6 The Rationality and Objectivity of Scientific Inference

1 Introduction

The rationality and objectivity of scientific inference is a topic that has been much on the minds of philosophers of science over the past few decades. But all too often these minds have been abuzz with incommensurability, relativism, duck-rabbit gestalt switches, the theory-ladenness of observation, and the like. I will have more to say about these matters in chapter 8, but for present purposes I will assume that we have all taken a magic pill that has cured this particular buzz. While this strategy will be unappealing to some readers, I think it serves to make the problem of rationality and objectivity more interesting, since, if I am right, the problem arises even if the "new fuzziness" (as Clark Glymour has dubbed it) is set aside.

For the purest of the Bayesian personalists, the constraints of rationality begin and end with the axioms of probability. Less extreme personalists want to impose the conditionalization model for learning from experience and perhaps some form of David Lewis's principal principle (chapter 2). If these procedural constraints are satisfied, will the resulting degrees of belief count as rational? And if not—if there is a mismatch between the rationality of the result and the alleged rationality of the procedure—can we properly say that the procedure is fully rational? To put the matter in concrete terms, if in the face of the currently available evidence you assign a high degree of belief to the propositions that Velikovsky's *Worlds in Collision* scenario is basically correct, that there are canals on Mars, that the earth is flat, etc., you will rightly be labeled as having an irrational belief system. And if you arrived at your present beliefs within the framework of Bayesian personalism, then the temptation is to say that at worst there is something rotten at the core of Bayesian personalism and at best there is an essential incompleteness in its account of procedural rationality.

Here the issue of rationality merges with that of objectivity. We use epithets like 'irrational' and 'crackpot' in the case I just described and in other cases where there is objectivity of opinion in the sense of a tight intersubjective agreement in the relevant scientific community and where the stigmatized person has opinions that differ radically from the consensus. Such objectivity exists in science not only concerning the roundness of the earth but for more theoretically interesting propositions as well, e.g., that matter has an atomic structure, that space and time are relativistic rather than absolute, etc. One's expectation, or hope if you will, is that the explanation of the intersubjective agreement on such matters is not merely historical or sociological but has a justificatory character—otherwise labeling as 'irrational' or a 'crackpot' those whose opinions depart radically from the consensus would be unfounded.

This sort of talk has traditionally been thought to presuppose a metaphysical thesis of objectivity to the effect that the object of scientific inquiry, "the world," exists independently of knowing subjects and that it makes sense to say of our opinions that either they match or fail to match this objective reality. In recent years the social constructivists have ridiculed this thesis as an outdated piece of ideology. I will not dignify constructivism with a response but will instead concentrate on what I take to be a more interesting but unmet challenge to the objectivity of scientific inference.

The popular image of science would have it that science provides us with a methodology for generating objective opinion: apply the scientific method faithfully and long enough and eventually it will produce certitudes that match reality, and before certainty is reached, the faithful use of the method by all the members of the scientific community guarantees objectivity qua intersubjective agreement on degrees of belief. These degrees of belief are rational because they are produced by an objective method of inquiry: it is value-free and presupposition-free, it is evidence-driven, and it sanctions no inference not strictly warranted by the evidence.[1] This straw man, or as I would prefer to say, this wish list, has been criticized on the grounds that nothing remotely like it can be instantiated because of incommensurability, the theory-ladenness of observation, and the like. Again I am postponing this challenge until chapter 8 to take up a more fundamental challenge: even if we leave aside incommensurability and its fellow travelers, there is still no obvious candidate for this objective method of inquiry. Certainly Bayesianism doesn't qualify, since it not only allows but requires presuppositions in the form of prior probabilities.

Failure to fulfill the pop-science specifications for an objective method of inquiry does not mean that Bayesianism cannot deliver on objectivity. We will see that the long-run use of the Bayesian method does produce certainties that "almost surely" match reality. At least this is so for observational hypotheses; theoretical hypotheses are another matter, one that requires careful discussion. Less satisfactory are attempts to ground objectivity as intersubjective agreement for opinions that fall short of certainty. Two of the most widely used attempts to provide the grounding within the

Bayesian framework are (1) constraining priors and (2) the washing out of priors. Both attempts, I will argue, have only limited success. Several other alternatives present themselves: (3) definitional solution, (4) socialism, (5) evolutionary solution, (6) modest but realistic solutions, (7) non-Bayesian solutions, and (8) retrenchment. Each of these alternatives will also be found wanting.

2 Constraining Priors

The first strategy for grounding objectivity as intersubjective agreement proceeds by supplementing the personalist form of Bayesianism by adding constraints on prior probabilities. The literature contains numerous and telling objections to various attempts to implement this strategy. I will not review this literature here but will simply indicate briefly why such attempts are unworkable and why they would not solve the problem of rationality and objectivity even if they were workable.

There are two reasons why such principles are unworkable. The first is that different applications of these principles are possible, and the different applications can yield conflicting results. This phenomenon was illustrated in chapter 1 for Bayes's particular application of the principle of insufficient reason. Nor is the problem of choosing the "correct" application much easier than the general problem of deciding what to believe. This is again illustrated by Bayes's case, since the application favored by Bayes leads to an inductivist probability with a high rate of learning from experience, whereas another application led to a noninductivist probability with no learning from experience. What holds for Bayes's rule for assigning priors holds quite generally: there are different ways of conceptualizing an inference problem, and the application of the rule to the different conceptualizations leads to different results. The problem of choosing among the different results is no less difficult than the original problem of assigning priors.

Even if there were no ambiguity in the conditions of application of these principles, there would remain the problem that the conditions are rarely satisfied in real-life cases. Recall that Bayes assumed a condition of complete ignorance regarding the unknown event. If we ignore the potential ambiguities in this notion, the point is that such a condition will be realized only in never-never-land Bayesianism, where an agent begins as a *tabula rasa*, chooses her priors, and forever after changes her probabilities only by conditionalization. A more realistic Bayesianism would recognize the local and episodic character of problem solving. In Bayesian terms, we use different probability functions for different problem-solving contexts, and within a context we may change probabilities not by conditionalization but by some more radical means.[2] Thus, far from being a *tabula rasa*, the typical scientist comes burdened with a wealth of information in trying to make what the Bayesian would describe as decisions about prior probabilities. E. T. Jaynes's modern version of the principle of indifference tries to take into account some of this information, since it enjoins us to maximize a quantity he calls "entropy" subject to known constraints that can be expressed in terms of moments of the probability distribution.[3] But only a small part of prior information can be expressed in these terms.

How, then, is prior information taken into account? What one finds running through the scientific literature are plausibility arguments. In Bayesian terms, these arguments are designed to persuade us to assign high priors to some alternatives and low priors to others.[4] Part of what it means to be an "expert" in a field is to possess the ability to recognize when such persuasions are good and when they are not. But it is highly doubtful that this ability can be codified in simple formal rules. And even if it could, why is or should only expert opinion be tolerated?

