) for which one of the *possible outcomes* (e.g., o) bears the following relationship to h_i and h_j :

$$P[o \mid c_k \cdot h_i \cdot b] > \delta_1, \text{ and either } LR[o \mid i/j] > \delta_2 \text{ or } LR[o \mid j/i] > \delta_2.$$

(The supposition that *such observations* are made “periodically” just means that in the sequence of observations C for the induction structure, the frequency with which such observations occur does not diminish to 0 as n increases. It would be best if *every* observation *could* distinguish the two hypotheses by at least some minimal amount; but, it will suffice if such observations do not become increasingly scarce among the sequence of all observations.)

When the outcomes of different observations are stochastically independent of one another (relative to the hypotheses), the *variance* for the observations c^n is just the sum of the *variances* for the individual observations, c_k . That is, the individual variances for $QI[c_k \mid i/j]$, for each c_k (given that h_i is true), are given by:

$$VQI_k[i/j \mid i] = \sum_{[o_{k,s}]} (QI[o_{k,s} \mid i/j] - EQI_k[i/j \mid i])^2 \times P[o_{k,s} \mid c_k \cdot h_i \cdot b].$$

If the evidence consists of independent events (relative to h_i and to h_j), then it can be shown that $VQI^n[i/j \mid i] = \sum_{k=1}^n VQI_k[i/j \mid i]$, hence, $\underline{VQI}^n[i/j \mid i] = \sum_{k=1}^n VQI_k[i/j \mid i] \div n$. Thus, the assumption in Theorem 6 that $\underline{VQI}^n[i/j \mid i]$ doesn't get too large (compared to $\underline{EQI}^n[i/j \mid i]^2$) will be satisfied if the *variance in the quality of information* for *each* independent observation does not become larger than some fixed (perhaps quite large) upper bound.

When the possible outcomes for the sequence of observations are not only independent, but also identically distributed, Theorem 6 reduces to a version of the theorem proved by Savage. Identically distributed outcomes most commonly result from the repetition of identical statistical experiments (e.g., repeated tosses of a coin, or a series of identical measurements on quantum systems prepared in identical states). For such experiments a hypothesis should specify the same likelihoods for the same kinds of outcomes from one observation to the next. If each independent observation is distributed in the same way, then $EQI_k[i/j \mid i]$ and $VQI_k[i/j \mid i]$ are positive constants (for all k). So, $\underline{EQI}^n[i/j \mid i]$ and $\underline{VQI}^n[i/j \mid i]$ will be positive constants, equal to $EQI_k[i/j \mid i]$ and $VQI_k[i/j \mid i]$, respectively. Thus, the conditions for Theorem 6 are satisfied automatically for a sequence of independent, identically distributed outcomes. However, Theorem 6 is *much more general*. The theorem holds even when the sequence of observations encompasses completely unrelated observations and experiments, observations with nothing in common but that they are empirically related to a common set of competing theories or hypotheses.

Through Theorems 1, 5, and 6, each hypothesis says that, given enough observations, *it* will very likely dominate its empirically distinct competitors in a contest of likelihoods. And even a sequence of observations with an extremely low *average expected quality of information* is likely to do the job. Presumably the true hypothesis, if present, speaks truthfully about this, and its competitors lie.

CONCLUSION

I have argued that Bayesian induction, in so far as it is objective, is essentially a form of induction by elimination. The likelihood ratios are the mechanism through which evidence can eliminate false hypotheses. Theorem 4 (together with Theorems 2 and 3) implies that the only general way in which the influence of the values of prior probabilities for hypotheses can be overcome by the evidence (i.e., washed out) is for evidence to increasingly refute hypotheses. If all competitors of a hypothesis become increasingly refuted, then its posterior probability will be driven to 1, but this is the only general way in which agreement can be achieved for the posterior probability of a hypothesis that is not increasingly refuted. Theorems 5 and 6 assure us that if hypotheses are empirically different enough (so that the quality of information for the sequence of observations does not become arbitrarily small), then, very probably, the evidence *will* increasingly refute false hypotheses relative to the true hypothesis.

