Department of Philosophy   Dietrich College of Humanities and Social Sciences

2004

Why Probability Does Not Capture the Logic of Scientific Justification

Kevin T. Kelly
Carnegie Mellon University

Clark Glymour
Carnegie Mellon University

Why Probability Does Not Capture the Logic of Scientific Justification

1. Introduction

Here is the usual way philosophers think about science and induction. Scientists do many things— aspire, probe, theorize, conclude, retract, and refine— but successful research culminates in a published research report that presents an argument for some empirical conclusion. In mathematics and logic there are sound deductive arguments that fully justify their conclusions, but such proofs are unavailable in the empirical domain because empirical hypotheses outrun the evidence adduced for them. Inductive skeptics insist that such conclusions cannot be justified. But "justification" is a vague term— if empirical conclusions cannot be established fully, as mathematical conclusions are, perhaps they are justified in the sense that they are partially supported or *confirmed* by the available evidence. To respond to the skeptic, one merely has to *explicate* the concept of confirmation or partial justification in a systematic manner that agrees, more or less, with common usage and to observe that our scientific conclusions are confirmed in the explicated sense. This process of explication is widely thought to culminate in some version of Bayesian confirmation theory.

Although there are nearly as many Bayesianisms as there are Bayesians, the basic idea behind Bayesian confirmation theory is simple enough. At any given moment a rational agent is required to assign a unique degree of belief to each proposition in some collection of propositions closed under "and", "or", and "not". Furthermore, it is required that degrees of belief satisfy the axioms of probability. *Conditional probability* is defined as follows:

$$P(h|e) = \frac{P(h \text{ and } e)}{P(e)}.$$

Confirmation can then be explicated like this:

evidence *e* *confirms* hypothesis *h* (for agent *P*) if and only if $P(h|e) > P(h)$.

In other words, confirmation is just positive statistical dependence with respect to one's

degrees of belief prior to their modification in light of e. In a similar spirit, the *degree* of confirmation can be explicated as the difference $P(h|e) - P(h)$.

So defined, confirmation depends on the structure of the prior probability function P to the extent that for some choice of P, the price of tea in China strongly confirms that the moon is green cheese. Personalists embrace this subjectivity, whereas objective Bayesians impose further restrictions on the form of P to combat it (cf. P. Maher's Chapter 3).

Confirmation theory's attractiveness to philosophers is obvious. First, it responds to the skeptic's challenge not with a proof, but with a conceptual analysis, conceptual analysis being the special skill claimed by analytic philosophers. Second, confirmation theorists seem to derive the pure "ought" of scientific conduct from the "is" of manifest practice and sentiment. No further argument is required that confirmation helps us *accomplish* anything, like finding the truth, for confirmation is (analytically) justification of belief and justification of belief that h is justification of belief that h is true. Third, in spite of the dependence on prior probability, Bayesian confirmation provides a simple unification of a variety of qualitative judgments of evidential relevance, some of which will be described below (cf. P. Maher's Chapter 3). Fourth, explications are hard to refute. Celebrated divergences between human behavior and the Bayesian ideal can be chalked up as "fallacies" due to psychological foible or computational infeasibility. What matters is that the explication provides a unified, simple explanation of a wide range of practice and that when violations are called to attention, the violator (or at least you, as a third party), will agree that the violation should be corrected (Savage 1951). Fifth, there are unexpected, a priori arguments in favor of Bayesian principles called Dutch Book arguments (DeFinetti 1937, Teller 1973). The basic idea is that you can't guard against possible disasters when you bet on the future, but at least you can guard against *necessary* disasters (i.e., combinations of bets in which one loses no matter what). It is then argued that Bayesian methodology is the unique way to avoid preferences for sure-loss bets. That isn't what anybody ever thought scientific method is for, but who ever said that philosophy can't make novel discoveries? Finally, confirmation theory restricts philosophical attention to a tractable subdomain of scientific practice. Confirmation stands to science as proof stands to mathematics. There are interesting psychological questions about how we find

proofs, but regardless of their intrinsic psychological and sociological interest, such issues are irrelevant to the resulting proof's validity. Similarly, science is an ongoing social process that retracts, repairs, and revises earlier theories; but the philosophical relevance of these social, psychological, and historical details is "screened off" by confirmation (Hempel 1965, Laudan 1980).

So what's not to like? One might dwell upon the fact that scientists from Newton through Einstein produced, modified, and refined theories without attaching probabilities to them, even though they were capable of doing so had they so desired (Glymour 1980). Or on the fact that the sweeping consistency conditions implied by Bayesian ideals are computationally and mathematically intractable even for simple logical and statistical examples (Kelly and Schulte 1995). Or on the fact that, in Bayesian statistical practice, the selection of prior probabilities often has more to do with mathematical and computational tractability than with anyone's genuine degrees of belief (Lee 1989). Or on gaps in the Bayesians' pragmatic Dutch book arguments (Kyburg 1978, Maher 1997, Levi 2002). Or on the fact that Bayesian ideals are systematically rejected by human subjects even when cognitive loading is not at issue (e.g., Ellsberg 1961 , Allais 1953, Kahneman and Tversky 1972). Or on the fact that a serious attempt to explicate real scientific practice today would reveal classical (non-Bayesian) statistics texts on the scientist's bookshelf and classical statistical packages running on her laboratory's desktop computer.

Our objection, however, is different from all of the above. It is that Bayesian confirmation is not even the right *sort* of thing to serve as an explication of scientific justification. Bayesian confirmation is just a change in the current output of a particular strategy or method for updating degrees of belief, whereas scientific justification depends on the truth-finding *performance* of the methods we use, whatever they might be. The argument goes like this.

1. Science has many aims, but its most characteristic aim is to find true answers to one's questions about nature.

2. So scientific justification should reflect how intrinsically difficult it is to find the truth and how efficient one's methods are at finding it. Difficulty and efficiency can be understood

3

in terms of such cognitive costs as errors or retractions of earlier conclusions prior to convergence to the truth.

3. But Bayesian confirmation captures neither: conditional probabilities can fluctuate between high and low values any number of times as evidence accumulates, so an arbitrarily high degree of confirmation tells us nothing about how many fluctuations might be forthcoming in the future or about whether an alternative method might have required fewer.

4. Therefore, Bayesian confirmation cannot explicate the concept of scientific justification. It is better to say that Bayesian updating is just one method or strategy among many that may or may not be justified depending on how efficiently it answers the question at hand.

The reference to truth in the first premise is not to be taken too seriously. We are not concerned here with the metaphysics of truth, but with the problem of induction. Perhaps science aims only at theories consistent with all future experience or at theories that explain all future experience of a certain kind or at webs of belief that don't continually have to be repaired as they encounter novel surface irritations in the future. Each of these aims outruns any finite amount of experience and, therefore, occasions skeptical concerns and a confirmation theoretic response. To keep the following discussion idiomatic, let "truth" range over all such finite-evidence-transcending cognitive goals.