Suppose now for the sake of argument that there are workable rules for assigning priors. There are still two reasons why these rules will not suffice to explain objectivity. First, the explanation would have the sought-after justificatory character only if the rules in question were accepted as norms of rational behavior. But their normative status is highly controversial; indeed, these rules are either explicitly rejected or else ignored by a large segment of the Bayesian camp. Second, even if these rules were uniformly accepted, they would not be sufficient to explain objectivity unless they sufficed to fix the likelihoods $\Pr(E/H_i)$ needed to implement the right-hand side of Bayes's theorem in the form (2.2). But these rules are typically intended to fix the prior probabilities on a partition $\{H_i\}$ of hypotheses and are not intended to apply to partitions such as $\{E \,\&\, H_i\}$, where E is a possible data report.[5] There are, of course, cases where the likelihoods do have an objective status. The HD case, where for each i either $H_i \,\&\, K \models E$ or $H_i \,\&\, K \models \neg E$, is one such. Another obtains when all the Bayesian agents agree on a statistical model for a chance experiment, E reports outcomes of the experiment, the H_i are alternative hypotheses about the

objective-probability parameters of the chance setup, and Lewis's principal principle applies. But these cases hardly exhaust the domain that would have to be covered by an adequate theory of scientific inference. Consider, for example, as astronomers of the seventeenth century were forced to, what probability should be assigned to stellar parallax of various magnitudes on the assumptions of a Copernican cosmology and the then accepted background knowledge, which contained scanty and uncertain information about the distance between the earth and the fixed stars.[6]

3 The Washing Out of Priors: Some Bayesian Folklore

Many Bayesians analyze objectivity in terms of the washing out of priors. Thus Adrian F. M. Smith writes,

> I personally am only able to make sense of the concept [scientific objectivity] in the context of a Bayesian philosophy that predisposes one to seek to report, openly and accessibly, a rich range of possible belief mappings induced by a given data set, the range being chosen to reflect and potentially to challenge the initial perceptions of a broad class of interested parties. If a fairly sharp consensus of views emerges from a rather wide spread of initial opinions, then, and only then, might it be meaningful to refer to "objectivity." (1986, p. 10)

The implication of Smith's suggestion is that even if there were workable principles for constraining priors, it would be a mistake to impose them. It is a fact of life that scientists start with different opinions. To try to quash this fact is to miss the essence of scientific objectivity: the emergence of an evidence-driven consensus from widely differing initial opinions.

It is part of the Bayesian folklore that the emergence of such a consensus is routine. Differences in prior probabilities do not matter much, at least not in the long run; for (the story goes) as more and more evidence accumulates, these differences wash out in the sense that the posterior probabilities merge, typically because they all converge to 1 on the true hypothesis. Here are two passages that have given currency to this folklore. The first comes from the now classic review article "Bayesian Statistical Inference for Psychological Research" by Edwards, Lindman, and Savage:

> Although your initial opinion about future behavior of a coin may differ radically from your neighbor's, your opinion and his will ordinarily be transformed by application of Bayes' theorem to the results of a long sequence of experimental flips as to become nearly indistinguishable. (1963, p. 197)

A similar sentiment has been expressed by Suppes:

> It is of fundamental importance to any deep appreciation of the Bayesian viewpoint to realize the particular form of the prior distribution expressing beliefs held before the experiment is conducted is not a crucial matter.... For the Bayesian, concerned as he is to deal with the real world of ordinary and scientific experience, the existence of a systematic method for reaching agreement is important.... The well-designed experiment is one that will swamp divergent prior distributions with the clarity and sharpness of its results, and thereby render insignificant the diversity of prior opinion. (1966, p. 204)

I take it that if this folklore were correct, the explanation of objectivity would have a justificatory resonance. The consensual degrees of belief are justified because they are the inevitable results of a rational process: let the Bayesian agents start off with whatever initial degrees of belief they like, as long as they conform to the probability calculus and as long as they don't differ too radically (as explained below), and let them update their opinions via the rule of conditionalization; as a result, they will all be driven by the accumulating evidence to the same final degrees of belief.

The folklore is based on more than pious hope and promissory notes. There are in fact hard mathematical results on merger of opinion that can be proved within the framework of a moderately tempered Bayesian personalism, characterized by the following principles:

P1 Degrees of belief satisfy the axioms of probability.

P2 Learning from experience is modeled as change of probability via strict conditionalization.

P3 All the agents of concern begin as *equally dogmatic* in that they initially assign 0's to the same elements of the probability space.

Principle (P3) can be motivated by a rule of mutual respect that enjoins members of a scientific community to accord a nonzero prior to any hypothesis seriously proposed by a member of the community.[7] Alternatively, it could be held that decisions on zero priors help to define scientific communities and that an account of scientific inference must be relativized to a community.[8]

The sort of result that Edwards, Lindman, Savage, and Suppes had in mind can be illustrated by an example adapted from Savage's *Foundations of Statistics* (1954). In this example it suffices to explicate (P1) in terms of (A1) to (A3) from chapter 1; countable additivity for Pr plays no role here,

although it does figure essentially in the more sophisticated results discussed below in sections 4 and 5. Consider a coin-flipping experiment, and suppose that all of the Bayesian agents of concern accept the posit K of independently and identically distributed (IID) trials. Suppose further, in concert with (P3), that they all assign nonzero priors to the hypotheses $\{H_i\}$, where H_i states that the objective probability of heads is p_i ($p_i \neq p_j$ if $i \neq j$). And finally, assume that in conformity with Lewis's principal principle they all evaluate the likelihoods as

$$\Pr\left(\underset{j \leqslant n}{\&} E_j/H_i \,\&\, K\right) = p_i^m(1-p_i)^{n-m},$$

where E_j reports the result of the jth flip and m is the total number of heads in the first n flips. Choose any one of the agents in the Bayesian community, and apply Bayes's theorem to conclude that for her the ratio of the posterior probabilities of two of the competing hypotheses is

$$\frac{\Pr(H_i/\&_{j\leqslant n} E_j \,\&\, K)}{\Pr(H_k/\&_{j\leqslant n} E_j \,\&\, K)} = \frac{p_i^m(1-p_i)^{n-m}\Pr(H_i/K)}{p_k^m(1-p_k)^{n-m}\Pr(H_k/K)}.$$

The strong form of the law of large numbers assures us that in almost every endless repetition of the experiment, the relative frequency of heads approaches the true value of the chance, say p_3, in the limit as $n \to \infty$.[9] As a consequence, the likelihood ratio for $i = 3$ and $k \neq 3$ almost surely blows up, which implies that $\Pr(H_3/\&_{j\leqslant n} E_j \,\&\, K) \to 1$ as $n \to \infty$. By (P2), the probability function $\Pr_n(\cdot)$ at stage n for an agent with starting probability $\Pr_0(\cdot) = \Pr(\cdot/K)$ is $\Pr(\cdot/\&_{j\leqslant n} E_j \,\&\, K)$. Thus as $n \to \infty$, the opinions of all of the agents regarding the H_i will almost surely merge, since each agent almost surely converges to certainty on the true hypothesis H_3.

