

***Giving up Judgment Empiricism:  
The Bayesian Epistemology of  
Bertrand Russell and Grover Maxwell***

*Human Knowledge: Its Scope and Limits* was first published in 1948.<sup>1</sup> The view on inductive inference that Russell develops there has received relatively little careful study and, I believe, has been largely misunderstood. Grover Maxwell was one of the few philosophers who understood and carried forward the program of Russell's later work. Maxwell considered *Human Knowledge* to be one of Russell's most significant works for its treatment of perception, the event ontology, the theory of space-time, the philosophy of mind, and especially for its solution of the mind-body problem. Furthermore, Maxwell wholly agreed with Russell's rejection there of judgment empiricism (the doctrine that all contingent knowledge can be validated on the basis of experience and [noncontingent] logic alone). But with regard to a positive account of inductive inference Maxwell, with his Bayesian version of inductive logic, parted ways with Russell, or so he believed. Of Russell's positive account Maxwell says:

He takes the untested, untestable, and, in this sense, non-empirical (though nevertheless, contingent) assumptions upon which our significant knowledge of the world and ourselves rest to be his notorious six "Postulates of Scientific Inference." But, in spite of my boundless admiration for Russell's later work, I do not think that he ever used these postulates significantly or ever showed how they could do much for anyone, be he scientist, philosopher, or man-in-the-street.<sup>2</sup>

Maxwell's comment is typical of how Russell's positive account of induction is usually understood. Everyone knows that Russell's account involves the contingent postulates, but there is little understanding of the precise role they are supposed to play.

This essay is an attempt to gain better insight into Russell's positive account of inductive inference. I contend that Russell's postulates play only a supporting role in his overall account. At the center of Russell's positive view is a probabilistic, Bayesian model of inductive inference. Indeed, Russell and Maxwell actually held very similar Bayesian views. But the Bayesian component of Russell's view

in *Human Knowledge* is sparse and easily overlooked.<sup>3</sup> Maxwell was not aware of it when he developed his own view, and I believe he was never fully aware of the extent to which Russell's account anticipates his own. The primary focus of this paper will be the explication of the Bayesian component of the Russell-Maxwell view, and the way in which it undermines judgment empiricism.

Bayes's theorem is an equality derivable in probability theory. By a Bayesian inductive logic I mean any view of theory confirmation that takes Bayes's theorem as expressing the essential features of the logic of theory confirmation. Bayes' Theorem may be written as follows:

$$P(T_i/E) = \frac{P(E/T_i) P(T_i)}{\sum_i P(E/T_i) P(T_i)}$$

where  $\sum_i P(T_i) = 1$ , and where for each pair  $T_i$  and  $T_j$ ,  $i \neq j$ ,  $P(T_i \cdot T_j) = 0$ . If we think of  $T_1, T_2, \dots$ , as an exhaustive list of mutually incompatible theories and  $E$  as a statement of the evidence, then Bayes's theorem says that the probability of a theory, say  $T_1$ , on the evidence  $E$  is determined by the likelihood which each  $T_i$  assigns to  $E$ ,  $P(E/T_i)$ , and the probability of each  $T_i$  prior to the evidence,  $P(T_i)$ . The terms  $P(T_i)$  are often called the prior (or a priori) probabilities of the  $T_i$ 's.

By my criteria many probabilistic accounts of inductive logic are Bayesian, but not all. Among Bayesian views are subjectivist and personalist versions (see, e.g., Frank P. Ramsey, Bruno De Finetti, and L. J. Savage),<sup>4</sup> logicist accounts (e.g., John Maynard Keynes, Harold Jeffreys, Rudolf Carnap),<sup>5</sup> and frequentist views (e.g., Hans Reichenbach, Wesley Salmon, and Grover Maxwell).<sup>6</sup> Maxwell's Bayesian views derive chiefly from Salmon. Russell's account comes from Keynes. Among the non-Bayesian probabilistic models of inductive inference are the classical statistical accounts of R. A. Fisher, J. Neyman and E. S. Pearson, and the logicist account of Henry Kyburg.<sup>7</sup>

What I call the Russell-Maxwell view on inductive inference may be expressed as steps in an enthymematic argument. The view takes the major premise to be fairly obvious:

1. Inductive inference is probabilistic.

A corollary of this premise is:

1a. Any adequate logic of inductive inference must be formally expressible in probability theory and justified with an account of its soundness (i.e., of how and why it may be expected to lead from true premises to probably true conclusions). In particular, the logic should give an account of how the confirmation of scientific theories by collections of singular evident events is possible.

Next an investigation of the logic of probability theory leads to an intermediate conclusion:

2. The theory of probabilistic inference, and Bayes's theorem in particular, shows that valid inductive inference may be formulated in terms of a purely logical component (i.e., the logic of probability theory) and an appropriate assignment of values for prior probabilities.