The second premise reflects a common attitude toward procedures in general: means are justified insofar as they efficiently achieve our ends. But immature ends are often infeasible— we want everything yesterday with an ironclad warranty. Growing up is the painful process of learning to settle for what is possible and learning to achieve it efficiently. Truth is no exception to the rule: there is no foolproof, mechanical process that terminates in true theories. Faced with this dilemma, one must either give up on finding the truth or abandon the infeasible requirement that inductive procedures must *halt* or signal success when they succeed. Confirmation theorists adopt the first course, substituting confirmation, which can be obtained for sure, for truth, which cannot. We prefer the latter option, which retains a clear connection between method and truth-finding (Kelly 2000). As William James (1948)

wryly observed, no bell rings when science succeeds, but science may, nonetheless, slip across the finish line unnoticed.[1] Although one cannot demand that it do so smoothly and without a hitch, one may hope that due diligence will at least minimize the number of ugly surprises we might encounter as well as the elapsed time to their occurrence. Empirical justification is not a static Form in Plato's heaven waiting to be recollected through philosophical analysis. If it is anything at all worth bothering about, it is grounded in the intrinsic difficulty of finding the truths we seek and in the relative efficiency of our means for doing so.

The third premise is the crucial one. Confirmation theory encourages hope for more than efficient convergence to the truth by promising some sort of "partial justification" the short run. The terms "partial justification" and "partial support" hint at something permanent, albeit incomplete. But high degrees of belief are not permanent at all: arbitrarily high confirmation can evaporate in a heartbeat and can fluctuate between extremes repeatedly, never providing a hint about how many bumps might be encountered in the future or about whether some other method could guarantee fewer. Hence, Bayesian confirmation, or any other notion of confirmation that can be arbitrarily high irrespective of considerations of truth-finding efficacy, cannot explicate scientific justification.

Shifting the focus from confirmation relations to the feasibility and efficiency of truth-finding turns traditional, confirmation-based philosophy of science on its head. Process, generation, refinement and retraction — the topics relegated to the ash-heap of history (or sociology or psychology) by confirmation theorists — are placed squarely in the limelight. Confirmation, on the other hand, is demoted to the status of a cog in the overall truth-finding process that must earn its keep like all the other parts. At best, it is a useful heuristic or defeasible pattern for designing efficient methods addressed to a wide range of scientific questions. At worst, rigid adherence to a preconceived standard of confirmation may prevent one from finding truths that might have been found efficiently by other means (Kelly and Schulte 1995, Osherson and Weinstein 1988).

---

[1]The fallible convergence viewpoint on inquiry was urged in philosophy by James, Peirce, Popper, Von Mises, Reichenbach, and Putnam. Outside of philosophy, it shows up in classical estimation theory, Bayesian convergence theorems, and computational learning theory. Several of the points just made (e.g., non-cumulativity) were argued on purely historical grounds by Kuhn and others.

2. Inductive Performance and Complexity

Our approach focuses on problems rather than on methods. An *empirical problem* is a pair (q, k) consisting of a *question q*, which specifies a unique, correct answer for each possible world and a *presupposition k* that restricts the range of possible worlds in which success is required. Let the presupposition be that you are watching a light that is either blue or green at each stage and that is otherwise unconstrained. The question is what color the light will be at the very next stage. Then there is an easy method for *deciding* the problem with certainty: simply wait and report what you see. This procedure has the attractive property that it halts with the right answer whatever the answer happens to be. But that is clearly because the next observation entails the right answer, so the problem is not really inductive when the right answer is obtained.

Next, consider the properly inductive question whether the color will remain green forever. The obvious procedure guesses that the color won't change until it does and then halts with the certain output that it does. This procedure is guaranteed to converge to the right answer whatever it is, but never yields certainty if the color never changes. It simply keeps waiting for a possible color change. We may say that such a method *refutes* the unchanging color hypothesis with certainty. This "one-sided" performance is reminiscent of Karl Popper's (1959) "anti-inductivist" philosophy of science— we arrive at the truth, but there is no such thing as accumulated "support" for our conviction, aside from the fact that we must leap, after a sufficiently long run of unchanging colors, to the conclusion that the color will never change *if* we are to converge to the right answer in the limit. In other words, the inductive leap is not pushed upward or supported by evidence; it is *pulled* upward by the aim of answering the question correctly.

One would prefer a "two-sided" decision procedure to the "one-sided" refutation procedure just described, but no such procedure exists for the problem at hand. For suppose it is claimed that a given method can decide the question under consideration with certainty. Nature can then feed the method constantly green experience until it halts with the answer that experience will always remain green (on pain of not halting with the true answer in that case). The method's decision to halt cannot be reversed, but Nature remains free to

present a color change thereafter. In the stream of experience so presented, the method converges to the wrong answer "forever green", which contradicts the reductio hypothesis that the method converges to the truth. So by reductio ad absurdum, no possible method decides the question with certainty. This is essentially the classical argument for inductive skepticism. We recommend the opposite conclusion: since no decision procedure is feasible in this case, a refutation procedure yields the best feasible sort of performance and, hence, is justified in light of the intrinsic difficulty of the problem addressed. Of course, this "best-we-can-do" justification is not as satisfying as a "two-sided" decision procedure would be, but *that* kind of performance is impossible and the grown-up attitude is to obtain the best possible performance as efficiently as possible rather than to opine the impossible.

Next, suppose that the question is whether the color will never change, changes exactly once, or at least twice. The obvious method here is to say "never" until the color changes, "once" after it changes the first time and "at least twice" thereafter. The first problem (about the color tomorrow) requires no retractions of one's initial answer. The second question (about unchanging color) requires one, and this question requires two. It is easy to extend the idea to questions requiring three, four, etc. retractions. Since retractions, or non-cumulative breaks in the scientific tradition, are the observable signs of the problem of induction in scientific inquiry, one can measure the *intrinsic difficulty* or *complexity* of an empirical question by the least number of retractions that Nature could exact from an arbitrary method that converges to the right answer.

Some problems are not solvable under any fixed, finite bound on retractions. For example, suppose it is known a priori that the color will change only finitely often and the question is how many times it will change. Then Nature can lead us to change our minds any number of times by adding another color change just after the point at which we are sure we will never see another one. But the question is still *decidable in the limit* in the sense that it is possible to converge to the right answer, whatever it might be (e.g., by concluding at each stage that the color will never change again).

There are also problems for which no possible method can even converge to the truth in the limit of inquiry. One of them is a Kantian antinomy of pure reason: the question whether

7

matter is infinitely divisible. Let the experiment of attempting to cut matter be successively performed (failures to achieve a cut are met with particle accelerators of ever higher energy). Nature can withhold successful cuts until the method guesses that matter is finitely divisible. Then she can reveal cuts until the method guesses that matter is infinitely divisible. In the limit, matter is infinitely divisible (new cuts are revealed in each "fooling cycle"), but the method does not converge to "infinitely divisible". Nonetheless, it is still possible to converge to "finitely divisible" if and only if matter is only finitely divisible: just answer "finitely divisible" while no new cuts are performed and "infinitely divisible" each time a new cut is performed. Say that this method *verifies* finite divisibility in the limit. The same method may be said to *refute* infinite divisibility in the limit, in the sense that it converges to the alternative hypothesis just in case infinite divisibility is false. These "one-sided" concepts stand to decision in the limit as verificiation and refutation with certainty stand to decision with certainty. Of course, we would prefer a two-sided, convergent, solution to this problem, but none exists, so one-sided procedures are justified insofar as they are the best possible.