Mary Hesse has objected that "the conditions of [Savage's type of convergence] theorem...are not valid for typical examples of scientific inference" (1975, p. 78). In particular, the crucial IID assumption certainly does not apply to the results of experiments addressed to nonstatistical hypotheses. Nor, as already noted above, is the assumption of objective likelihoods justified in these nonstatistical cases, save when HD testing is applicable.

While these objections are well taken, it is nevertheless true that more powerful convergence-to-certainty and merger-of-opinion results, none of which uses the questionable assumptions tagged by Hesse, are available in the statistics literature. Since the most elegant of these results use Doob's theory of martingales, I will briefly outline some of the leading ideas of this theory in the following section.

4 Convergence to Certainty and Merger of Opinion as a Consequence of Martingale Convergence

Consider a probability space in the mathematician's sense, that is, a triple $(\Omega, \mathscr{F}, \mathscr{P}\!r)$ consisting of a sample space Ω, a collection $\mathscr{F}$ of measurable subsets of Ω, and a countably additive function $\mathscr{P}\!r\colon \mathscr{F} \to [0,1]$ such that $\mathscr{P}\!r(\Omega) = 1$. Let $\{X_n\}$ be a sequence of random variables (rv's) and let $\{\mathscr{F}_n\}$ be a sequence of σ fields such that $\mathscr{F}_n \subseteq \mathscr{F}_{n+1} \subseteq \mathscr{F}$.[10] The set $\{X_n\}$ is said to be a *martingale* with respect to $\{\mathscr{F}_n\}$ just in case for every n, $E(|X_n|) < \infty$, X_n is measurable with respect to $\mathscr{F}_n$, and $E(X_{n+1}/\mathscr{F}_n) = X_n$.[11] Doob's basic *martingale convergence theorem* states that for such a martingale, if $\sup_n E(|X_n|) < \infty$, then $\lim_{n\to\infty} X_n$ is finite and exists almost everywhere (a.e.) (see Doob 1971).

Doob's application of this result is simple but ingenious. Let X be an rv such that $E(|X|) < \infty$. Then the $X_n \equiv E(X/\mathscr{F}_n)$ form a martingale ("Doob's martingale") with respect to the $\mathscr{F}_n$. If we think of the $\mathscr{F}_n$ as corresponding to the information gathered up to and including stage n, then successive conditional expectations of X as we come to know more and more yield a martingale. If the particular Doob martingale satisfies $\sup_n E(|X_n|) < \infty$, the convergence theorem guarantees that $\lim_{n\to\infty} E(X/\mathscr{F}_n)$ is finite and exists a.e. Further, if $\mathscr{F}_\infty$ denotes the smallest σ field that contains all of the $\mathscr{F}_n$, then $\lim_{n\to\infty} E(X/\mathscr{F}_n) = E(X/\mathscr{F}_\infty)$ a.e. And if $\mathscr{F}_\infty = \mathscr{F}$, then $E(X/\mathscr{F}_\infty) = E(X/\mathscr{F}) = X$.

The final step was, to my knowledge, not explicitly noted by Doob himself, but probabilists took the step to be so obvious as not to require explicit mention. Take X to be the characteristic function $[H]$ corresponding to some hypothesis H, i.e., $[H](\omega) = 1$ if H is true at $\omega \in \Omega$, 0 otherwise. $E([H]) < \infty$, and $\sup_n E([H]/\mathscr{F}_n) < \infty$. So if $[H]$ is measurable and $\mathscr{F}_\infty = \mathscr{F}$, $\lim_{n\to\infty} E([H]/\mathscr{F}_n)(\omega) = [H](\omega)$ a.e. But $E([H]/\mathscr{F}_n)$ is just the conditional probability of H on the evidence gathered through stage n. So the upshot is that if the information gathered is complete enough ($\mathscr{F}_\infty = \mathscr{F}$), then almost surely the posterior probability of H will go to 1 if H is true and to 0 if H is false.

Hesse's complaints against Savage do not apply here, since IID trials and objective likelihoods have not been assumed. In effect, the washing out launders not only different estimates of priors but also different estimates of likelihoods. As with the Savage result, merger of opinion takes place because of the almost sure convergence to certainty. In both cases, however, the merger is of a very weak form. All that is guaranteed is that for almost any world, any pair of equally dogmatic Bayesian conditionalizers, any hypothesis H, and any desired $\varepsilon > 0$, there is an N such that after the agents have seen at least N pieces of data, their opinions regarding H will differ by no more than ε. Since N may depend not only on the world and on ε but also on H and on the pair of agents chosen, the merger can be far from uniform. Stronger results on merger of opinion can be derived, as will be discussed in the following section.

5 The Results of Gaifman and Snir

Gaifman and Snir (1982) have shown how to translate the results of section 4 into a setting more in harmony with the standard philosophical discussions of confirmation theory, where probabilities are assigned to sentences of some formal language and the results of experiment and observation are reported in the form of atomic sentences or truth-functional compounds of atomic sentences. Specifically, Gaifman and Snir work in a language $\mathscr{L}$ obtained by adding empirical predicates and empirical function symbols to first-order arithmetic, assumed to contain names for each of the natural numbers $\mathbb{N}$. The Gaifman and Snir *models* $\mathrm{Mod}_{\mathscr{L}}$ for $\mathscr{L}$ consist of interpretations of the quantifiers as ranging over $\mathbb{N}$, and interpretations of the k-ary empirical predicates and k-ary function symbols respectively as subsets of $\mathbb{N}^k$ and functions from $\mathbb{N}^k$ to $\mathbb{N}$. (So, for example, if 'P' is an atomic empirical predicate, Pi might be taken to assert that the ith flip in a coin flipping experiment is heads.) A sentence φ of $\mathscr{L}$ is said to be *valid* in $\mathscr{L}(\models \varphi)$ just in case φ is true in all $w \in \mathrm{Mod}_{\mathscr{L}}$.

We can now make our starting assumption (P1) more precise by requiring that the probability axioms (A1) to (A3) from chapter 2 hold for Gaifman and Snir's $\models$ and that countable additivity holds in the form

$$\Pr((\forall_i)\eta(i)) = \lim_{n\to\infty} \Pr\left(\underset{i\leqslant n}{\&}\,\eta(i)\right), \tag{A4$'$}$$

where $\eta(i)$ is an open formula whose only free variable is i. Assumption (A4$'$) is needed for the application of the martingale theorems.

For a sentence φ of $\mathscr{L}$,

$\text{mod}(\varphi) \equiv \{w \in \text{Mod}_{\mathscr{L}}\colon \varphi \text{ is true in } w\}$.