Russell and Maxwell investigate non-Bayesian inductive methods as well (e.g., induction by simple enumeration). They conclude that the success of any inductive method must depend for its soundness on contingent presuppositions. But both of them see Bayes's theorem as expressing the most comprehensive account of the logic of the confirmation of scientific theories. If purely logical considerations could furnish the right kind of prior probabilities, judgment empiricism might be saved. This leads to the next step in the view. A minor premise is argued for:

3. Logic alone neither guides nor justifies assigning the kinds of values for prior probabilities that are required for the soundness of inductive inferences.

This leads to the conclusion:

4. The valid application of the logic of inductive inference must presuppose some contingent principles to guide and justify the right kinds of prior probabilities. And, on pain of circularity, these principles cannot be justified inductively.

Russell formulated his postulates as an attempt to state the principles presupposed by inductive inference. Maxwell fills the role of the postulates with a different contingent principle, which I will briefly describe later. Their disagreement over the form of the contingent principles is the only major difference in their views. There are, of course, relatively minor differences in the details of their arguments (e.g., they have different accounts of the interpretation of probability). In this essay I will primarily follow Russell's version.

The first hint at the Bayesian flavor of Russell's view, and the role his postulates will play in it, surfaces in the introduction to *Human Knowledge*. The final paragraph of the introduction begins:

That scientific inference requires, for its validity, principles which experience cannot render even probable is, I believe, an inescapable conclusion from the logic of probability. (HK, p. xv)

How does the logic of probability require that inductive inference depend on unconfirmable contingent principles? Russell's answer comes in two parts. The first part claims that inductive inference must be probabilistic; the second claims

that probabilistic inference, where valid, depends on an appropriate selection of prior probabilities, and logic alone cannot suffice for this selection.

Regarding the first part Russell writes:

It is generally recognized that the inferences of science and common sense differ from those of deductive logic and mathematics in a very important respect, namely, that when the premises are true and the reasoning correct, the conclusion is only probable. (HK, p. 335)

He then proceeds to investigate various accounts of what probability *is*. He concludes that there are two kinds: *mathematical probability* and *degree of credibility*. Mathematical probabilities assert the relative frequency with which members of one class occur in another class. Degree of credibility, Russell says, is much more widely applicable than mathematical probability:

Any proposition concerning which we have rational grounds for some degree of belief or disbelief can, in theory, be placed in a scale between certain truth and certain falsehood. (HK, p. 381)

Probability as degree of credibility represents this scale of *rational* belief. It is a logical tool that should guide our degree of subjective (i.e., psychological) certainty in the truth of propositions. The totality of a person's actual beliefs may not be logically consistent. A logic is a normative standard of rationality, not a description of one's actual belief set. A person falls short of the standard for, among other reasons, lack of logical omniscience. Russell sees the relationship between subjective certainty (which he also calls psychological probability) and degrees of credibility to be strictly analogous to this relationship between actual belief and logical consistency. Just as two sets of beliefs may be inconsistent with each other but internally consistent, so two rational credibility functions may assign different, but internally consistent (with the laws of probability theory), degrees of credibility to propositions. The logic of credibility functions, as expressed by the laws of probability, describes necessary conditions for one's subjective degree of certainty in various propositions to hang together in a coherent way.

Modern Bayesians bolster the view that a notion like degree of credibility is a rational guide to belief and action by arguing that any betting system that violates probabilistic laws can be "Dutch booked" (i.e., the system will accept as fair a system of bets that logically cannot win, no matter what the outcome of the gamble). If one's subjective degree of certainty is to guide one's actions in a rational way, then one may only be certain of avoiding Dutch book by bringing one's subjective degrees of certainty into line with some rational credibility function. In this way Dutch book arguments tie together the ideas of probabilistic credibility as a logic and a guide to life. But Russell seems unaware of such arguments.

Russell finds a certain connection between mathematical probability and degree of credibility:

The connection is this: When, in relation to all available evidence, a proposition has a certain mathematical probability, then this measures its degree of credibility. For instance, if you are about to throw dice, the proposition "Double sixes will be thrown" has only one thirty-fifth of the credibility attaching to the proposition "Double sixes will not be thrown." Thus the rational man, who attaches to each proposition the right degree of credibility, will be guided by the mathematical theory of probability *when it is applicable*. (HK, p. 381)

This connection is an important one. Carnap calls it *direct inference*. A direct inference is an inference from a statistical description of a whole population to the likelihood that a sample from the population exhibits some specified attributes. The relationship is expressed in terms of conditional credibility functions.