What do these formal results portend for inductive practice? They do seem to suggest a certain inductive strategy for the evaluation of scientific theories. The strategy they suggest, though, is nothing any fancier than common sense scientific method: continually develop and test alternative hypotheses against one another. In testing one hypothesis against another, one should make the kinds of observations and conduct the kinds of experiments that, if *either* hypothesis is true, are likely to produce outcomes that are much more unlikely according to one hypothesis than according to the other (e.g., observations and experiments for which $EQI[i/j \mid i]$ and $EQI[j/i \mid j]$ are large, and $VQI[i/j \mid i]$ and $VQI[j/i \mid j]$ are small). If the true hypothesis is never considered, then of course no mode of reasoning can come to support it. But if the true hypothesis ever does come under consideration, then Theorems 5 and 6 assure us that empirically distinct competitors will almost surely be eliminated by a long enough sequence of observations. Thus, for any way of assigning *relative* initial probabilities to hypotheses, the true hypothesis (perhaps along with some of its empirically equivalent competitors) will eventually obtain *relative* posterior probabilities much larger than empirically distinct alternatives, as described by equation (I). In this way the true hypothesis (and its empirical equivalents) will

eventually become the most likely hypothesis, and will continue to survive all challenges.

In Bayesian induction, empirically equivalent hypotheses maintain the relative values for posterior probabilities established by their initial probabilities, according to equation (I). Thus, a true hypothesis can become highly confirmed only if the initial plausibilities of its empirically equivalent competitors are judged to be much lower than the initial plausibility of the true hypothesis. Some sort of plausibility considerations always do play a role in the decisions scientists make about which hypotheses to develop and test. These considerations often involve ontological presuppositions native to the subject matter.¹⁸ Prior probabilities merely provide a slot in the Bayesian machine into which initial plausibilities can be inserted. The values of initial plausibilities continue to significantly influence the values of posterior probabilities unless (and until) sufficient evidence is forthcoming to eliminate alternatives. Indeed, initial plausibilities for hypotheses may even be changed over time (contrary to the usual recommendation of Bayesians); i.e., one may switch from probability function P_α to P_β if new plausibility considerations dictate. Nothing in Theorems 5 and 6 requires that one hold to fixed values for initial plausibilities. The evidence does the same work on *likelihood ratios* regardless of the values for priors. Provided that such shifts among prior probability assignments do not repeatedly discount the true hypothesis by assigning it ever lower prior probabilities (approaching 0), the true hypothesis will very probably come to be rated as much more plausible on the evidence than any of its empirically distinct competitors.

APPENDIX: PROOFS OF THEOREMS 4, 5, AND 6

THEOREM 4: *Non-Zero Convergence is Convergence to One.*

Consider any *probabilistic induction structure* $\langle S, P \rangle$. Let h_i be some hypothesis in H , and suppose there is a P_α in P that satisfies the following conditions:

- i) there is a real number r such that, for all n , $P_\alpha[h_i \mid e^n \cdot c^n \cdot b] \geq r > 0$, and
- ii) there is a P_β that *modestly differs* with P_α relative to h_i , and
- iii) $\lim_n |P_\alpha[h_i \mid e^n \cdot c^n \cdot b] - P_\beta[h_i \mid e^n \cdot c^n \cdot b]| = 0$.

Then $\lim_n P_\alpha[h_i \mid e^n \cdot c^n \cdot b] = 1$.

Proof: Assume that the antecedent of the theorem holds. Without loss of generality, we may suppose that P_β modestly differs from P_α in the following way, for some real number $M > 1$:

$P_\beta[h_i \mid b] > 0$; and
for each alternative h_j to h_i (in H), $P_\beta[h_j \mid b] \geq M \times P_\alpha[h_j \mid b] > 0$.