There are even questions that are neither refutable in the limit nor verifiable in the limit, such as whether the limiting relative frequency of green observations exists. This problem has the property that a method could output "degrees of belief" or "confirmation values" that converge to unity if and only if the limiting relative frequency exists, but no possible method of this kind converges to unity if and only if the limiting relative frequency does not exist. And then there are problems that have neither of these properties. At this point it starts to sound artificial to speak of convergent success in any sense.

Changes in background information can affect solvability. For example, let the question be whether a sequence of observed colors will converge to blue. Absent further background knowledge, the best one can do is to verify convergence to blue in the limit (the analysis is parallel to that of the finite divisibility example). But if we know a priori that the sequence of colors will eventually stabilize (e.g., the current color corresponds to which magnet a damped pendulum is nearest to), then the question is decidable in the limit: just respond "yes" while the color is blue and respond "no" otherwise. This shows that extra assumptions may make a problem intrinsically easier to solve without giving the game away altogether.

In mathematical logic and computability theory, it is a commonplace that formal problems have *intrinsic complexities* and that a problem's intrinsic complexity determines the best possible sense in which a procedure can solve it. For example, the unavailability of a decision procedure for the first-order predicate calculus justifies the use of a one-sided verification procedure for inconsistency and of a one-sided refutation procedure for consistency. The notion that background presuppositions can make a problem easier is also familiar, for the predicate calculus is decidable if it is known in advance that all encountered instances will involve only monadic (one-place) predicates. All we have done so far is to apply this now-familiar computational perspective, which has proven so salutary in the philosophy of deductive reasoning, to the empirical problem of induction.

Philosophers of science are accustomed to think in terms of confirmation and underdetermination rather than in terms of methods and complexity, but the ideas are related. Bayesian updating, on our view, is just one method among the infinitely many possible methods for attaching numbers to possible answers. Underdetermination is a vague idea about the difficulty of discerning the truth of the matter from data. We propose that a problem's intrinsic complexity is a good explication for this vague notion, since it determines the best possible sense in which the problem is solvable.

Using retractions to measure inductive complexity is a more natural idea than it might first appear. First, the concepts of refutability and verifiability have a long standing in philosophy and these concepts constititute just the first step in the retraction hierarchy (verifiability is success with one retraction starting with initial guess $\neg h$ and refutability is success with one retraction starting with h). Second, Thomas Kuhn (1970) emphasized that science is not cumulative because in episodes of major scientific change some content of rejected theories is lost. These are retractions. Kuhn also emphasized the tremendous cost of cognitive retooling that these retractions occasion. Unfortunately, he did not take the next logical step of viewing the minimization of retractions as a natural aim that might provide alternative explanations of features of scientific practice routinely explained along confirmation-theoretic lines. Third, by generalizing resource bounds in a fairly natural way (Freivalds and Smith 1993, Kelly 2002), one can obtain the equation that each retraction is worth infinitely many errors (ω many, to be

precise), so that the aim of minimizing the number of errors committed prior to convergence generates exactly the same complexity classes as minimizing retractions. Fourth, the concept of minimizing retractions is already familiar in logic and computability. In analysis, retraction complexity is called *difference* complexity (Kuratowski 1966) and in computability, it is known as "n-trial" complexity, a notion invented by Hilary Putnam (1965). Finally, the idea has been extensively studied in empirical applications by computational learning theorists (for an extended summary and bibliography, cf. Jain et al. 1999).

3. Explanations of Practice

Bayesian confirmation theorists have some foundational arguments for their methods (e.g., derivation from axioms of "rational" preference, Dutch book theorems) but one gets the impression that these are not taken too seriously, even by the faithful. What really impresses confirmation theorists is that Bayesian updating provides a unified, if highly idealized, explanation of a wide range of short-run judgments of evidential relevance. Such explanations are facilitated *Bayes' theorem*, a trivial logical consequence of the definition of conditional probability:

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}.$$

It follows immediately, for example, that initial plausibility of h is good ($P(h)$ is upstairs), that prediction of e is good and refutation by e is bad ($P(e|h)$ is upstairs) and that surprising predictions are good ($P(e)$ is downstairs). Successive confirmation by instances has diminishing returns simply because the sum of the increases is bounded by unity. These explanations are *robust*: they work for any prior probability assignment such that $P(e) > 0$. Other explanations depend on prior probability. For example, it seems that black ravens confirm "all ravens are black" better than white shoes, and under some plausible assignments of prior probability, this judgment is accommodated. Under others, it isn't (cf. P. Maher's Chapter 3).

The trouble with this naively hypothetico-deductive case for Bayesianism (even by Bayesian standards) is that it ignores competing explanations. In particular, it ignores the possibility that some of the same intuitions might follow from truth-finding efficiency itself, rather than from the details of a particular method. For an easy example, consider the maxim that scientific hypotheses should be consistent with the available evidence. If we assume that the data

are true (as the Bayesian does), then any method that produces a refuted answer obviously hasn't converged to the truth *yet* and it is possible to do better (Schulte 1999a, 1999b). To see how, suppose that a method produces an answer h inconsistent with current evidence e but eventually converges to the truth. Since the answer is inconsistent with true data e and the method converges to the truth, the method eventually converges to an answer other than h in each world compatible with e. Let n be the least stage by which the method converges to the true answer h' (distinct from h) in some world w compatible with e. Now construct a new method that returns h' in w from the end of e onward. This method converges to the truth immediately in w, but converges no more slowly in any other world, so in decision theoretic jargon one says that it *weakly dominates* the inconsistent method in convergence time or that the inconsistent method is *inadmissable* with respect to convergence time. Since science is concerned primarily with finding the truth, avoiding needless delays is a natural and direct motive for consistency.

The preceding argument does not take computability into account. In some problems, a computable method can maintain consistency at each stage only by timidly failing to venture substantive answers, so it fails to converge to the truth in some worlds (Kelly and Schulte 1995, Kelly 1996). In other words, computable methods may have to produce refuted theories if they are to converge to the truth. In that case, a committed truth-seeker could rationally side with convergence over consistency, so the Bayesian's blanket insistence on idealized consistency as a necessary condition for "rationality" is too strong.

Consider next the maxim that it is better to predict the data than to merely accommodate them. Recall the example whether the color will change no times, exactly once, or at least twice and suppose we have seen ten green observations. The only answer that predicts the next datum in light of past data is "the color never changes". Neither of the other answers is refuted, however, so why not choose one of them instead? Here is a reason based on efficiency: doing so would result in a needless retraction. For Nature can continue to present green inputs until, on pain of converging to the wrong answer, we cave in and conclude "the color never changes". Thereafter, Nature can exhibit one color change followed by constant experience until we revise to "the color changes exactly once" and can then present another color change

to make us revise again to "the color changes at least twice", for a total of three retractions. Had we favored "the color never changes" on constant experience, we could have succeeded with just two retractions in the worst case. Furthermore, after seeing a color change, we should prefer the answer "one color change", which is the only answer compatible with experience that entails the data until another color change occurs. To do otherwise would result in the possibility of two retractions from that point onward when one retraction should have sufficed in the worst case.