The family of sets $\{\text{mod}(\varphi)\colon \varphi$ is a sentence of $\mathscr{L}\}$ is a field that generates a σ field $\mathscr{G}$. It is shown that for every Pr on $\mathscr{L}$ satisfying (A1) to (A4′) there is a unique countably additive $\mathscr{P}\!r$ on $\mathscr{G}$ such that $\mathscr{P}\!r(\text{mod}(\varphi)) = \Pr(\varphi)$ for every φ. Then $(\text{Mod}_{\mathscr{L}}, \mathscr{G}, \mathscr{P}\!r)$ is the mathematical probability space. For a given Pr on $\mathscr{L}$, a property is said to hold a.e. just in case it holds for a set $K \subseteq \text{Mod}_{\mathscr{L}}$ such that $\mathscr{P}\!r(K) = 1$. Now for $w \in \text{Mod}_{\mathscr{L}}$ and a sentence φ, define φ^w as φ or $\neg\varphi$ according as $w \in \text{mod}(\varphi)$ or $w \in \text{mod}(\neg\varphi)$. A class of sentences Φ is said to *separate* a set $K \subseteq \text{Mod}_{\mathscr{L}}$ just in case for any two distinct $w_1, w_2 \in K$, there is a $\varphi \in \Phi$ such that $w_1 \in \text{mod}(\varphi)$ and $w_2 \in \text{mod}(\neg\varphi)$. (If $\Phi = \{\varphi_i\}, i = 1, 2, 3, \ldots$, separates $\text{Mod}_{\mathscr{L}}$ and if $\mathscr{G}_n$ are the fields generated by $\{\text{mod}(\varphi_i)\colon i \leqslant n\}$, then $\mathscr{G}_n$ generates $\mathscr{G}$. Thus it is the separating power of the accumulating evidence that makes applicable the Doob martingale convergence results.) Finally, $\Pr_1$ and $\Pr_2$ are said to be *equally dogmatic* just in case $\mathscr{P}\!r_1$ and $\mathscr{P}\!r_2$ are mutually absolutely continuous (i.e., $\mathscr{P}\!r_1(A) = 0$ iff $\mathscr{P}\!r_2(A) = 0$ for any $A \in \mathscr{G}$). This implies the equal dogmatism of (P3) assumed above, but the converse does not necessarily hold unless $\Pr_1$ and $\Pr_2$ are definable in $\mathscr{L}$. For simplicity, then, assume that the Pr functions of the agents in the Bayesian community are definable in $\mathscr{L}$.

Using the standard martingale convergence theorems for $(\text{Mod}_{\mathscr{L}}, \mathscr{G}, \mathscr{P}\!r)$ and then transferring the results down to $\mathscr{L}$, we can establish the following result.

Gaifman and Snir Theorem Let $\Phi = \{\varphi_i\}$, $i = 1, 2, \ldots$, separate $\text{Mod}_{\mathscr{L}}$. Then for any sentence ψ of $\mathscr{L}$

1. $\Pr(\psi / \&_{i \leqslant n} \varphi_i^w) \to [\psi](w)$ a.e. as $n \to \infty$,
2. if Pr′ is as equally dogmatic as Pr, then

$$\sup_{\psi} \left| \Pr'\left(\psi \middle/ \underset{i \leqslant n}{\&} \varphi_i^w\right) - \Pr\left(\psi \middle/ \underset{i \leqslant n}{\&} \varphi_i^w\right) \right| \to 0$$

a.e. as $n \to \infty$.

Call Φ in the hypothesis of the theorem an *evidence matrix*. If Φ is separating, part 1 of the theorem shows that the evidence accumulated in almost any world by successively checking the elements of the evidence matrix serves to drive the posterior probability to certainty in the limit, and this certainty is reliable in that in almost any world where the probability goes to 1 (respectively, 0), the hypothesis ψ is true (respectively, false).[12] The rate of convergence to certainty cannot be expected to be uniform over ψ. For example, take ψ_n to assert that in a countable sequence of balls drawn from a bottomless urn, all the balls up to and including the nth are red, while the rest are green. For a reasonable Pr function one would expect that as n gets larger and larger, it takes longer and longer for certainty to set in.

This makes all the more remarkable part 2 of the theorem, which says that merger of opinion between two equally dogmatic agents is uniform over ψ, or in mathematical jargon, that the distance between two equally dogmatic Pr functions, as measured in the uniform distance metric, goes to 0. Note, however, that without further restrictions one cannot hope to show that there is merger of opinion in the strong sense of uniform convergence over the set of equally dogmatic Pr functions.[13] This would be the case, for instance, if the collection of equally dogmatic Pr functions formed a closed convex set with a finite number of extremal points (see Schervish and Seidenfeld 1990). But such additional assumptions markedly reduce the scope of the explanation of objectivity.

These results do not rest on those presuppositions of Savage's type of result, which, though plausible for the coin flipping case, are highly implausible when applied to the testing of nonstatistical hypotheses. Also the distinguishability hypothesis of the theorem is satisfied if the empirical predicates and function symbols of $\mathscr{L}$ all stand for observable properties and functions and if the evidence matrix consists of a complete enumeration of the atomic observation sentences.[14] In this case the successive checking by direct observation of the elements of the evidence matrix serves to drive the convergence to certainty and the merger of opinion.

The theorem is undeniably impressive. Indeed, it seems almost too good to be true, especially when one reflects on the fact that ψ may have a quantifier structure as complicated as you like. In chapter 9 we will learn that there is a sense in which it is too good to be true.[15]

6 Evaluation of the Convergence-to-Certainty and Merger-of-Opinion Results

Some of the prima facie impressiveness of these results disappears in the light of their narcissistic character, i.e., the fact that the notion of 'almost surely' is judged by Pr. Sentence ψ may be true in the actual world w_a and in some intuitively natural neighborhood of worlds near w_a. But if $\Pr(\psi) = 0$, $\Pr(\psi/\&_{i \leq n} \varphi_i^w)$ is also 0 in all these worlds. This does not contradict the theorem, since these worlds form a set of measure 0, as judged by Pr. From the point of view of an omniscient observer, the self-congratulatory success of the Bayesian method is hollow if the zeros of the prior distribution are incorrectly assigned. The personalist will no doubt respond by noting that in real life there are no omniscient observers and by asserting that flesh-and-blood observers have no metastandard by which to judge the correctness of Pr. Be that as it may, 'almost surely' sometimes serves as a rug under which some unpleasant facts are swept, as we will see in chapter 9.

Another qualm concerns Gaifman and Snir's semantics for $\mathscr{L}$. In the usual semantics, the models $\widetilde{\mathrm{Mod}}_{\mathscr{L}}$ are not separated by the empirical atomic sentences, so the straightforward application of the theorem to empirical testing is lost. Perhaps, however, one should not worry about the models that lie in $\widetilde{\mathrm{Mod}}_{\mathscr{L}}$ but not in $\mathrm{Mod}_{\mathscr{L}}$, since they contain nonstandard integers and thus are in some sense "impossible worlds."

Leaving aside these qualms, the convergence-to-certainty results do ground that aspect of the objectivity of Bayesian inference concerned with the long-run match between opinion and reality; at least this is so for observational hypotheses. But the merger-of-opinion results do not serve to ground objectivity qua intersubjective agreement for opinions that fall short of certainty, and this for two different sorts of reasons. The first has to do with the limit character of the results. Keynes's lament that in the long run we are all dead has no sting in the present context if we can know in advance how long the long run is. But what is lacking in the results before us is any estimate of the rate of convergence. Nor does it seem possible to derive informative estimates in the present general setting. In Savage's type of example in section 3, results about the rate of concentration of the posterior distribution are readily derivable, since all the agents agree on the statistical model that serves to fix the form of the likelihoods. In IID experiments, for example, the concentration of the posterior, as measured by the reciprocal of the variance, can be expected to grow as $\sqrt{n}$. This happy circumstance will not obtain in general, especially when the hypotheses at issue are nonstatistical.