Clearly, $P_\alpha[h_i | b] > P_\beta[h_i | b]$. Also notice that

$$0 < P_\beta[h_i | b] = 1 - \sum_{j \neq i} P_\beta[h_j | b] \leq 1 - M \times \sum_{j \neq i} P_\alpha[h_j | b] = 1 - M \times (1 - P_\alpha[h_i | b]) = M \times P_\alpha[h_i | b] - (M-1).$$

Now calculate the differences in the ratios of initial probabilities, as follows:

$$\begin{aligned} \frac{P_\beta[h_j | b]}{P_\beta[h_i | b]} - \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} &\geq \frac{M \times P_\alpha[h_j | b]}{M \times P_\alpha[h_i | b] - (M-1)} - \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} \\ &= \frac{(M-1)}{M \times P_\alpha[h_i | b] - (M-1)} \times \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} \geq 0. \end{aligned}$$

The factor $(M-1) \div [M \times P_\alpha[h_i | b] - (M-1)]$ is a positive constant; call it 'K'.

Now the strategy of the proof is to show that for any $\varepsilon > 0$, there is a $\delta = K \times r^2 \times \varepsilon / (1 + K \times r \times \varepsilon) > 0$ such that

$$\text{if } |P_\alpha[h_i | e^n \cdot c^n \cdot b] - P_\beta[h_i | e^n \cdot c^n \cdot b]| \leq \delta, \text{ then } \Omega_\alpha[\neg h_i | e^n \cdot c^n \cdot b] \leq \varepsilon.$$

Once this is proved, it follows immediately that

$$\text{if } \lim_n |P_\alpha[h_i | e^n \cdot c^n \cdot b] - P_\beta[h_i | e^n \cdot c^n \cdot b]| = 0, \text{ then } \lim_n \Omega_\alpha[\neg h_i | e^n \cdot c^n \cdot b] = 0, \text{ so } \lim_n P_\alpha[h_i | e^n \cdot c^n \cdot b] = 1.$$

Suppose $\delta \geq |P_\alpha[h_i | e^n \cdot c^n \cdot b] - P_\beta[h_i | e^n \cdot c^n \cdot b]|$, where δ is as above. Then $P_\beta[h_i | e^n \cdot c^n \cdot b] \geq P_\alpha[h_i | e^n \cdot c^n \cdot b] - \delta \geq r - \delta = r/(1 + K \times r \times \varepsilon)$ (since $P_\alpha[h_i | e^n \cdot c^n \cdot b] \geq r$). So, $1/(P_\alpha[h_i | e^n \cdot c^n \cdot b] \times P_\beta[h_i | e^n \cdot c^n \cdot b]) \leq (1 + K \times r \times \varepsilon)/r^2$. The latter relationship will play a role in the following derivation.

$$\begin{aligned} \text{It is easy to check that } |P_\alpha[h_i | e^n \cdot c^n \cdot b] - P_\beta[h_i | e^n \cdot c^n \cdot b]| &= |\Omega_\beta[\neg h_i | e^n \cdot c^n \cdot b] - \Omega_\alpha[\neg h_i | e^n \cdot c^n \cdot b]| \times P_\alpha[h_i | e^n \cdot c^n \cdot b] \times P_\beta[h_i | e^n \cdot c^n \cdot b]. \\ \text{So, } K \times \varepsilon \geq \delta / (P_\alpha[h_i | e^n \cdot c^n \cdot b] \times P_\beta[h_i | e^n \cdot c^n \cdot b]) &\geq |\Omega_\beta[\neg h_i | e^n \cdot c^n \cdot b] - \Omega_\alpha[\neg h_i | e^n \cdot c^n \cdot b]| \\ &= \left| \sum_{j \neq i} L[e^n | j/i] \times \left[\frac{P_\beta[h_j | b]}{P_\alpha[h_i | b]} - \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} \right] \times \frac{P_\alpha[c^n | h_j \cdot b]}{P_\alpha[c^n | h_i \cdot b]} \right| \\ &\geq \left| \sum_{j \neq i} L[e^n | j/i] \times K \times \frac{P_\alpha[h_j | b]}{P_\alpha[h_i | b]} \times \frac{P_\alpha[c^n | h_j \cdot b]}{P_\alpha[c^n | h_i \cdot b]} \right|. \end{aligned}$$

Thus, $\varepsilon \geq \Omega_\alpha[\neg h_i | e^n \cdot c^n \cdot b]$. \square

THEOREM 5: The Special Refutation Theorem.