Consider the question whether "all ravens are black", and suppose that Nature is obligated to show us a black raven, eventually, if one exists. Then the most efficient possible method (in terms of retractions and convergence time) is to assume that the hypothesis is true until a counterexample is encountered and to conclude the contrary thereafter, since this method uses just one retraction and is not weakly dominated in convergence time by any other method. Now suppose one were to filter shoes out of the data stream and to reject "all ravens are black" as soon as a non-black raven is encountered. We would succeed just as soon and with no more retractions than if we were to look at the unfiltered data. If one were to filter out ravens, however, no possible method could converge to the truth, even in the limit. More generally, a kind of datum is *irrelevant* (for the purposes of efficient inquiry) if systematically filtering data of that kind does not adversely affect efficiency. Suppose we know in advance that all observed ravens are within one meter of a white paint can. Then by the time the sphere with that radius is filled with positive instances, "all ravens are black" is conclusively refuted and an efficient method must reject it immediately. No mystery there: different problems call for different solutions. Indeed, the performance viewpoint explains what background information "is for": extra background constraints tend to make an empirical problem easier to solve.

4. Ockham's Razor and Efficiency

In this section, we show how efficiency explains one of the great mysteries of scientific method better than Bayesian confirmation can. A quick survey of the major scientific revolutions (e.g., the Copernican, the Newtonian, the Lavoisierian, the Darwinian, etc.) reveals an unmistakable pattern, described already by William Whewell (1840). The received theory of some domain achieves broad, shallow coverage over a range of phenomena by positing a

large number of free parameters and then tweaking them until the various phenomena are accounted for. Then another, narrower, but more *unified* explanation is proposed that involves fewer parameters. The new theory appears implausible to those trained in the older tradition, but its ability to unify previously unrelated facts makes it ultimately irresistable.

Twenty years ago, one of us (Glymour 1980) proposed that the unified theory is better *confirmed* because it is cross-tested in more different ways than the disunified theory by the same data. This has a tough, Popperian ring: the simpler or more unified theory survives a more rigorous, self-inflicted, cross-testing ordeal. But a theory is not a long distance runner who needs training and character development to win— it just has to be *true*. Since reality might be disunified and complex (indeed, it *is* more complex than we used to suspect), how is the quest for *truth* furthered by presuming the true theory to be simple and severely cross-testable? If there is no clear answer to this question, then science starts to look like an extended exercise in sour grapes (if the world isn't the way I want it to be, I don't *care* what it is like) or in wishful thinking (I like simplicity, so the world must be simple).

Of course, Bayesians have no trouble *accommodating* simplicity biases: just assign greater prior probability to simpler hypotheses (e.g., Jeffreys 1985). But that approach evidently presupposes the very bias whose special status is to be explained. One doesn't have to favor the simple theory outright, however; one need only "leave the door open" to it (by assigning it nonzero prior probability) and it *still* wins against a strong a priori bias toward its complex competitor (cf. Rosencrantz 1983). For suppose that one merely assigns nonzero prior probabilitly to the simple, unified theory s, which entails evidence e without extra assumptions, so that $P(e|s) = 1$. The complex competitor $c = \exists\theta.q(\theta)$ has a free parameter θ to wiggle in order to account for future data. In the strongest possible version of the argument, there is a unique value θ_0 of the parameter such that $P(e|q(\theta_0)) = 1$ and at every other value of θ, $P(e|q(\theta)) = 0$. Finally, a "free parameter" isn't really free if we have sharp a priori ideas about the best way to set it, so suppose that $P(q(\theta)|c)$ is zero for each value of θ including θ_0. Given these assumptions, $P(e|c) = \int P(e|q(\theta))P(q(\theta)|c)d\theta = 0$. Hence,

$$\frac{P(c|e)}{P(s|e)} = \frac{P(c)}{P(s)}\frac{P(e|c)}{P(e|s)} = \frac{P(c)}{P(s)}\frac{0}{1} = 0.$$

So the simple theory trounces its complex competitor, as long as $P(s) > 0$. This accounts for

the temptation to say that it would be a *miracle* if the parameters of the complex theory were carefully adjusted by nature to reproduce the effects of s. If the hard-edged assumptions of the preceding argument are softened a bit, then the complex theory may end up victorious, but only if it is assigned a much greater prior probability than the simple theory.

This improved argument merely postpones the objectionable circularity of the first version, however. For focus not on the contest between c and s, but on the contest between $q(\theta_0)$ and s. Since both of these theories account for e equally well, there is no external or objective reason to prefer one to the other. In fact, the only reason s wins is because $P(s) > 0$ whereas $P(q(\theta_0)) = P(q(\theta_0)|c))P(c) + P(q(\theta_0)|s))P(s) = 0 \cdot P(c) + 0 \cdot P(s) = 0$. In other words, probabilistic "fairness" to s in the contest against c necessarily induces an infinite bias for s in the contest against $q(\theta_0)$. But one could just as well insist upon "fairness" in the contest between s and $q(\theta)$. Since we don't think $q(\theta)$ is more plausible a priori than $q(\theta')$, for any other value θ', it follows that $P(q(\theta)) = P(s) = 0$. Then $P(c) = 1 - P(s) = 1$, so the complex theory c wins a priori (because it covers so many more possibilities). The moral is that *neither* Bayesian prior is really "open minded"— we are being offered a fool's choice between two extreme biases. What open-mindedness really dictates in this case is to reject the Bayesian's forced choice between prior probabilities altogether; but then the Bayesian explanation of Ockham's razor evaporates, since it is grounded entirely in prior probability.

Here is an alternative, efficiency-based explanation that presupposes no prior bias for or against simplicity and that doesn't even mention prior probabilities. Recall the question whether the observed color will change zero times, once, or at least twice. Three different intuitions about simplicity lead to the same simplicity ranking over these answers. First, the hypothesis that the color never changes is more *uniform* than the hypotheses that allow for color changes. Second, the hypothesis that there are no color changes is the most *testable* since it is refutable in isolation, whereas the other answers are refutable only given extra auxiliary hypotheses (e.g., the hypothesis that the color changes exactly once is refutable only under the extra assumption that the color will change at least once). Third, the theory that there are no color changes has *no free parameters*. The theory that there is one color change has one free parameter (the time of the change). The theory that there are at least two has at least

14

two parameters (one for each change). So the constant color hypothesis seems to carry many of the intuitive marks of simplicity. Now suppose we prefer the needlessly complex theory that the color will change prior to seeing it do so. Nature can withhold all color changes until, on pain of converging to the wrong answer, our method outputs "no changes". Then Nature can exact two more retractions, for a total of three, when two would have sufficed had we always sided with the simplest hypothesis compatible with experience (once for the first color change and another for the second).