It is not just that different Bayesian agents will give different estimates of rates of convergence but that there may be no useful way to form the estimates. To form an estimate for a given possible world we need to know what kind of evidence is received and also what bits are received in what order. A statistical model in effect specifies the relevant evidence (e.g., the outcomes of repeatedly flipping a coin), and the assumption of independent or exchangeable trials says that the order does not matter. But in the general case, the relevant evidence can come in myriad forms, and within a form the order can matter crucially. Some sort of estimate of rate of convergence could be produced by averaging over the rates for different sequences of evidence strings. However, this requires a weighting of different sequences, and it is problematic whether in the general setting there exists a weighting function that will gain the allegiance of all the Bayesian agents.

The second reason that the formal merger results do not serve to ground objectivity derives from the observation that for some aspects of the objectivity problem not only is the long run irrelevant, so is the short run. Scientists often agree that a particular bit of evidence supports one theory better than another or that a particular theory is better supported by one experimental finding than another (e.g., the data from the perihelion of Mercury better confirm Einstein's general theory of relativity than either the red-shift data or the bending-of-light data). What happens in the long or the short run when additional pieces of evidence are added is irrelevant to the explanation of shared judgments about the evidential value of present evidence.

Finally, the theorem does not suffice to demonstrate even long-run convergence to certainty and merger of opinion for theoretical hypotheses, at least not if one form of the antirealists' argument from underdetermination is correct, for the failure of the crucial distinguishability premise corresponds to one plausible explication of the notion of underdetermination of theory by evidence. This topic will be explored in the following section.

7 Underdetermination and Antirealism

The twin goals of this section are to discuss merger-of-opinion results for theoretical hypotheses and to assess a popular argument for antirealism. I begin with a discussion of the latter argument.

The underdetermination of theory by observational evidence is widely thought to weigh in favor of a nonrealist interpretation of scientific theories. But upon first reflection, it is not easy to see how underdetermination supports *semantic antirealism*, i.e., the doctrine that theoretical terms lack referential status.[16] Nor is it obvious why underdetermination supports *epistemological antirealism*, i.e., the doctrine that observational evidence gives no good reason to believe theoretical propositions, even if their constituent terms are referential.[17] After all, observational assertions about the elsewhere are underdetermined by all possible observations that can be made here, while observational assertions about the future are underdetermined by all possible past observations. But nevertheless, we may have good reason to believe observational predictions about the elsewhere and elsewhen.[18] Is there, then, something special about theoretical propositions that allows the epistemological antirealist to take a principled stand that differs from a form of blanket skepticism?

I will explore one possible answer that can be given within the confines of Bayesian confirmation theory. Antirealists have typically been leery of Bayesianism, and seemingly with good reason, since there is nothing in the Bayesian machinery to prevent the assignment of high probabilities to theoretical propositions. If the Bayesian account of scientific inference should imply that inferences to unobserved observables stand or fall together with inferences to unobservables, then in Bayesian eyes, at least, epistemological antirealism would be reduced to general inductive skepticism.

The beginnings of an antirealist response are suggested by the merger-of-opinion results discussed above. The mere assignment of a high personal probability to a proposition, theoretical or observational, by some member of the scientific community does not constitute the good reasons for belief that we want from scientific inference. In particular, the supposed objectivity of scientific inference is missing. To explore this matter further, let me say that the degree of belief in a hypothesis H is *objectifiable* with respect to a class $\{\mathrm{Pr}\}$ of probability functions just in case for a.e. $w \in \mathrm{Mod}_{\mathscr{L}}$, there is a number r such that for every suitable evidence matrix $\Phi = \{\varphi_i\}$ and every $\mathrm{Pr} \in \{\mathrm{Pr}\}$, $\mathrm{Pr}(H/\&_{i \leqslant n}\, \varphi_i^w) \to r$ as $n \to \infty$. What constitutes a "suitable" Φ may be open to dispute among empiricists of different stripes, but for present purposes, let us take suitable Φ's to consist of enumerations of the atomic observation sentences of $\mathscr{L}$. Then the convergence-to-certainty results show that for any community of scientists who operate with equally dogmatic Pr functions and for any observational H, the degree of belief in H is objectifiable (for any such H, r may be taken to be 1 or 0). Whether or not the objectification sets in within the lifetime of the average scientist is something that the convergence results do not tell us. But at least in principle there is a long-run notion of objective degree of belief for observational propositions, whether or not we are around in the long run to achieve it.

For theoretical propositions the situation is altogether different. For a start, once theoretical terms are added to the language $\mathscr{L}$, the suitable evidence matrices will no longer serve to separate $\mathrm{Mod}_{\mathscr{L}}$, and consequently, the condition for the application of the convergence result fails.[19] To extend the convergence results to theoretical hypotheses, some assumption about observational distinguishability is needed. Call the incompatible theories T_1 and T_2 *weakly observationally distinguishable* (wod) for the models MOD just in case for any $w_1, w_2 \in \mathrm{MOD}$ such that $w_1 \in \mathrm{mod}(T_1)$ and $w_2 \in \mathrm{mod}(T_2)$, there exists a (possibly quantified) observation sentence O such that $w_1 \in \mathrm{mod}(O)$ and $w_2 \in \mathrm{mod}(\neg O)$. If $\{T_i\}$ is a partition of theories that are pairwise wod for Gaifman and Snir's $\mathrm{MOD} = \mathrm{Mod}_{\mathscr{L}}$, then the degrees of belief in these theories will be objectifiable. For given any $T_j \in \{T_i\}$, $\mathrm{Pr}(T_j/\&_{i \leqslant n}\, \varphi_i^w) \to [T_j](w)$ a.e. for any suitable $\Phi = \{\varphi_i\}$.[20] But at this juncture the antirealist can interpose that the failure of wod is precisely what the underdetermination of theory by observation means (in at least one precise sense). Hence underdetermination does constitute an argument for epistemological antirealism by way of undermining the conditions needed to demonstrate the objectification of belief in theories.