If for all e^k such that $P[e^k | c^k \cdot b \cdot h_j] > 0$, the following two conditions hold:

- 1) $P[e^k | c^{k+1} \cdot b \cdot h_j] = P[e^k | c^k \cdot b \cdot h_j]$, and similarly for h_i ;
- 2) $P[\neg r[o_{k,r} | c^k \cdot e^{k-1} \cdot b \cdot h_j] = 0] | c^k \cdot e^{k-1} \cdot h_i \cdot b] \geq \delta > 0$,

then $P[\vee\{e^n \mid LR[e^n \mid j/i] = 0\} \mid c^n \cdot h_i \cdot b] \geq 1 - (1-\delta)^n$,

and thus, $\lim_n P[\vee\{e^n \mid LR[e^n \mid j/i] = 0\} \mid c^n \cdot h_i \cdot b] = 1$.

Proof: Assume that the antecedent of the theorem holds. Notice that, for each n , $P[e^n \mid c^n \cdot b \cdot h_i] = \prod_{k=1}^n P[e_k \mid c^k \cdot e^{k-1} \cdot b \cdot h_i]$, and similarly for h_j . So (with all terms for which $P[e^n \mid c^n \cdot b \cdot h_i] = 0$ deleted from the sums):

$$\begin{aligned} P[\vee\{e^n \mid LR[e^n \mid j/i] \neq 0\} \mid c^n \cdot h_i \cdot b] &= \sum_{\{e^n \mid LR[e^n \mid j/i] \neq 0\}} P[e^n \mid c^n \cdot h_i \cdot b] \\ &= \sum_{\{e^n \mid P[e^n \mid c^n \cdot h_i \cdot b] \neq 0\}} (\prod_{k=1}^n P[e_k \mid c^k \cdot e^{k-1} \cdot b \cdot h_i]) \\ &= \sum_{\{e^n \mid \prod_{k=1}^n P[e_k \mid c^k \cdot e^{k-1} \cdot h_j \cdot b] \neq 0\}} (\prod_{k=1}^n P[e_k \mid c^k \cdot e^{k-1} \cdot b \cdot h_i]) \\ &= \sum_{\{e^n \mid \text{for all } ok,r \text{ a conjunct of } e^n, P[ok,r \mid c^k \cdot e^{k-1} \cdot h_j \cdot b] \neq 0\}} \prod_{k=1}^n P[e_k \mid c^k \cdot e^{k-1} \cdot b \cdot h_i] \\ &= \prod_{k=1}^n (\sum_{\{ok,r \mid P[ok,r \mid c^k \cdot e^{k-1} \cdot b \cdot h_j] \neq 0\}} P[ok,r \mid c^k \cdot e^{k-1} \cdot b \cdot h_i]) \\ &= \prod_{k=1}^n P[\vee_r \{ok,r \mid P[ok,r \mid c^k \cdot e^{k-1} \cdot b \cdot h_j] \neq 0\} \mid c^k \cdot e^{k-1} \cdot h_i \cdot b] \\ &\leq \prod_{k=1}^n (1-\delta) = (1-\delta)^n. \end{aligned}$$

The rest follows easily. \square

THEOREM 6: *The General Refutation Theorem.*

Suppose that there is a positive (lower bound) δ such that, for some (large enough) positive integer N , for all $n \geq N$, $\underline{EQI}^n[i/j \mid i] \geq \delta$. Then, for any $m > 1$ (as large as you please), there is a positive integer M such that, for all $n \geq M$,

$$P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n \cdot h_i \cdot b] \geq 1 - \frac{1}{n} \times \frac{\underline{VQI}^n[i/j \mid i]}{(\underline{EQI}^n[i/j \mid i] - (m/n))^2}$$

Suppose, in addition, that there is some constant $K > 1$ (K as large as you like) and some positive integer L such that, for all $n \geq L$, $\underline{VQI}^n[i/j \mid i] \leq (K-1) \times \underline{EQI}^n[i/j \mid i]^2$. Then, for any $m > 1$ (as large as you please), there is a positive integer M such that, for all $n \geq M$, $P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n \cdot h_i \cdot b] > 1 - K/n$, and $\lim_n P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n \cdot h_i \cdot b] = 1$.