Easy as it is, this argument suggests a general, performance-based understanding of the role of simplicity in science that sheds new light on the philosophical stalemate over scientific realism. The anti-realist is right that Ockham's razor doesn't point at or *indicate* the truth, since the truth might be simple or complex, whereas Ockham's razor points at simplicity no matter what. But the realist is also right that simplicity is more than an arbitrary, subjective bias that is washed out, eventually, by future experience. It is something *in between*: a necessary condition for minimizing the number of surprises prior to convergence. So choosing the simplest answer compatible with experience is better justified (in terms of truth-finding efficacy) than choosing competing answers , but such justification provides *no security whatever* against multiple, horrible surprises in the future. The two theses are consistent, so the realism debate isn't a real debate. It is a situation. Our situation.

Suppose one were to ask whether the "grolor" changes no times, once, or at least twice, where the "grolor" of an observation is either "grue" or "bleen", where "grue" means "green prior to n_0 and blue thereafter" and "bleen" means "blue prior to n_0 and green thereafter" (Goodman 1983). The preceding argument now requires that one guess "no grolor change" until a grolor change occurs. But there is no grolor change if and only if there is a color change, so it seems that the whole approach is inconsistent. The right moral, however, is that simplicity is relative to the problem addressed. That is as it must be, for if simplicity is to facilitate inquiry over a wide range of problems, simplicity must somehow adapt itself to the contours of the particular problem addressed. There must be some general, structural concept of simplicity that yields distinct simplicity rankings in different problems.[2]

---

[2]Goodman's own response to this issue was that there is a special family of *projectible* predicates out of which confirmable generalizations may be formulated. That approach grounds simplicity and justification in

Here it is. Simple worlds are those in which Nature has opportunities to fool us but never exercises them. The least simple worlds are those in which all the opportunities are actually used up, so eventually no further inductive uncertainty remains. More precisely, say that a finite chunk of experience *verifies* an answer to a question relative to background information k just in case each world satisfying k that presents the same experience also satisfies the answer. Then a world has *simplicity degree* zero in a problem just in case it eventually presents experience that verifies some answer to the problem. A world has simplicity degree $n+1$ just in case it eventually presents experience that verifies some answer given the assumption that the world has simplicity at least $n+1$. For example, recall the problem in which one must say whether the color changes zero times, once or at least twice. In this problem, worlds in which the color changes twice verify the answer "at least two color changes", so they have simplicity degree zero. Worlds in which the color changes exactly once verify the answer "exactly once" *given* that the world has at least unit simplicity, so they have unit simplicity, and so forth.

The simplicity of an answer is defined as the maximum simplicity degree over all worlds satisfying the answer and the *structural complexity* of a problem can be defined as the supremum of the simplicities of the worlds in the problem (recall that simplicity is a matter of avoiding dirty tricks by Nature forever, so problems with extremely simple worlds afford lots of retractions, and hence are complex). Ockham's razor can now be stated in the obvious way: never output an answer unless it is the uniquely simplest answer compatible with current experience. The important point is this: it is a mathematical theorem that any violation of Ockham's razor implies either that one fails to decide the question at hand in the limit or that one uses more retractions than necessary in the subproblem entered when Ockham's razor is violated. Hence, *efficiency in each subproblem requires that one follow Ockham's razor at each stage of inquiry* (the proofs of the claims in this paragraph are in Kelly 2002). Furthermore, one can show that following Ockham's razor is *equivalent* to minimizing errors prior to convergence, assuming that the method converges to the truth in the limit and that

personal sentiment, which strikes us as wrong-headed. For us, sentiment is relevant only to the selection of problems. Justification then supervenes on objective efficiency with respect to the problems sentiment selects. Hence, our approach involves a middle term (problems) that "screens off" sentiment from justification, allowing us to give an objective proof of the truth-finding efficacy of Ockham's razor over a broad range of problems.

success under a (transfinite) error bound is possible at all. Finally, success under a transfinite error bound is possible in each problem (q, k) whose background presupposition k is itself decidable in the limit.

To see how the idea applies in a different context, consider an idealized version of the problem of inferring conservation laws in particle physics (Schulte 2000). The standard practice in this domain has been to infer the most restrictive conservation laws compatible with the current reactions (Ford 1963). Retraction efficiency demands this very practice, for suppose one were to propose looser conservation laws than necessary. Then Nature could withhold the unobserved reactions incompatible with the most restrictive laws until we give in (on pain of converging to the wrong laws) and propose the most restrictive laws. Thereafter, Nature can exhibit reactions excluded by these laws, forcing a retraction, and so forth for the remaining degrees of restrictiveness. Notice that tighter conservation laws are "simpler" in our general sense than are looser laws, for in worlds in which the tighter laws hold, Nature forever reserves her right to exhibit reactions violating these laws, but in worlds in which looser laws are true, eventually Nature has to reveal reactions refuting simpler laws, assuming that all the reactions are observable.

5. Statistical Retractions

There is something admittedly artificial about examples involving ravens and discrete color changes. Both the world and the measurements we perform on it are widely thought to involve chance, and where chance is involved, nothing is strictly verified or refuted: it is always *possible* for a fair coin to come up heads every time or for a measurement of weight to be far from the true value due to a chance conspiracy of disturbances. In this section, we extend the preceding efficiency concepts, for the first time, to properly statistical problems. Doing so illustrates clearly how the problem of induction arises in statistical problems and allows one to derive Ockham's razor from efficiency in statistical settings, with applications to curve fitting and causal inference. Readers who are willing to take our word for it are invited to skip to the next section, in which the applications are sketched.

In a statistical problem concerning just one continuous, stochastic measurement X, each possible statistical world w determines a *probability density* function p_w over possible values

of X. If we repeatedly sample values of X for n trials, we arrive at a sample sequence $(X_1 = x_1, \ldots, X_n = x_n)$ in which $X_i = x_i$ is the outcome of the ith trial. If the sampling process is independent and identically distributed, then samples are distributed according to the product density:

$$p_w^n(X_1 = x_1, \ldots, X_n = x_n) = p_w(X_1 = x_1) \cdot \ldots \cdot p_w(X_n = x_n).$$

Increasing sample size will serve as the statistical analogue of the notion of accumulating experience through time.

A *statistical question* partitions the possible statistical worlds into mutually incompatible potential answers, a *statistical presupposition* delimits the set of worlds under consideration and a *statistical problem* consists of a question paired with a presupposition. A *statistical method* is a rule that responds to an arbitrary sample of arbitrary size with some guess at the correct answer to the question or with '?', which indicates a refusal to commit at the current time. In a familiar, textbook example, the background presupposition is that the observed value of X is normally distributed with known variance σ^2 and unknown mean μ. Then each possible value of the mean μ determines the normal sampling density with mean μ and variance σ^2. The question might be whether $\mu = 0$. A method for this problem returns $\mu = 0$, $\mu \neq 0$ or '?' for an arbitrary sample of arbitrary size.

There is always some small probability that the sample will be highly unrepresentative, in which case the most sensible of statistical methods will produce spurious results. Hence, there is no way to guarantee that one's method actually converges to the truth. It is better to focus on how the probability of producing the right answer evolves as the sample size increases. Accordingly, say that M *solves* a problem *in the limit* (in probability) just in case in each world satisfying the problem's presupposition, the probability of producing the right answer for that world approaches unity as the sample size increases.