This last move requires some comment. Consider the more usual and apparently stronger sense of observational distinguishability; namely, T_1 and T_2 are *strongly observationally distinguishable* (sod) for MOD just in case there is a (possibly quantified) observation sentence O such that for any $w_1, w_2 \in \mathrm{MOD}$, if $w_1 \in \mathrm{mod}(T_1)$ and $w_2 \in \mathrm{mod}(T_2)$, then $w_1 \in \mathrm{mod}(O)$ and $w_2 \in \mathrm{mod}(\neg O)$, i.e., relative to MOD, O is a consequence of T_1 and $\neg O$ is a consequence of T_2. Trivially, sod implies wod. If MOD is taken to be the usual models $\widetilde{\mathrm{Mod}}_{\mathscr{L}}$ for $\mathscr{L}$, then a simple compactness argument shows that the implication goes in the opposite direction.[21] The parallel implication is not quite so obvious if MOD is taken to be Gaifman and Snir's $\mathrm{Mod}_{\mathscr{L}}$, since $\mathrm{Mod}_{\mathscr{L}}$ is not compact even if $\mathscr{L}$ contains empirical predicate symbols but no empirical function symbols (for example, there is no model in $\mathrm{Mod}_{\mathscr{L}}$ for $\{(\exists i)Pi, \neg P1, \neg P2, \ldots\}$, even though there is a model in $\mathrm{Mod}_{\mathscr{L}}$ for every finite subset). And in fact, if T_1 and T_2 are allowed to consist of the closure under logical implication of arbitrary sets of sentences, then the implication does not hold. But if we restrict attention to the case where T_1 and T_2 are sentences, which is the case at issue, then the implication does hold.[22]

This discussion raises problems for both the Bayesian who wants the merger-of-opinion results to have bite and for the would-be epistemological realist. To take the first problem first, it might seem that the convergence-to-certainty results for theoretical hypotheses are bootless. Either wod holds for pairs of $\{T_i\}$ or not. If it does not hold, then the convergence results do not apply. If it does hold, then the convergence results do apply but are useless, for wod entails that distinct pairs of the $\{T_i\}$ have incompatible observational consequences, so one can arrive at the true theory by simple eliminative induction without using the Bayesian apparatus. In fact, however, the latter horn of this dilemma is flawed, since sod does not necessarily mean that the observational consequences of the $\{T_i\}$ are finitely verifiable or falsifiable. And if finite verifiability and falsifiability fail, the convergence results do have some bite: one converges to certainty on T_{34}, say, by making more and more atomic observations and thereby converging to 0 on the (possibly multiply quantified) observation sentences that separate T_{34} from its rivals.[23] Of course, the worries about rates of convergence raised above apply here as well.

I now turn to a discussion of how the would-be epistemological realist might respond to the underdetermination argument. First, he could grant the force of the move from underdetermination to antirealism but maintain that underdetermination does not pose a serious threat, because either it is not widespread or else occurs in uninteresting varieties. Starting from a theory and tacking on theoretical epicycles that add no new observational predictions would produce an endless string of observationally indistinguishable theories, but this form of underdetermination is uninteresting, since the core theoretical content is the same in every case. Theories of gravitation that are observationally indistinguishable and that make interestingly different theoretical commitments can be created if they are permitted to remain silent about classical gravitational tests. But completeness (in the sense of yielding definite predictions) with respect to the phenomena belonging to the commonly agreed-upon explanatory domain of gravitation would seem to be a reasonable demand to impose on theories of gravitation worthy of consideration (see, for example, Will 1972). Indeed, it could be held to be a necessary condition for calling a set of axioms a theory of gravitation. Whether there are explanatorily complete and observationally indistinguishable theories of gravity that make

interestingly different theoretical posits is a question that will be taken up in chapter 7.[24]

Second, the realist could deny that underdetermination does support epistemological antirealism by denying the antirealist's identification of good reasons to believe with objectified degree of belief in the Bayesian sense of merged posterior opinion. To repeat, past observations, even if they stretch infinitely far into the past, do not serve to objectify observational predictions about the future for a broad class $\{Pr\}$ of equally dogmatic probability functions.[25] But nonetheless, one might claim that past experience does give good reason to believe that the sun will rise tomorrow and that the emeralds seen in the dawn of this new light will be green. Similarly, the realist may hold that we can have good reasons to believe theoretical claims even if the degree of belief is not objectifiable in the technical sense offered above. I am sympathetic to this point of view, but it is unavailing in the present context, which seeks to discern the implications of Bayesianism for the realism versus antirealism controversy. For in its current stage of development, the Bayesian account of scientific inference contains no explication of objective good reasons other than the forced merger of subjective opinion or the apparently unworkable schemes for objectifying assignments of priors. The Bayesianized version of the realist versus antirealist debate thus grinds to a halt over the unresolved problem of objectivity.

8 Confirmability and Cognitive Meaningfulness

I suggested above that the epistemological antirealist who does not wish to be a vulgar skeptic may run afoul of Bayesianism, since quashing Bayesian inferences to unobservables threatens to quash inferences to unobserved observables. The strength of this objection is open to debate, but we need not settle the debate to recognize that the objection can be turned around to cast doubt on Bayesian inference. If Bayesians can assign nonzero priors to hypotheses about such unobservable entities as quarks, then it would seem that they can also assign nonzero priors to hypotheses about vital forces, devils, and deities. Consequently, Bayesianism faces the embarrassment of countenancing inductive arguments in favor of (or against) such hypotheses.

Perhaps the embarrassment can be faced down with a divide-and-conquer strategy.

Case 1. The hypothesis 'Jehovah exists and rules the world' (J) is so construed that it does make a difference for the probabilities of pieces of observational evidence E about, say, the amount of suffering in the world ($\Pr(E/J \,\&\, K) \neq \Pr(E/K)$). Then Bayes's theorem shows how and why (J) is confirmed (or disconfirmed) by E. So contrary to first impressions, we can properly speak of inductive arguments for the existence of God (see Swinburne 1979).
Case 2. 'Jehovah exists and rules the world' is construed so that it doesn't make a probabilistic difference for any observational evidence $E(\Pr(E/J \,\&\, K) = \Pr(E/K))$. Then Bayes's theorem shows why (J) is immune to inductive considerations. In this case the embarrassment doesn't need to be explained away, since it doesn't arise.

The positivists and logical empiricists held that 'Jehovah exists and rules the world' and its like are not real hypotheses, since (in their jargon) these inscriptions are "cognitively meaningless." Initially the positivists favored verifiability/falsifiability as the identifying mark of the cognitively meaningful, but when this criterion ran into difficulties, they switched to confirmability/disconfirmability.[26] If the latter criterion is to be implemented through Bayesian personalism, then it must be conceded after all that 'Jehovah exists and rules the world' can be cognitively meaningful. To the extent that positivists and logical empiricists balk at such a conclusion, their views clash with the Bayesianism promoted here. Whether the clash is just another nail in the coffin of a dying philosophical movement or whether it is a mark against Bayesianism is a matter that I will leave to the reader to decide.

9 Alternative Explanations of Objectivity

I turn now to an examination of some of the alternative explanations (3) to (8) listed in section 1. The idea behind (3) is that a definitional ploy may succeed where honest theorem proving has failed. The notions of rationality and objectivity are relativized to a scientific community and 'community' is defined in terms of merger of opinion over the relevant range of hypotheses. This move threatens to reactivate the buzz of relativism I assumed at the beginning of the chapter to have been cured. Therefore, further discussion of this alternative will be postponed to chapter 8. Chapter 8 will also take up one form of (4), socialism in the guise of a rule for manufacturing a consensus by means of a prescription for aggregating opinions. It is not giving away too much to anticipate the conclusion that neither (3) nor (4) holds the answer to our prayers.

The remainder of this chapter will be devoted to a discussion of (5), the evolutionary solution; (6), modest but realistic solutions; (7), non-Bayesian solutions; and (8), retrenchment.