Proof: I will derive the inequality in the first part of the theorem, and then treat the second part.

Given any m , choose M large enough so that *both* $M \geq N$ (i.e., for all $n \geq M$, $\underline{EQI}^n[i/j \mid i] \geq \delta$) *and* $\delta - (m/M) > 0$ (i.e., $M > m/\delta$). Then for each $n \geq M$, we have $\underline{EQI}^n[i/j \mid i] - (m/n) \geq \underline{EQI}^n[i/j \mid i] - (m/M) \geq \delta - (m/M) > 0$.

Now, by definition,

$$\underline{VQI}^n[i/j \mid i] = \sum_{[e^n]} (QI[e^n \mid i/j] - \underline{EQI}^n[i/j \mid i])^2 \times P[e^n \mid c^n \cdot h_i \cdot b].$$

So, $(1/n) \times \underline{VQI}^n[i/j \mid i]$

$$= \sum_{[e^n]} ((QI[e^n \mid i/j]/n) - \underline{EQI}^n[i/j \mid i])^2 \times P[e^n \mid c^n \cdot h_i \cdot b]$$

$$\begin{aligned} &\geq \sum_{\{e^n \mid QI[e^n \mid i/j] \leq m\}} (\underline{EQI}^n[i/j \mid i] - (QI[e^n \mid i/j]/n))^2 \times P[e^n \mid c^n \cdot h_i \cdot b] \\ &\geq \sum_{\{e^n \mid QI[e^n \mid i/j] \leq m\}} (\underline{EQI}^n[i/j \mid i] - (m/n))^2 \times P[e^n \mid c^n \cdot h_i \cdot b]. \end{aligned}$$

Then, $(1/n) \times \underline{YQI}^n[i/j \mid i] / (\underline{EQI}^n[i/j \mid i] - (m/n))^2$

$$\begin{aligned} &\geq \sum_{\{e^n \mid QI[e^n \mid i/j] \leq m\}} P[e^n \mid c^n \cdot h_i \cdot b] \\ &= P[\vee\{e^n \mid QI[e^n \mid i/j] \leq m\} \mid c^n \cdot h_i \cdot b] \\ &= P[\vee\{e^n \mid LR[e^n \mid i/j] \leq 2^m\} \mid c^n \cdot h_i \cdot b] \\ &= 1 - P[\vee\{e^n \mid LR[e^n \mid i/j] > 2^m\} \mid c^n \cdot h_i \cdot b] \\ &= 1 - P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n \cdot h_i \cdot b]. \end{aligned}$$

This completes the proof of the first part of the theorem.

For the proof of the second part, assume the antecedent of the second part of the theorem. We can make M (as defined earlier in the proof) large enough (i.e., $M \geq L$) so that, for some positive integer K , for every $n \geq M$, $\underline{YQI}^n[i/j \mid i] \leq (K-1) \times \underline{EQI}^n[i/j \mid i]^2$.

Since δ , m , and K are constants, we can also make sure to choose M large enough that $0 < 2 \times (m/M) \times K \leq \delta \leq \underline{EQI}^n[i/j \mid i]$ for every $n \geq M$. Then, for every $n \geq M$, $0 < 2 \times (m/n) \times K \leq \delta \leq \underline{EQI}^n[i/j \mid i]$. It follows that $\underline{EQI}^n[i/j \mid i] - (m/n) \times K \geq (m/n) \times K > 0$, so $(\underline{EQI}^n[i/j \mid i] - (m/n) \times K)^2 \geq ((m/n) \times K)^2 > (m/n)^2 \times K \times (K-1)$;

$$\underline{EQI}^n[i/j \mid i]^2 - 2(m/n) \times K \times \underline{EQI}^n[i/j \mid i] + (m/n)^2 \times K^2 > (m/n)^2 \times K^2 - (m/n)^2 \times K; [(m/n)^2 - 2(m/n) \times \underline{EQI}^n[i/j \mid i]] \times K > -\underline{EQI}^n[i/j \mid i]^2.$$