In Russell's example the population is the set of (possible) throws of the dice, and the statistical description says that the relative frequency of double sixes among throws is one thirty-sixth. Let  $T$  be the set of throws of the dice, and let  $D$  be the subset of  $T$  which consists of all double sixes. The statistical description may be represented by " $F(D, T) = 1/36$ " (i.e., the frequency of  $D$ 's among  $T$ 's is  $1/36$ ). Then, where  $P$  is a rational credibility function and  $n$  is the next toss, " $P(n \in D/n \in T \cdot F(D, T) = 1/36) = 1/36$ " expresses a direct inference. It says that the rational degree of credibility of the next throw,  $n$ , being a double six, given the statistical hypothesis,  $F(D, T) = 1/36$  (along with the fact that  $n$  is a throw, i.e.,  $n \in T$ ) is  $1/36$ . There are, of course, more complex cases of direct inference. For example, let  $m$  be another toss of the dice:  $P(n \in D \cdot m \notin D/n \in T \cdot m \in T \cdot F(D, T) = 1/36) = 1/36 \times 35/36 = .027$ .

Direct inference is sometimes called statistical syllogism. It is a straightforward extension of the deductive entailment relation. It is a relation of partial entailment of an instance (or conjunction of instances) by a more general hypothesis.

There are a multitude of differing rational credibility functions. Given two arbitrary sentences  $A$  and  $B$ , in general there will be rational credibility functions  $P_\alpha$  and  $P_\beta$  such that  $P_\alpha(A/B) \neq P_\beta(A/B)$ . In general the degree to which one sentence makes another credible is not logically determinate. Two rational people can honestly disagree on the degree to which a sentence is credible without either being internally inconsistent. The logic of probability only supplies constraints that must be satisfied for a credibility function to be internally consistent. But all rational credibility functions should agree in cases of direct inference. For, supposing the truth of a statistical hypothesis,  $H$ , about a population and given an event description,  $E$ , which describes an instance of the statistical population, the only rational way to assign  $E$  credibility on the basis of  $H$  is to assign it the value  $H$  specifies.

Recent work on the logic of statistical inference largely consists of attempts to supplement the standard probabilistic axioms for credibility functions with axioms that would require all credibility functions to agree on direct inferences. In these investigations a precise syntactic system (like the language of first-order logic, or an extension of it) is constructed so that statistical statements have a perspicuous syntactic structure. Then the direct inferences are specified in terms of the logical structure of the sentences involved. For our purposes it is not important whether such formal attempts succeed. What is important is that any statistical scientific theory worth its salt plays the role of premises (on the right-hand side of conditional credibility functions) in direct inferences about the evidence. All rational credibility functions should agree on the degree of credibility afforded the evidence by a given statistical theory or hypothesis. For nonstatistical theories the only direct inferences come from logical entailments, and the degree of credibility (in a direct inference) afforded evidence by a theory is either 1 or 0 (depending on whether the theory entails the evidence or its negation, respectively).

In this essay I will express credibility functions as a subscripted letter “*P*” (e.g.,  $P_\alpha, P_\beta, P_\gamma, \dots$ ). The subscript is to remind us that credibility functions may differ on the degree of credibility they assign a sentence. But since all credibility functions agree on the degree of credibility for a direct inference, I will mark direct inferences by dropping the subscript. So if  $P_\alpha(E/H)$  is a direct inference I will write “ $P(E/H)$ .”

We are now prepared to see how the theory of probability shows that inductive inference may be formulated so as to depend for its soundness only on logic and appropriate values for prior probabilities. I will explicate two Bayesian arguments investigated by Russell. Russell draws the first from Keynes.<sup>8</sup> I will generalize it slightly, and employ a more readable notation (see HK, pp. 408–10 and 435–37).

Let  $T_1$  be a theory that taken together with initial condition statements  $a_1, a_2, \dots$ , and other relevant background knowledge  $g$ , logically entails the observable evidential statements  $e_1, e_2, \dots$ , respectively. After  $n$  observations Bayes’s formula for the credibility function  $P_\alpha$  yields

$$P_\alpha(T_1/e^n \cdot a^n \cdot g) = \frac{P(e^n/a^n \cdot g \cdot T_1) P_\alpha(T_1/a^n \cdot g)}{P(e^n/a^n \cdot g \cdot T_1) P_\alpha(T_1/a^n \cdot g) + P_\alpha(e^n/a^n \cdot g \cdot \neg T_1) P_\alpha(\neg T_1/a^n \cdot g)}$$

Here “ $e^n$ ” represents the conjunction “ $(e_1 \cdot e_2 \dots e_n)$ ,” and the same convention applies to “ $a^n$ .”