10 The Evolutionary Solution

The results of Savage and Doob discussed above have exercised a fascination not only because they entail merger of opinion but also because they reveal a link that joins Bayesian methods to truth and reliability. But because it is forged only in the infinite limit, the link revealed in the formal theorems is too weak.

A partnership between Darwin and Bayes might be thought to supply the missing link for the medium and short runs. The idea of the partnership is, first, that evolution has produced a species for which rapid merger of opinion (not necessarily to 1 or 0) takes place and, second, that the evolutionary story of this merger has the sought-after justificatory character in that our degrees of belief are reliable estimates of the actual frequencies of relevant events, since otherwise we would not have survived.

The ideas of van Fraassen (1983a) and Shimony (1988) mentioned in chapter 2 can be used to give an account of what it means for degrees of belief to be reliable estimates of frequencies, at least for simple atomic hypotheses. It is far from clear, however, what is meant by saying that my degree of belief of .75 in Einstein's GTR is a reliable estimate of a frequency. Talk about the frequency with which hypotheses of this sort have proven to be true is vague, but insofar as I understand it, the relevant frequency would seem to be 0. I can calibrate my degree of belief in Einstein's GTR with frequencies by finding an H for which my $\Pr(H)$ is naturally interpreted as an estimate of a frequency and for which I set $\Pr(H) = \Pr(\text{GTR})$. But such calibration involves subjective judgments.

Even if the Darwin and Bayes partnership had an unproblematic statement, there would still be two obstacles to implementing it. In the first place, there is no obvious Darwinian edge to reliability of beliefs about the esoteric matters that lie at the core of modern science. Case after case could be cited from the history of science where scientists developed a strong consensus that the hidden springs of nature followed, at least approximately, the dictates of a certain theory only to become convinced at a later stage that the theory was badly flawed. In the second place, while there may be a class of propositions for which a rapid and accurate process for fixing degrees of belief was essential to survival during humankind's formative stages (e.g., 'Tiger near'), it isn't clear how far this class extends even into the realm of mundane affairs. Thus, despite the importance of weather to prosperity and even survival itself, historically, our weather forecasts have been notoriously unreliable. Perhaps we have prospered as a species not because of any general reliability of belief-fixing processes but because we are robust enough to tolerate or creative enough to maneuver around the consequences of the unreliabilities in this process.[27]

11 Modest but Realistic Solutions

The washout theorems studied above had the lofty aim of underwriting a global consensus, but because of their limit character, they proved to be incapable of explaining the consensus that exists now. This actual consensus is partial rather than sharp and spotty rather than global. Its partial and spotty nature make it at once easier and more difficult to explain—easier because there is less to explain, and more difficult because the explanation will not be uniform but will consist of disparate pieces. Here I will concentrate on explaining such comparative judgments as evidence E confirms H_1 more than it confirms H_2, or E_1 confirms H more than does E_2.

The former case seems especially difficult to deal with. Suppose, for example, that H_1 and H_2 are both hypotheticodeductively confirmed by E relative to the background knowledge K (i.e., $\{H_1, K\} \models E$ and $\{H_2, K\} \models E$). The incremental Bayesian confirmations of H_1 and H_2 are respectively

$$\Pr(H_1/K)[(1/\Pr(E/K)) - 1]$$

$$\Pr(H_2/K)[(1/\Pr(E/K)) - 1],$$

and the absolute confirmations are respectively $\Pr(H_2/K)/\Pr(E/K)$ and $\Pr(H_2/K)/\Pr(E/K)$. Thus, on the Bayesian analysis, any judgment to the effect that E is better evidence (in either the incremental or absolute sense)

for H_1 than for H_2 boils down to the judgment that $\Pr(H_1/K) > \Pr(H_2/K)$, and we are back in the middle of the swamp of the problem of priors.

The hope burns brighter when the case concerns the way in which different pieces of evidence bear on the same hypothesis or theory. Consider the three classical tests of Einstein's GTR. As noted in chapter 5, it is generally agreed by physicists that the evidence E_P of the advance of Mercury's perihelion gives more support to GTR than does the evidence E_B of the bending of light or the evidence E_R of the red shift. On the Bayesian analysis, the incremental confirmation values are $\Pr(\text{GTR}/K) \times [(1/\Pr(E_{P,B,R}/K)) - 1]$. Since the prior probability factor is the same in all three cases, the focus shifts to the prior-likelihood factors $\Pr(E_{P,B,R}/K)$. (Here we run smack into the problem of old evidence [see chapter 5], which is a thorn in the side of Bayesianism confirmation theory. I am just going to ignore it for present purposes.) Can we show that judgments about these prior likelihoods have an objective basis?

Here is one attempt. Imagine a complete enumeration $\{T_i\}$ of alternative theories of gravity, and suppose that each theory yields a definite prediction about the three classical tests.[28] By total probability,

$$\Pr(E_{P,B,R}/K) = \sum_i \Pr(E_{P,B,R}/T_i \,\&\, K) \times \Pr(T_i/K).$$

By assumption, the first factors in the sum on the right-hand side are all either 0 or 1, so the sum reduces to the sum of the priors of those theories that successfully explain the results of the test in question. Thus if it could be shown that the set of theories that succeed with respect to E_P is a proper subset of each of the sets of theories successful with respect to E_B and E_R, it would follow that, independently of judgments of the prior probabilities of the theories, E_P gives a better confirmational value than either E_B or E_R.

As mentioned in chapter 5, to first-order approximation, the most general stationary spherically symmetric line element can be written as

$$ds^2 = [1 - (\alpha m/r) + (2\beta m^2/r^2)]dt^2 - [1 + (2\gamma m/r)](dx^2 + dy^2 + dz^2).$$

GTR sets the parameters α, β, and γ equal to 1. The perihelion shift depends on all three parameters, while the red shift depends only on α and the bending of light only on α and γ. Does it follow that any theory that successfully explains the perihelion shift must also explain the red shift and the bending of light? Not necessarily, for a theory can get the red shift and bending of light wrong but by compensating errors get the perihelion right. So it seems that our judgments in this case cannot be divorced from judgments about prior probabilities of theories.

Still, the Bayesian might claim a partial victory here on the grounds that he has to explain not why E_P gives better confirmational value than E_R or E_B (for in fact it may not) but only why it was thought that this was so. The long history of failures to explain the perihelion phenomenon (see Earman and Glymour 1991) coupled with the ready availability of multiple alternative explanations of the red shift perhaps explains why, around the time GTR was introduced, most physicists would have set $\Pr(E_P/K) < \Pr(E_R/K)$ and thus $\Pr(\mathrm{GTR}/E_P \,\&\, K) > \Pr(\mathrm{GTR}/E_P \,\&\, K)$.This explanation doesn't hold today, when many of the presently available members of the zoo of alternative theories of gravity explain the perihelion shift (see chapter 7).