Therefore, $(\underline{EQI}^n[i/j \mid i] - (m/n))^2 \times K =$

$$\begin{aligned} &\underline{EQI}^n[i/j \mid i]^2 \times K + [(m/n)^2 - 2(m/n) \times \underline{EQI}^n[i/j \mid i]] \times K \\ &> \underline{EQI}^n[i/j \mid i]^2 \times K - \underline{EQI}^n[i/j \mid i]^2 = (K-1) \times \underline{EQI}^n[i/j \mid i]^2 \geq \\ &\underline{YQI}^n[i/j \mid i]. \end{aligned}$$

And so, $P[\vee\{e^n \mid LR[e^n \mid j/i] < 1/2^m\} \mid c^n \cdot h_i \cdot b]$

$$\begin{aligned} &\geq 1 - (1/n) \times \underline{YQI}^n[i/j \mid i] / (\underline{EQI}^n[i/j \mid i] - (m/n))^2 \\ &> 1 - (1/n) \times K. \quad \square \end{aligned}$$

NOTES

1. I wish to thank Chris Swoyer and Chris Hill for their many helpful comments on drafts of this paper. I owe an older and deeper debt to John Earman, Daniel Garber, William Hanson, and Grover Maxwell for discussions of Bayesian induction that go back over fifteen years. Their insights largely shaped the research that led to this paper.
2. C. Howson and P. Urbach, *Scientific Reasoning: the Bayesian Approach* (La Salle: Open Court, 1989).
3. See L. J. Savage, *The Foundations of Statistics* (New York: Dover, 1972).

4. J. Earman, *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory* (Cambridge, Mass.: MIT Press 1992).
5. See M. Hesse, "Bayesian Methods and the Initial Probability Of Theories," in G. Maxwell and R. Anderson, eds., *Induction, Probability, and Confirmation*, Minnesota Studies in the Philosophy of Science, vol. 6 (Minneapolis: University of Minnesota Press, 1975).
6. See *Bayes or Bust?*, 163.
7. See *The Foundations of Statistics*, 46–50.
8. See K. Popper, *The Logic of Scientific Discovery* (New York: Harper & Row, 1968).
9. See, for instance, the works of Savage (note 3) and Howson and Urbach (note 2) for Bayesian justifications of classical probability theory.
10. See B. de Finetti, "La Prevision: Ses Lois Logiques, Ses Sources Subjectives," *Annales de L'Institut Henri Poincare* 7 (1937): 1–68. English translation in H. Kyburg and H. Sokler, eds., *Studies in Subjective Probability* (New York: John Wiley, 1964). Also, see the works by Savage (note 3) and Howson and Urbach (note 2).
11. D. Lewis, "A Subjectivist's Guide to Objective Chance," in R. Jeffrey, ed., *Studies in Inductive Logic and Probability*, vol. 2 (Berkeley: University of California Press, 1980).
12. See R. Carnap, *The Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950).
13. H. Kyburg, Jr., *The Logical Foundations of Statistical Inference* (Dordrecht: Reidel, 1974).
14. See W. Harper, "Kyburg on Direct Inference," in R. Bodgan, ed., *Henry E. Kyburg, Jr. and Isaac Levi* (Dordrecht: Reidel, 1981), 97–127.
15. See Kyburg's reply to Harper in his "Reply," in *Henry E. Kyburg, Jr. and Isaac Levi*. For an insightful exchange on the logic of direct inference probabilities also see I. Levi, "Direct Inference," *The Journal of Philosophy* 74 (1977): 5–29; H. Kyburg, Jr., "Randomness and the Right Reference Class," *The Journal of Philosophy* 74 (1977): 501–20; I. Levi, "Confirmational Conditionalization," *The Journal of Philosophy* 75 (1978): 730–37; and T. Seidenfeld, "Direct Inference and Inverse Inference," *The Journal of Philosophy* 75 (1978): 709–30.
16. See *The Foundations of Statistics*, 46–50.
17. See *The Foundations of Statistics*, 49 and 268.
18. See Earman's discussion of eliminative induction applied to gravitational theories in *Bayes Or Bust?*, ch. 7.