Taking the initial condition statements as irrelevant to  $T_1$  on  $g$  alone (i.e.,  $P_\alpha(T_1/a^n \cdot g) = P_\alpha(T_1/g)$ , since the initial conditions should only be relevant to  $T_1$  in light of the outcome of the test), noting that the value of the direct inferences

$P(e^n/a^n \cdot g \cdot T_1) = 1$  (since  $T_1 \cdot g \cdot a^n$  entails  $e^n$ ), and dividing numerator and denominator by the prior probability of  $T_1$  (i.e.,  $P(T_1/g)$ ) we have

$$P_\alpha(T_1/e^n \cdot a^n \cdot g) = \frac{1}{1 + \frac{P_\alpha(e^n/a^n \cdot g \cdot -T_1)}{P_\alpha(T_1/g)} \times (1 - P_\alpha(T_1/g))}$$

Note that the term  $(1 - P_\alpha(T_1/g))$  is equal to  $P_\alpha(-T_1/g)$ . So, as evidence increases (i.e., as  $n$  increases),  $T_1$  becomes highly confirmed (i.e.,  $P_\alpha(T_1/e^n \cdot a^n \cdot g)$  goes to 1) if and only if two conditions are met:  $P_\alpha(T_1/g) > 0$ , and  $P_\alpha(e^n/a^n \cdot g \cdot -T_1)$  goes to 0 as evidence (i.e.,  $n$ ) increases. If these conditions can be satisfied, then a whole class of inductive inferences will be justified.

Russell finds the second condition unproblematic in many cases dealing with empirical material. He notes (following Keynes) that

$$\begin{aligned} P_\alpha(e^n/a^n \cdot g \cdot -T_1) &= P_\alpha(e_n/a_n \cdot e^{n-1} \cdot a^{n-1} \cdot g \cdot -T_1) \times \\ P_\alpha(e_{n-1}/a_{n-1} \cdot e^{n-2} \cdot a^{n-2} \cdot g \cdot -T_1) \times \dots \times \\ P_\alpha(e_1/a_1 \cdot g \cdot -T_1). \end{aligned}$$

If each of the terms on the right is less than some fixed value  $q < 1$  for arbitrarily large  $n$ , then  $P_\alpha(e^n/a^n \cdot g \cdot -T_1) < q^n$ , which goes to 0 as  $n$  increases. This satisfies Russell that the second condition can often be met.

Russell finds the first condition, the question of which theories should get non-zero prior probabilities, to be the primary problem. He finds it difficult to see how any empirical hypothesis can be found initially credible (or incredible) independent of experience. And there is an additional difficulty. If the prior probability for a true theory, say  $T_1$ , is too low (but nonzero) on credibility function  $P_\alpha$ , then the whole lifespan of the human race may not be of sufficient length to gather the amount of evidence needed to make  $P_\alpha(T_1/e^n \cdot a^n \cdot g)$  appreciably large (i.e., by making  $n$  large enough for

$$\frac{P_\alpha(e^n/a^n \cdot g \cdot -T_1)}{P_\alpha(T_1/g)}$$

to become appreciably small).

We will return to difficulties with prior probabilities. But first I want to consider another, more general Bayesian analysis that arises in *Human Knowledge* (see pp. 406–7) in a section on induction by simple enumeration. I am going to generalize the analysis substantially, but the seeds of it are clearly there, in the text (see also pp. 410–12). This analysis leads back to the same kind of problem with prior probabilities. Let  $T_1, T_2, \dots$ , be a (possibly infinite) list of competing theories such that  $P_\alpha(T_i \cdot T_j/g) = 0$  whenever  $i \neq j$  and  $\sum_i P_\alpha(T_i/g) = 1$ . Let  $e$  be a possible outcome of initial condition  $a$  (taken together with background  $g$ ). Induction by simple enumeration is the process of determining the probability of

$e$  on an accumulating body of relevant evidence  $e^n \cdot a^n$  (where superscripts are understood as before). The following is a theorem of probability theory:

$$\begin{aligned} P_\alpha(e/a \cdot e^n \cdot a^n \cdot g) &= \sum_i P_\alpha(e/a \cdot e^n \cdot a^n \cdot g \cdot T_i) P_\alpha(T_i/a \cdot e^n \cdot a^n \cdot g) \\ &= \sum_i P_\alpha(e/a \cdot g \cdot T_i) P_\alpha(T_i/e^n \cdot a^n \cdot g). \end{aligned}$$

(Note that  $\sum_i P_\alpha(T_i/e^n \cdot a^n \cdot g) = 1$ .) The second line of the equality depends only on two additional assumptions: each  $T_i$ , together with background knowledge  $g$  and initial condition  $a$ , bears a direct inference relation to  $e$  independent of the other data ( $e^n \cdot a^n$ ); initial conditions are relevant to the  $T_i$ , given  $g$ , only in light of their associated outcomes. This formula expresses a kind of inductive systematization that theories impose on simple inductions. Notice that if increasing evidence (for large  $n$ ) drives the posterior probability of one of the theories toward 1, then the simple induction takes on the logical value imposed by that theory in direct inference (i.e., if (as  $n$  increases)  $P_\alpha(T_i/e^n \cdot a^n \cdot g) \rightarrow 1$ , then (as  $n$  increases)  $P_\alpha(e/a \cdot e^n \cdot a^n \cdot g) \rightarrow P(e/a \cdot g \cdot T_i)$ ). Thus, the probability of the next event given by simple induction depends entirely on the probability various theories assign it and the credibility of each theory on past evidence.