12 Non-Bayesian Solutions

At present this is an empty label, since there aren't any extant non-Bayesian accounts of scientific inference that have proved to be viable across the broad range of cases. As one example of dashed hopes, I would cite Hempel's account of qualitative confirmation and Glymour's attempt to extend Hempel's ideas to the confirmation of theoretical hypotheses by means of bootstrapping relations. One might have hoped that Hempel's confirmation relations and Glymour's boostrapping relations, which are purely logicostructural relations, could provide at least part of the basis for objectivity. Alas, as we saw in chapter 3, the Bayesian apparatus is needed before any conclusions can be drawn about the bearing of these relations on the credibility of hypotheses. Other examples of dashed hopes could be cited, but enough tears have already been shed.

13 Retrenchment

If (1) through (7) of section 1 all fail, the only resort would seem to be a retrenchment to a more modest set of goals for a theory of confirmation and scientific inference, as suggested by conceiving the theory as constituting an inductive logic that parallels deductive logic. Deductive logic provides a neutral framework for evaluating deductive arguments. It is

neutral in the sense that it doesn't tell us which contingent statements to accept as true. But it is not lacking in bite, since it does tell us that if we accept certain statements as true, then on pain of inconsistency we must accept certain other statements as true and reject still others as not true. On this analogy, inductive logic can be thought to provide a neutral framework for evaluating inductive arguments. It is neutral in that it doesn't tell us what degrees of belief to assign to contingent propositions. But it does have bite in that it tells us that if we assign such and such degrees of belief to such and such propositions, then on pain of inconsistency (i.e., incoherency) we must also assign specified degrees of belief to other propositions.

One might hope for a bit more than this from a theory of confirmation, although the more calls for work on our part. Consider the EUREKA! cartoon that appeared recently in the *Toronto Globe*. Why is the cartoon amusing? The part of the explanation of the humor relevant to present concerns is simply that there is in fact a sharp consensus about the outcomes of the "unnecessary experiments"—that is what makes them unnecessary. However, the basis of this consensus remains to be investigated. The worst-case result of the investigation would find a consensus built on sand, a consensus that obtains not because of merger of opinion forced by the accumulated evidence but because members of our community have given in to social pressures to conform. A better-case result would find a de facto washing out of priors. That is, the actually accumulated evidence does force a merger of opinion for the class of actual belief functions with which the members of our community have been endowed. But this class is relatively narrow, and when it is expanded with additional belief functions

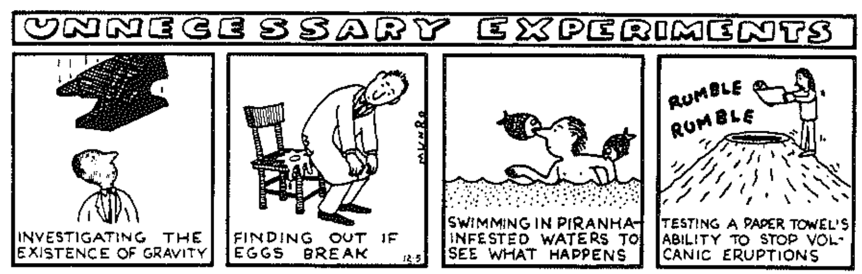

Figure 6.1
EUREKA! created by Munro Ferguson, copyright 1989, distributed by Universal Press Syndicate

equally dogmatic with respect to ours, merger no longer takes place. Ever better cases are reached as this class expands until we reach the best-case result, where the consensus is very solid in that it arises for a maximal class of equally dogmatic belief functions that assign nonzero priors to the phenomena in question.

The results of such investigations will color the Bayesian interpretation of consensual degrees of belief. Where the consensus is one of the best-case types, the degrees of belief may deservedly be labeled as objective. But as we shade toward the worse-case end of the spectrum, scare quotes will need to be added to the label, and eventually the label may be withdrawn altogether. Whatever the decision, the Bayesian will insist that his apparatus is equal to drawing the relevant distinctions. And hankering after some form of objectivity beyond the ken of these distinctions is to hanker after the unobtainable.

Without taking any final stand on this issue, I want to agree partly with the Bayesians in insisting that investigations of the kind outlined above need to be carried out in detail. What I very much fear, however, is that these investigations will not reveal any strong Bayesian basis for claiming objectivity for the opinions so confidently announced at the beginning of this chapter or for the opinions implicitly endorsed in the EUREKA! cartoon. Certainly the discussion of the general problem of induction and the grue problem in particular shows that merger of opinion on hypotheses about the future cannot be forced even in a limit sense for a maximal class of equally dogmatic belief functions by any amount of evidence about the past, since these hypotheses are underdetermined by all past evidence. And I suspect that the evidence actually gathered forces the current consensuses in science only for a very circumscribed class of belief functions. Unless reasons can be found to privilege this class over others, the door is open to relativism, social constructivism, and other equally horrific isms.

14 Conclusion

In a certain mood I am all for upholding scientific common sense and for proclaiming that the presently available evidence does justify high confidence in the propositions that Velikovsky's *Worlds in Collision* is humbug, that space and time are relativistic rather than absolute, that the next emerald we examine will be green, etc. For those in a like mood, the drift

of this chapter indicates that Bayesian personalism must either be supplemented or else rejected altogether as an account of scientific inference.

Some Bayesians would respond, "Scientific common sense be damned!" For them, there is no question of rejecting Bayesianism as an account of scientific inference, since (they proclaim) such an account must be couched in terms of degrees of belief and since what Bayesianism provides is rationality constraints on degrees of belief. Nor is there any question of supplementing Bayesianism, since to go beyond Bayesianism is to go beyond the "logic" of inductive inference. The supplementing principles must, therefore, be substantive in nature, and as Hume taught us, any justification for such principles must produce a regress or a vicious circle.

I trust that the reader of previous chapters will be convinced that the first part of this response is unacceptable. Bayesianism without a rule of conditionalization is hamstrung, but the attempted demonstrations of conditionalization do not succeed in showing that it is a constraint of rationality. And in chapter 9, I will argue that Bayesians cannot consistently maintain an attitude of evenhanded neutrality and at the same time prove merger-of-opinion and convergence-to-certainty results, for a Bayesianism strong enough to yield these results can be shown to embody what look suspiciously like substantive assumptions about the world. The principle at issue here is countable additivity. But even finite additivity does not enjoy an unquestioned status as a sine qua non of rationality (see Schick 1986 and the discussion of chapter 2 above).

I am enough of a non-Bayesian that I do not think that any a priori considerations block a non-Bayesian account of scientific inference. But when I survey the shortcomings of the non-Bayesian accounts that have been attempted, I despair that any such approach will work. In the grip of such despair, one might seek refuge in Goodman's circle: "An inductive inference... is justified by conformity to general rules, and a general rule by conformity to accepted inductive inferences. Predictions are justified if they conform to valid canons of induction; and the canons are valid if they accurately codify accepted inductive practice" (1983, p. 64). I do not doubt that this circle is virtuous rather than vicious. But the notion that the circle provides a resting place is an illusion. For the only uniformly accepted "general rules" or "canons" of induction are so near triviality as to make Goodman's circle so small that it cannot encompass any interesting scientific inferences. And it is unclear how to widen the circle without opening it to the full scope of rampant Bayesian personalism.