Statistical retractions occur when a method's chance of producing some answer drops from a high to a low value. Let $1 > \gamma > 0.5$. Method M γ-*retracts* h between stages n and n' in w just in case $P_w^n(M = h) > \gamma$ and $P_w^{n'}(M = h) < 1 - \gamma$. Then M γ-*retracts at least k times* in w iff there exist $n_0 < n_1 < \ldots < n_k$ such that M γ-retracts some answer to the question between n_0 and n_1, between n_1 and n_2, etc. Also, M γ-retracts exactly k times iff k is the greatest k'

such that M γ-retracts at least k' times in w. Moreover, M *solves* a given statistical problem *with at most k γ-retractions* just in case M solves the problem in the limit in probability and γ-retracts at most k times in each world. Finally, M *solves* a given statistical problem *with at most k γ-retractions starting with h* just in case M solves the problem with at most k retractions and in each world in which which M uses all k γ-retractions, h is the first answer produced by M with probability $> \gamma$.

The γ-*retraction complexity* of a statistical problem *starting with h* is the least γ-retraction bound under which some method can solve the problem starting with h. The γ-*retraction complexity* of a statistical problem is the least γ-retraction bound under which some method can solve it. As before, γ-*verifiability* is solvability with one γ-retraction starting with $\neg h$, γ-*refutability* is solvability with one γ-retraction starting with h and γ-*decidability* is solvability with zero γ-retractions.

To see how it all works, recall the textbook problem described earlier, in which observed variable X is known to be normally distributed with variance σ^2 and the question is whether or not h is true, where h says that the mean of X is zero. Let M_α^n be the standard statistical test of the point null hypothesis h at sample size n and signifance level α. In this test, one rejects h if the average of the sampled values of X deviates sufficiently from zero. The significance level of the test is just the probability of mistakenly rejecting h when h is true (i.e., when the true sampling distribution is p_μ). It won't do to hold the significance level fixed over increasing samples, for then the probability of producing h when h is true will not go to unity as n increases. That is readily corrected, however, by "tuning down" α according to a monotone schedule $\alpha(n)$ that decreases so slowly that the successive tests $M_{\alpha(n)}^n$ have ever-narrower acceptance zones.

It is a familiar fact that $M_{\alpha(n)}^n$ solves the preceding problem in the limit (in probability), but the current idea is to attend to γ-retractions as well, where $1 > \gamma > 0.5$. Suppose that the initial significance level is low— less than $1 - \gamma$. Then $M_{\alpha(n)}^n$ starts out producing h with high probability in w. Also, since the significance level drops monotonically to zero as the sample size increases, the probability of producing h rises monotonically to unity in w, so there are no γ-retractions of h. If w' satisfies $\neg h$, then since the sample mean's density peaks

monotonically around the true mean in w' and the acceptance zone shrinks monotonically around w, the probability that $M^n_{\alpha(n)}$ produces $\neg h$ approaches unity monotonically. If w' is very close to w, then the method may start out producing h with high probability in w', because p_w will be very similar to $p_{w'}$. But since the probability of producing $\neg h$ rises monotonically in w', h is γ-retracted just once. Far from w there are no γ-retractions at all, because h is never produced with high probability. So $M^n_{\alpha(n)}$ succeeds with one γ-retraction starting with h and, hence, *h is γ-refutable, for arbitrary γ such that $1 > \gamma > 0.5$.*

That doesn't suffice to justify the proposed method. We still have to argue that *no possible method can decide the problem* in the more desirable, two-sided sense. For this, it suffices to show that no possible method γ-verifies h in probability when $1 > \gamma > 0.5$. Suppose, for reductio, that M solves the problem with one retraction starting with $\neg h$. Suppose w satisfies h. Since M succeeds in the limit and $\gamma < 1$, there exists an n_0 in w such that $P^{n_0}_w(M = h) > \gamma$. There exists a small, open interval I around w such that for each world w' in I, $P^{n_0}_{w'}(M = h) > \gamma$.[3] Choose $w' \neq w$ in I. Then since M succeeds in the limit and w' does not satisfy h, there exists $n_1 > n_0$ such that $P^{n_1}_{w'}(M = \neg h) > \gamma$. Since w' is in I, we also have that $P^{n_0}_{w'}(M = h) > \gamma$, so at least one γ-retraction occurs in w'. By the reductio hypothesis, the first answer output with probability $> \gamma$ in w' is $\neg h$. So another retraction occurs by stage n_0, for a total of two retractions. This contradicts the reductio hypothesis and closes the proof. A corollary is that *no possible method solves the problem with zero γ-retractions if $1 > \gamma > 0.5$*, for any such method would count as a γ-verifier of h. Hence, h is γ-refutable but is not γ-verifiable or γ-decidable.

This asymmetry is not so surprising in light of the familiar, statistical admonition that rejections of tests are to be taken seriously whereas acceptances are not. Less familiar is the question whether statistical problems can require more than one retraction, so that *neither* side of the question is refutable (in probability). In fact, such problems are easy to construct. Suppose that we have two independent, normally distributed variables X and Y and we want to know which of the variables has zero mean (both, one or the other or neither). *This problem is solvable with two γ-retractions starting with "both zero", but is not solvable with*

---

[3] This follows from the Lebesgue convergence theorem (cf. Royden 1988, p. 267).

two γ-retractions starting with any other answer, as long as $1 > \gamma > 0.5$. The negative claim can be shown by the following extension of the preceding argument. Let h_S be the answer that exactly the variables in S have zero means, so we have possible answers $h_\emptyset, h_{\{X\}}, h_{\{Y\}}$ and $h_{\{X,Y\}}$. Suppose, for reductio, that M solves the problem with two γ-retractions starting with some answer other than $h_{\{X,Y\}}$. Each possible world corresponds to a possible value of the joint mean (x, y). Since M succeeds in the limit and $\gamma < 1$, there exists an n_0 such that $P_{(0,0)}^{n_0}(M = h_{\{X,Y\}}) > \gamma$. There exists a small open disk B_0 around world $(0,0)$ such that for each world w' in B_0, $P_{w'}^{n_0}(M = h_{\{X,Y\}}) > \gamma$.[4] Choose $(0, r)$ in B_0 so that $(0, r)$ satisfies $h_{\{X\}}$. Since M succeeds in the limit, there exists an $n_1 > n_0$ such that $P_{(0,r)}^{n_1}(M = h_{\{X\}}) > \gamma$. Again, there exists a small open disk B_1 around $(0, r)$ such that for each world w' in B_1, $P_{w'}^{n_1}(M = h_{\{X\}}) > \gamma$. Choose (r', r) in B_1 so that (r', r) satisfies $h_\emptyset$. Since M succeeds in the limit, there exists an $n_2 > n_1$ such that $P_{(r',r)}^{n_2}(M = h_\emptyset) > \gamma$. Since world (r', r) is in both B_0 and B_1, we also have that $P_{(r',r)}^{n_0}(M = h_{\{X,Y\}}) > \gamma$ and that $P_{(r',r)}^{n_1}(M = h_{\{X\}}) > \gamma$, for a total of at least two retractions in (r', r). By the reductio hypothesis, the first answer output with probability $> \gamma$ in w' is not $h_{\{X,Y\}}$. But then another retraction occurs by n_0, for a total of three. This contradicts the reductio hypothesis and closes the proof. It follows as an immediate corollary that *no possible method solves the problem with one γ-retraction if $1 > \gamma > 0.5$*, for any such method would count as succeeding with two retractions starting with an arbitrary answer different from $h_{\{X,Y\}}$.