The degree of credibility of each theory is given by Bayes's theorem. For example, the credibility of  $T_1$  is given by

$$P_\alpha(T_1/e^n \cdot a^n \cdot g) = \frac{P(e^n/a^n \cdot g \cdot T_1) P_\alpha(T_1/g)}{\sum_i P(e^n/a^n \cdot g \cdot T_i) P_\alpha(T_i/g)}$$

where  $\sum_i P_\alpha(T_i/g) = 1$ . Notice that the posterior probability of  $T_1$  on  $e^n \cdot a^n \cdot g$  depends only on the logically determined values  $P(e^n/a^n \cdot g \cdot T_i)$  and on the values of the prior probabilities of the various possible theories. In general the  $T_i$  may be statistical theories, yielding values for  $P(e^n/a^n \cdot g \cdot T_i)$  other than 1 or 0. But they are direct logical inferences on which all rational credibility functions agree.

To get a clearer picture of what Bayes's formula does, consider the case where the  $T_i$  are all deterministic theories. That is, suppose for each  $T_i$ ,  $T_i \cdot g \cdot a^n$  either entails  $e^n$  or it entails  $\neg(e^n)$ . Then for each  $T_i$  and each outcome  $e^n$ , either  $P(e^n/a^n \cdot g \cdot T_i) = 1$  or 0. If every theory entails the evidence, then it follows from Bayes's theorem that  $P_\alpha(T_i/e^n \cdot a^n \cdot g) = P_\alpha(T_i/g)$ . The only way evidence can change the credibility of a particular theory, say  $T_1$ , is either by making  $P(e^n/a^n \cdot g \cdot T_1) = 0$  so that  $P_\alpha(T_1/e^n \cdot a^n \cdot g) = 0$  and  $T_1$  is falsified, or by making  $P(e^n/a^n \cdot g \cdot T_i) = 0$  for some  $T_i$  other than  $T_1$ . In the latter case

$$P(T_1/e^n \cdot a^n \cdot g) = \frac{P_\alpha(T_1/g)}{\sum_j P_\alpha(T_j/g)}$$

where  $j$  ranges only over those theories not falsified by  $e^n \cdot a^n \cdot g$ . This is just hypothetico-deductive theory confirmation with weights assigned to the various alternative theories. The prior probabilities assign each theory a weight, and the



weights all add to 1. Imagine these theories lined up in the order of magnitude of their prior probabilities. When some theories are falsified by the evidence, eliminate them from the line and renormalize the weights of the remaining theories to 1. The true theory, represented here by  $T_1$ , can become highly confirmed only if all theories with greater weight and the most substantial of the less weighty theories are falsified. Bayes's formula measures the relative size of a theory's prior probability to the collective size of the priors of all as yet unrefuted theories.

Cases in which the  $T_i$ 's include statistical theories are only a little more complicated. Dividing out the numerator from the earlier Bayes equation we have

$$P(T_1/e^n \cdot a^n \cdot g) = \frac{1}{1 + \sum_{i \neq 1} \frac{P(e^n/a^n \cdot g \cdot T_i)}{P(e^n/a^n \cdot g \cdot T_1)} \frac{P_\alpha(T_i/g)}{P_\alpha(T_1/g)}}$$

The ratios

$$\frac{P(e^n/a^n \cdot g \cdot T_i)}{P(e^n/a^n \cdot g \cdot T_1)}$$

are called the likelihood ratios.  $P_\alpha(T_1/e^n \cdot a^n \cdot g)$  goes to 1 if and only if the likelihood ratios all go to 0 (with increasing evidence). And  $P_\alpha(T_1/e^n \cdot a^n \cdot g)$  goes to 0 just in case at least one of the likelihood ratios blows up to infinity (with increasing evidence).

In both the case of purely deterministic alternative theories and in the more general case involving statistical theories, the logic of probability leads to the same conclusion: probabilistic inference will assign a high degree of credibility to a true theory if and only if that theory is assigned a nonzero prior degree of credibility and, in addition, all of its competitors that have prior credibility of any appreciable size in comparison can be refuted (relative to it) by the evidence. If two theories agree on the evidence, then the relative size of their posterior probabilities remains proportional to the relative size of their priors. This establishes the second step of the Russell-Maxwell view, and we've specified what formal condition prior probabilities must satisfy if inductive logic is to make true theories highly confirmed.

The Bayesian analysis establishes this claim: inductive logic succeeds in making a true theory highly probable on the evidence within a reasonably short segment of human history *just in case* the prior probability assigned to it is not too low and the prior probabilities assigned to evidentially indistinguishable false competitors are significantly lower than the prior of the true theory.