So the best one can hope for is two retractions starting with the hypothesis that all the means are zero. The following, natural strategy does as well as possible. Choose the usual statistical tests for $\mu_X = 0$ and for $\mu_Y = 0$. Tune down the significance levels to make both tests γ-refute their respective hypotheses, as in the preceding example. Let M produce $h_\emptyset$ if both tests reject, $h_{\{Y\}}$ if only the X test rejects, $h_{\{X\}}$ if the Y test rejects and $h_{\{X,Y\}}$ if neither test rejects. The probability of producing the right answer rises monotonically toward unity in each test, as was described above. The probability that M produces the right answer is the product of the marginal probabilities that the component tests are right. In the worst case, the right answer is $h_\emptyset$ and the actual world (r, r') has the property that r is quite small

---

[4]Again, by the Lebesgue convergence theorem (Royden p. 267).

and r' is even smaller. In such a world, the probability that the X test rejects will rise late and the probability that the Y test rejects will rise later. Then at worst, there is a time at which both tests probably accept followed by a time at which the X test probably rejects and the Y test probably accepts followed by a time after which both tests probably reject. Since the joint probability is the product of the marginal probabilities, M γ-retracts at most twice. Furthermore, each worst-case world in which two retractions occur has $h_{\{X,Y\}}$ as the first output produced with probability $> \gamma$. Hence, this problem's complexity is exactly "two retractions starting with $h_{\{X,Y\}}$". It is clear that each new variable added to the problem would result in an extra retraction, so there is no limit to the number of retractions a statistical problem can require.

The multiple mean problem has suggestive features. In order to minimize retractions, one must start out with the hypothesis that both means are zero. This is the most uniform hypothesis (if a mean distinct from zero is close to zero, that fact will become apparent only at large sample sizes, resulting in a "break" in the signal from the environment as sample size increases). It is also the most testable hypothesis (the reader may verify that this answer is γ-refutable but none of the alternative answers is). Finally, $h_{\{X,Y\}}$ has no free parameters (both μ_X and μ_Y are fixed at zero) whereas $h_{\{X\}}$ allows adjustment of μ_Y and $h_\emptyset$ allows for adjustment of both both μ_X and μ_Y. These features suggest that retraction efficiency should explain intuitive simplicity preferences in more interesting statistical problems, such as curve fitting, model selection, and causal inference.

6. Curves and Causes

Suppose we know that the true law is of form

$$y = \alpha x^3 + \beta x^2 + \gamma x + \epsilon,$$

where ϵ is normally distributed measurement error and the question is whether the law is linear, quadratic, or cubic. Simplicity intuitions speak clearly in favor of linearity, but why should we agree? Minor variants of the preceding arguments show that the problem requires at least two retractions and requires more if the method starts with a non-linear answer. Moreover, any method that probably outputs a law of higher order than necessary in a world

that is simplest in some subproblem uses more retractions than necessary in the subproblem. We conjecture that the usual, nested sequence of tests (Jeffreys 1985) succeeds with two retractions starting with linearity.

Another sort of simplicity is minimal causal entanglement. The key idea behind the contemporary theory of causal inference is to axiomatize the appropriate connection between the true causal network and probability rather than to attempt to reduce the former to the latter (Spirtes et al. 2000, Pearl 2001). The principal axiom is the *causal Markov condition*, which states that each variable is probabilistically independent of its non-effects given its immediate causes. A more controversial assumption is *faithfulness*, which states that every conditional probabilistic independence follows from causal structure and the causal Markov condition alone (i.e., is not due to causal pathways that cancel one another out exactly). If all variables are observable and no common causes have been left out of consideration, it follows from the two axioms that there is a direct causal connection between two variables (one way or the other) just in case the two variables are statistically dependent conditional on each subset of the remaining variables.

The preceding principles relate the (unknown) causal truth to the (unknown) probabilistic truth. The methodological question is what to infer now, from a sample of the current size. Spirtes et al. have proposed the following method (which we now oversimplify— the actual method is much more efficient in terms of the number of tests performed). For each pair of variables X, Y, and for each subset of the remaining variables, perform a statistical test of independence of X and Y conditional on the subset. If every such test results in rejection of the null hypothesis of independence, add a direct causal link between X and Y (without specifying the direction). Otherwise, conclude provisionally that there is no direct causal connection. In other words, presume against a direct causal connection until rejections by tests verify that it should be added.

The proposed method is, again, a Boolean combination of standard statistical tests, because the edges in the output graph result from rejections by individual tests and missing edges correspond to acceptances. Since it uses familiar marginal tests, the procedure can be implemented on a laptop computer and it has been used with success in real problems.

23

However, it is neither Bayesian nor Neyman-Pearsonian: the significance levels and powers of the individual tests do not really pertain to the overall inference problem. For some years, the principal theoretical claim for the method has been that it solves the causal inference problem in the limit (in probability); a rather weak property. But now one can argue, as we have done above several times, that (a) the problem of inferring immediate causal connections requires as many probable retractions as there are possible edges in the graph and that (b) an extra retraction is required in the current subproblem if the method ever probably outputs a complex graph in one of the simplest worlds in the subproblem. Furthermore, we conjecture that the proposed method (or some near variant thereof) succeeds under the optimal retraction bound.

7. Confirmation Revisited

Bayesian methods assign numbers to answers instead of producing answers outright. This hedging is thought to be an especially appropriate attitude in the face of possible, nasty surprises in the future. It is, rather, a red herring, for there is a natural sense in which hedgers retract just as much and as painfully as methods that leap straight for the answers themselves. Say that Bayesian $P(.|.)$ *solves* a statistical problem just in case for each $\epsilon < 1$ and for each statistical world w satisfying the problem's presupposition, there exists a stage n such that for each stage $m \geq n$, $P_w^m(P(h_w|.) > \epsilon) > \epsilon$, where h_w is the correct answer in w to the statistical question posed by the problem. Let γ be strictly between 0.5 and unity and say that confirmation method $P(.|.)$ γ-*retracts* answer h between n and n' in w just in case $P_w^n(P(h|.) > \gamma) > \gamma$ and $P_w^{n'}(P(h|.) < 1 - \gamma) < 1 - \gamma$. Also, $P(.|.)$ *starts with* h (relative to k, γ) iff in each world in which k is realized, the first answer assigned more than γ probability by $P(.|.)$ with chance $> \gamma$ is h.