I don't pretend that Russell had in mind the precise details of the second Bayesian analysis when he wrote *Human Knowledge*, but closely related arguments drawn from Keynes are sketched there. The preceding version above has evolved

into its present form through the work of more recent Bayesians. I became aware of much of it through working with Grover Maxwell.

We now approach step 3 of the Russell-Maxwell view. If judgment empiricism is to be saved, then logical considerations alone must provide that prior credibilities are assigned to hypotheses in the right way. Can any noncontingent, purely logical principle guide the assignments of prior credibilities so as to determine that true hypotheses are assigned large enough values and their evidentially indistinguishable competitors are assigned small enough values?

Both Russell and Maxwell come to this issue already fairly convinced by more general Humean considerations that logic alone cannot justify any inductive method. The Bayesian analysis promises to surmount the Humean obstacles if purely logical considerations can furnish prior probabilities of the required sort. So both philosophers expect to find deficiencies in purely logical accounts. Russell investigates attempts related to Keynes's analysis.<sup>9</sup> Maxwell considers Carnap's attempt.<sup>10</sup> Rather than reconstructing the details of particular logicist attempts and their failures, I will outline a more general critique. It is closely related to issues raised by Russell and Maxwell.

First, then, consider that no purely logical account should assign a prior probability of zero to any contingent hypothesis. For every contingent hypothesis is true in some possible world, and prior to the evidence our world could be any one of those worlds. The logic of probability theory (and the Bayesian analyses in particular) shows that if a contingent hypothesis has a zero prior probability, then no evidence can give it a nonzero posterior probability. If logic alone is to provide priors that give a true hypothesis any chance at confirmation it must assign to it and all its contingent competitors nonzero priors.

Any reasonably sophisticated hypothesis has an infinite number of contingent competitors. They cannot all have the same nonzero prior probability, for then they would sum to infinity rather than 1. Indeed, since the sum of these priors must be 1, for any positive real number  $\epsilon$  (as close to zero as you wish) there must be an infinite number of these hypotheses with prior probabilities less than  $\epsilon$ . If inductive logic is to highly confirm the true hypothesis, then purely logical considerations must assign all but a relative few of its competitors arbitrarily small prior credibilities, priors so small that all the evidence obtainable in human history would not suffice to confirm the true hypothesis were it among them. How could purely logical considerations determine which subset of contingent hypotheses is so sure to contain the true one that only its members receive high enough priors to have a chance at confirmation? And how could these considerations eliminate those competitors to the true hypothesis that agree on all possible evidence? After all, they are contingent too.

Logical considerations cannot assign priors to hypotheses on the basis of their empirical content, for logical consideration alone cannot play favorites among competing contingent claims. So logicist accounts usually make their assignments

of priors on the basis of the syntactic structure of the hypotheses. The idea seems to be that ad hoc, implausible, and silly competitors will have syntactic logical structures that will expose them (in some precise way) as ad hoc, implausible, and silly. But, isn't it logically possible that some such hypotheses are true? That nature does not in fact instantiate such hypothesis would be a contingent (not a purely logical) assumption. Besides, examples a la Nelson Goodman's<sup>11</sup> grue-predicates show how the logician's bag of tricks can always furnish an infinite number of seemingly silly competitors with the same logical structure as the true hypothesis. This move, calling on Goodmanesque predicates, can be countered only by some sort of restriction to hypotheses formulated in a preferred vocabulary, a vocabulary in which predicate terms pick out the real properties and relations (natural kinds?) rather than hokey Goodmanesque composites. Maxwell, commenting on Carnap's system, points out the trouble with such a restriction<sup>12</sup>:

In order to apply Carnap's system, given any instance of evidence to be used, we must already know what are the relevant individuals and properties in the universe (as well as the cardinality of the relevant classes of individuals). In other words, we must already have the universe "carved at the joints" and spread out before us (otherwise we could not perform the necessary counting of stated descriptions, etc.). But just how to "carve at the joints," just what are the important relevant properties, and just which segments of the world are to be designated as individuals are all contingent questions—or, at least, they have crucial contingent components—and, moreover, are often the most important questions that arise in scientific inquiry. Before any application of Carnap's system can begin, we must assume that a great deal—perhaps most—of that for which a confirmation theory is needed has already been accomplished. We may, as might have been expected, put the objection in a form that, by now, is familiar ad nauseum: Given any body of evidence, there is an infinite number of mutually incompatible ways of "carving the world" (setting up state-descriptions, structure descriptions, etc.) each of which will give different results for predictions, "instance confirmations," etc.

Such considerations lead Russell and Maxwell to the conclusion of their argument. Logic and evidence alone cannot eliminate the myriad competing hypotheses (some of them evidentially indistinguishable) which vie with the true one for inductive confirmation. Science avoids them only because scientists find them silly and implausible (if they find them at all). If these hypotheses are false, they are only contingently false. Only contingent principles can rule them out. Inductive logic is sound only if some contingent principles guide and justify the assignment of prior degrees of credibility (assigning sufficiently large priors to true theories and sufficiently small priors to evidentially indistinguishable competitors) so that the evidence available to us in a reasonable stretch of human history can eliminate initially plausible false competitors and confirm true ones.

In an earlier chapter of *Human Knowledge* on causal laws Russell summarizes the conclusions he will draw from his later Keynesian analysis. Here he makes it abundantly clear that the role of his Postulates of Scientific Inference is to guide the assignment of prior probabilities to theories. I will quote at some length:

In the establishment of scientific laws experience plays a twofold part. There is the obvious confirming or confuting of a hypothesis by observing whether its calculated consequences take place, and there is the previous experience which determines what hypotheses we shall think antecedently probable. But behind these influences of experience there are certain vague general expectations, and unless these confer a finite a priori probability on certain kinds of hypotheses, scientific inferences are not valid. In clarifying scientific method it is essential to give as much precision as possible to these expectations, and to examine whether the success of science in any degree confirms their validity. After being made precise the expectations are, of course, no longer quite what they were while they remained vague, but so long as they remain vague the question whether they are true or false is also vague.

It seems to me that what may be called the "faith" of science is more or less of the following sort: There are formulas (causal laws) connecting events, both perceived and unperceived; these formulas exhibit spatio-temporal continuity; i.e., involve no direct unmediated relation between events at a finite distance from each other. A suggested formula having the above characteristics becomes highly probable if, in addition to fitting in with all past observations, it enables us to predict others which are subsequently confirmed and which would be very improbable if the formula were false. (p. 314)

I won't take time for a careful treatment of Russell's postulates here. I only want to suggest that the five postulates he sets down seem to be an attempt to state the basic assumptions underlying a version of a scientific realist view of the world. They state a metaphysical view that characterizes the world as composed of mind-independent events and spatiotemporally continuous causal sequences. After listing the five postulates at the beginning of a chapter that summarizes them, Russell clearly states their purpose:

The postulates collectively are intended to provide the antecedent probabilities required to justify inductions. (HK, p. 487)

The postulates seem to provide Russell with the required antecedent probabilities in two senses. They explain how it is that we can have relatively accurate knowledge of the (mostly distant) events that make up the world, i.e., through the mediation of spatiotemporally continuous causal processes. The relative constancy, stability, and similarity inherent in many of these processes together with our vast experience at the perceiving end of them account for our exceptionally good luck at discovering credible hypotheses that are true or nearly true. Second,

the postulates function as a guide in assessing the rational credibility of (other) contingent hypotheses. A hypothesis or theory is totally incredible (has probability 0) if it is inconsistent with the causal makeup of the world as described by the postulates. If a hypothesis is at all credible, its prior degree of credibility should be assessed, it seems, on the basis of how coherently (as compared with credible competitors) it fits together the postulated causal structure of the world.

The postulates are metaphysical in that they express basic contingent facts about the makeup of the world, they must be known if any empirical knowledge beyond immediate experience is possible, and they cannot themselves be known empirically (i.e., on the basis of experience and logic alone).

But although experience of barking dogs suffices to *cause* belief in the generalization "Dogs bark," it does not, by itself, give any grounds for believing this is true in untested cases. If experience is to give such a ground, it must be supplemented by causal principles such as will make certain kinds of generalization antecedently plausible. These principles, if assumed, lead to results which are in conformity with experience, but this fact does not logically suffice to make the principles even probable. (HK, p. 507)

Maxwell was unaware of Russell's Bayesian leanings, but he acknowledges his agreement with Russell that inductive inference needs a boost from some preinductive, contingently true postulate. He comments that Russell's postulates may reflect an important part of our commonsense knowledge, but he finds them neither necessary nor sufficient for a viable theory of induction. Maxwell (like most other philosophers) seems to have missed the Bayesian component of Russell's inductive logic. He (and others) finds the postulates perplexing because it is difficult to see how the postulates, standing alone, can furnish anything like inductive rules of inference. The postulates seem to suggest what kinds of theories to search for, but they furnish no machinery for testing or confirming theories with evidence. I think this is the main thrust of Maxwell's criticism of Russell's approach. But Maxwell found fault with the postulates on other grounds, too. He found them inadequate as a guide to the assignment of prior probabilities.

Maxwell's arguments suggest that even if the version of scientific realism expressed by Russell's postulates is true (as it may well be) there are a multitude of incompatible, but evidentially indistinguishable, theories that satisfy them. Given one such theory, the logician can easily concoct the rest. Appeals to *simplicity* will only help narrow the field if it is contingently true that nature favors simpler theories. This would be an additional contingent postulate. And, again, Goodman's grue-predicates illustrate that the logician can often get the desired syntactic simplicity by drawing on his bag of tricks. ("Theories gotten without tricks are more likely a priori" would be another plausible, but contingent, *assumption*.)