Given these concepts, one can re-run all the preceding arguments for exacting retractions from a statistical method, only now one forces the Bayesian's credence to drop by a large amount (from γ to $1 - \gamma$) with high chance (from γ to $1 - \gamma$). Indeed, you are invited to run your favorite choices of Bayesian prior probabilities through the negative arguments of the preceding section. Be as tricky as you like; assign point mass to simple answers or assign continuous priors. Either your agent fails to converge to the truth in probability in some

world or it realizes the worst case retraction bound in some world.

That doesn't mean Bayesian methods are *bad*, for the same arguments apply to *any* strategy for attaching partial credences to answers in light of samples or for inferring answers from samples. We are not like the classical statisticians who reject Bayesian methodology unless the prior corresponds to a known chance distribution. Nor are we like the idealistic extremists in the Bayesians camp, who call their arbitrary prior distributions "knowledge", even when nothing is known. We advocate the middle path of letting problems speak for themselves and of solving them as efficiently as possible by whatever means. Bayesian means may be as good as any others, but they are not and cannot be better than the best.

The moral for confirmation theory is that good Bayesian methods are *only* good methods. The probabilities they assign to hypotheses are just the current outputs of good methods which, at best, converge to the truth with the minimum of surprises. The same efficiency could be had by attaching the numbers in other ways or by dispensing with the numbers and producing theories outright, as scientists have always done until fairly recently. There is no special aptness about softening one's views in the face of uncertainty: Nature can wreak as much havoc on a high confirmation value as on outright acceptance of an answer. High confirmation provides no guarantee or partial guarantee of a smooth inductive future. There are just smooth problems, bumpy problems, methods that add extra bumps and methods that avoid all the avoidable ones.

Acknowledgements

Bibliography

Allais, M. (1953). "Le comportement de l'homme rationel devant le risque: Critiques des postulats et axiomes de l'école americaine". *Econometrika* 21: 503-546.

DeFinetti, B. (1937). "Foresight: its Logical Laws, its Subjective Sources". In *Studies in Subjective Probability*. H. Kyburg and H. Smokler, eds. New York: Wiley.

Ellsburg, D. (1961). "Risk, Ambiguity, and the Savage Axioms". *Quarterly Journal of Economics* 75: 643-669.

Freivalds, R. and Smith, C. (1993). "On the Role of Procrastination in Machine Learning". *Information and Computation*. 107: pp. 237-271.

Ford, K. (1963). *The World of Elementary Particles*, New York: Blaisdell.

Glymour, C. (1980). *Theory and Evidence*, Princeton: Princeton University Press.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. 4th ed. Cambridge: Harvard University Press.

Hempel, C. G. (1965). "Studies in the Logic of Confirmation". In *Aspects of Scientific Explanation*. New York: The Free Press, pp. 3-51.

Howson, C. and P. Urbach (1989). *Scientific Reasoning: The Bayesian Approach*. New York: Open Court.

Jain, S., Osherson, D., Royer, J., and Sharma, A. (1999) *Systems that Learn*. Second ed. Cambridge: M.I.T. Press.

James, W. (1948). "The Will to Believe". In *Essays in Pragmatism*. A. Castell, ed. New York: Collier.

Jeffreys, H. (1985). *Theory of Probability*. Third ed. Oxford: Clarendon Press.

Kelly, K. (1996). *The Logic of Reliable Inquiry*. New York: Oxford.

Kelly, K. (2000). "The Logic of Success", *British Journal for the Philosophy of Science* 51: 639-666.

Kelly, K. (2002). "A Close Shave with Realism: How Ockham's Razor Helps Us Find the Truth". CMU Philosophy Technical Report 137.

Kelly, K. and Schulte, O. (1995). "The Computable Testability of Theories with Uncomputable Predictions". *Erkenntnis* 42: pp. 29-66.

Kahneman, D. and A. Tversky (1972). "Subjective Probability: a Judgment of Representativeness". *Cognitive Psychology* 2: 430-454.

Kuhn, T. (1970). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuratowski, K. (1966). *Topology*, vol. 1. New York: Academic Press.

Kyburg, H. (1978). "Subjective Probability: Criticisms, Reflections and Problems". *Journal of Philosophical Logic* 7: pp. 157-180.

Laudan,L. (1980). "Why Abandon the Logic of Discovery?". In *Scientific Discovery, Logic, and Rationality.* T. Nickles, ed. Boston: D. Reidel.

Lee, P. (1989). *Bayesian Statistics: An Introduction.* London: Edward Arnold.

Levi, I. (1993). "Money Pumps and Diachronic Books". *Philosophy of Science.* Supplement to 69: s236-s237.

Maher, P. (1997). "Depragmatized Dutch Book Arguments". *Philosophy of Science* 64: 291-305.

Osherson, D. and Weinstein, S. (1988). "Mechanical Learners Pay a Price for Bayesianism". *Journal of Symbolic Logic* 53: 1245-1252.

Pearl, J. (2000). *Causation.* New York: Oxford University Press.

Popper, K. (1959). *The Logic of Scientific Discovery.* New York: Harper.

Putnam, H. (1965). *Trial and Error Predicates and a Solution to a Problem of Mostowski. Journal of Symbolic Logic* 30: 49-57.

Rosencrantz, R. (1983). "Why Glymour is a Bayesian". *Testing Scientific Theories*, John Earman ed., Minneapolis: *University of Minnesota Press.*

Royden, H. (1988). *Real Analysis.* Third edition. New York: MacMillan.

Savage, L. J. (1951). *The Foundations of Statistics*, New York: John Wiley & Sons.

Schulte, O. (1999a). "The Logic of Reliable and Efficient Inquiry", *The Journal of Philosophical Logic* 28: 399-438.

Schulte, O. (1999b). "Means-Ends Epistemology". *The British Journal for the Philosophy of Science* 51: 151-153.

Schulte, O. (2001). "Inferring Conservation Laws in Particle Physics: A Case Study in the Problem of Induction". *British Journal for the Philosophy of Science* 51: 771-806.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search.* Cambridge: M.I.T. Press.

Teller, P. (1973). "Conditionalization and Observation". *Synthese* 26: 218-258.

Whewell, W. (1840). *Philosophy of the Inductive Sciences* London: Parker.

Further Reading

Earman, J. (1992) *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge: M.I.T. Press.

Glymour, C. (1980). *Theory and Evidence*, Princeton: Princeton University Press.

Hempel, C. G. (1965). "Studies in the Logic of Confirmation". In *Aspects of Scientific Explanation.* New York: The Free Press, pp. 3-51.

Howson, C. and P. Urbach (1989). *Scientific Reasoning: The Bayesian Approach.* New York: Open Court.

Jain, S., Oshersion, D., Royer, J., and Sharma, A. (1999). *Systems that Learn.* Cambridge: M.I.T. Press.

Kelly, K. (1996). *The Logic of Reliable Inquiry.* New York: Oxford University Press.

Lee, P. (1989). *Bayesian Statistics: An Introduction.* London: Edward Arnold.

Martin, E. and Osherson, D. (1998) *Elements of Scientific Inquiry.* Cambridge: MIT Press.

Popper, K. (1959). *The Logic of Scientific Discovery.* New York: Harper.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search.* Cambridge: M.I.T. Press.

KEVIN T. KELLY

CLARK GLYMOUR