Maxwell's method for assigning prior probabilities to theories is simply this:

a person should assign prior probabilities by ordering and weighting the alternatives in accordance with his best carefully considered intuitive judgment of their respective credibilities (actually Maxwell used a frequency theory of probability instead of degree of credibility, so his version explicates “credibilities” in terms of “relative frequencies of truth among similar hypotheses”). Maxwell backs this recommendation with a contingent but (on pain of circularity) unconfirmable postulate. It says, roughly, that we have the innate (perhaps, naturally selected) capacities to develop the following abilities: first, to propose hypotheses such that a nonvanishing proportion of them are true; and second, to rank these hypotheses, by means of subjective estimates, in such a way that the evidence will pare away the initially most credible competitors of a true hypothesis. This is a rational reconstruction of how we actually proceed both in science and in daily life, and, Maxwell argued, no inductive logic can improve on it.<sup>13</sup>

I will conclude with a long quotation from the last two paragraphs of *Human Knowledge*. Maxwell believed that the view it expresses was one of Russell’s most significant contributions.

As mankind have advanced in intelligence, their inferential habits have come gradually nearer to agreement with the laws of nature which have made these habits, throughout, more often a source of true expectations than of false ones. The forming of inferential habits which lead to true expectations is part of the adaptation to the environment upon which biological survival depends.

But although our postulates can, in this way, be fitted into a framework which has what we may call an empiricist “flavor,” it remains undeniable that our knowledge of them, in so far as we do know them, cannot be based upon experience, though all their verifiable consequences are such as experience will confirm. In this sense, it must be admitted, empiricism as a theory of knowledge has proved inadequate, though less so than any other previous theory of knowledge. Indeed, such inadequacies as we have seemed to find in empiricism have been discovered by strict adherence to a doctrine by which empiricist philosophy has been inspired: that all human knowledge is uncertain, inexact, and partial. To this doctrine we have not found any limitation whatever. (p. 507)

#### Notes

1. Bertrand Russell, *Human Knowledge: Its Scope and Limits* (New York: Simon & Schuster, 1948).

2. Grover Maxwell, “The Later Russell: Philosophical Revolutionary,” in *Russell’s Philosophy*, ed. George Nakhnikian (London: Duckworth, 1974), pp. 169–82.

3. Kenneth Blackwell, Archivist of the Bertrand Russell Archives at McMaster University, reports that the Archives has a two foot stack of manuscripts related to *Human Knowledge*, and that much of this material is devoted to various probabilistic investigations and derivations. Perhaps more detail from Russell’s Bayesian investigations can be found there.

4. Frank P. Ramsey, “Truth and Probability” (1926), in Henry E. Kyburg and Howard E. Smokler

(eds.), *Studies in Subjective Probability* (New York: Wiley, 1964), pp. 61–92; Bruno De Finetti, “La Préviation: ses lois logiques, ses sources subjectives” (1937), translated in Kyburg and Smokler, *Studies in Subjective Probability*, pp. 93–158; L. J. Savage, *The Foundations of Statistics* (New York: Wiley, 1954).

5. John Maynard Keynes, *A Treatise on Probability* (London and New York: Macmillan, 1921; 2nd ed., 1929); Harold Jeffreys, *Theory of Probability*, 2nd ed. (Oxford: Clarendon Press, 1948); Rudolf Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950).

6. Hans Reichenbach, *The Theory of Probability* (Berkeley: University of California Press, 1949); Wesley Salmon, *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1966 & 1967); Grover Maxwell, “Induction and Empiricism: A Bayesian Frequentist Alternative,” in Grover Maxwell & Robert Anderson, Jr. (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 3 (Minneapolis: University of Minnesota Press, 1975), pp. 106–65.

7. R. A. Fisher, *Statistical Methods for Research Workers* (Edinburgh and London: Oliver & Boyd, 1925); J. Neyman and E. S. Pearson, *Joint Statistical Papers* (Berkeley: University of California Press, 1967); Henry E. Kyburg, Jr., *The Logical Foundations of Statistical Inference* (Boston: Reidel, 1974).

8. John Maynard Keynes, *Treatise on Probability*.

9. See HK, pp. 408–12 and 433–44.

10. See Maxwell, “Induction and Empiricism,” pp. 161–62.

11. Nelson Goodman, *Fact, Fiction and Forecast*, 2nd ed. (New York: Bobbs-Merrill, 1965).

12. Maxwell, “Induction and Empiricism.”

13. Grover Maxwell, “Induction and Empiricism,” and “Corroboration without Demarcation,” in P. A. Schlipp (ed.), *The Philosophy of Karl Popper: The Library of Living Philosophers* (LaSalle, IL: Open Court, 1974), pp. 292–321.