Argument, Inquiry, and the Unity of Science

Kevin T. Kelly
Department of Philosophy
Carnegie Mellon University
kk3n@andrew.cmu.edu

November 28, 2008

Abstract

Ockham's razor impels scientists to seek ever greater unity in nature. That seems to saddle science with a metaphysical presupposition of simplicity that might be false. The objection is apt if scientific method is understood as a system of inductive logic or proof, for then the unity of science must, somehow, function as an unjustified premise in scientific arguments. But if science is understood, instead, primarily as a process of discovery that aims at finding the truth as efficiently as possible, the unity of science can be understood as an optimally truth-conducive *heuristic* rather than as a metaphysical presupposition. Optimal truth conduciveness is what epistemic justification is for. Therefore, Ockham's razor is justified as a scientific heuristic even though it might be false.

1 Ockam's Razor and the Unity of Science

In the Preface to his *De Revolutionibus*, Nikolaus Copernicus did not cite any new or crucial experiments in favor of his heliocentric astronomical hypothesis: his argument was based squarely on what he called the *harmonies* of his system. What he had in mind is the fact that his theory is severely tested by data that Ptolemy's theory merely accommodates; for example, heliocentrism entails that planetary retrograde motion must happen either at solar conjunction or at solar opposition, whereas in Ptolemy's theory retrograde motion is entirely independent of solar position. Fresnel's argument for the wave theory of light centered on the ability of the theory to provide a *unified* explanation of diffraction bands around shadows and of the rings that appear when lenses are pressed together, which are qualitatively completely different. Each phenomenon allows one to derive the wave lengths of the various colors of light, which yields a sharp, testable prediction with respect to the other phenomena. Universal gravitation allowed Newton to estimate the gravitational constant from terrestrial pendula and then test the theory against the moon's orbit. Maxwell's electromagnetic equations unified magnetic and electrical forces. Darwin's theory of evolution uses common ancestry to explain homologies or similarities of structure across diverse environments.

And so on. The scientific preference for simple, harmonious, unified, severely testable explanations is now known as *Ockham's razor*.

Harmony, unity, and severe testability sound like wholesome *virtues*, but it is a further question whether such virtues count as reasons to believe in simple theories. Ockham's razor can result in error. Parmenides proposed that the universe is an immutable, perfectly symmetrical sphere. Aristotle set force proportional to the first derivative of distance rather than to the second. Johannes Kepler "unified" the diameters of planetary orbits by nesting the orbits in nested platonic solids. Particle physicists mistakenly assumed that reactions would conserve parity. Conspiracy theorists see diabolical common causes everywhere. In fact, nature is not perfectly simple and its complexities were not always so obvious as to be observable immediately (e.g., the violation of parity). To assume that reality is as simple as possible with respect to the data available can, therefore, result in error. So what entitles science to assume that the truth is simple? The simplicity of the universe would seem to be an empirical matter that should be investigated rather than pre-judged by science.

Every philosophical explanation starts with a simple, guiding paradigm or metaphor. One traditional, guiding metaphor for explaining scientific method is that science is a system of empirical or inductive *arguments*. Inductive arguments reflect a relation of support or partial entailment of theoretical hypotheses by the available data and other background assumptions. On that view, Ockham's razor must be understood either as a material premise or as a structural feature of inductive argumentation. But it is implausible to suppose that we know that hidden realities must be simple, so Ockham's razor does not appear to be a premise. Nor does the rhetorical force of Ockham's razor explain why simplicity *should* count toward empirical justification—is does not imply ought.

Alternatively, science can be viewed not as a system of arguments but as an ongoing process of inquiry directed at finding the truth. But how could any process systematically biased toward simplicity help one find possibly complex truths? Since Ockham's razor is a fixed bias, relying on its advice is like navigating a ship whose compass needle frozen into a fixed position by rust. Nonetheless, I will explain how Ockham's razor can be shown, in a sense, to be optimally truth conducive. The trick is that theory choice is an ongoing, essentially fallible procedure and, hence, cannot possibly be guaranteed to head *straight* for the truth, so the best Ockham's razor could do is to keep inquiry on the straightest possible path to the truth—much as directions to the freeway keep one on the best route home wherever home might happen to be. Therefore, the conception of science as an extended, fallible process of inquiry allows one to explain how Ockham's razor can be optimally truth conducive without assuming a priori that the universe is simple. The conception of science as a static system of arguments has not achieved a similar explanation and it is not clear how it could.

2 Simplicity and Proof

Philosophers of science tend to think of scientific justification in terms of *confirmation*. The idea is that confirmation is some sort of relation between theory and evidence that in some sense justifies or increases the justification of the theory. However you come up with a theory, at the end of the day it is the evidence reported in the final write-up rather than the Stürm und Drang of the scientist's individual psycho-history that matters (Hempel 1965). Kepler's neoplatonic fantasies and travails in deciphering the orbit of Mars and Kekule's dream of a snake biting its own tail prior to publishing the stere-structure of Benzene make fascinating and inspirational reading, but have nothing to do with the justification of the views for which these men are justly famous. Instead, it is the *bearing* of the available evidence that ultimately justifies a theory. Hence, a proper understanding of scientific justification reduces, in its cold fundamentals, to the nature of the relation of *confirmation* or *evidential bearing*, itself. And how does one study such a thing? As philosophers have done since Plato: by examining intuitive judgments of positive and negative evidential relevance to sketch out or *explicate* the underlying concept. The concept of confirmation, properly analyzed, distills the essence of scientific justification and, hence, screens off or renders irrelevant all of the other more vivid, procedural or, one might say, *tawdry* aspects of inquiry, like gathering data, generating hypothesis, testing them, formulating alternatives, and generally doing one's best to find the truth.

The plausibility of the preceding position is based squarely on a rough analogy between scientific justification and mathematical proof. Scientists and mathematicians both use proofs and arguments. Gottlob Frege, David Hilbert, and others seemed to explain mathematics as a system of theorems from formal axioms. It is tempting, therefore, to apply the same idea to natural science, substituting a fallible or "inductive" logic for the deductive logic of mathematics.

The logic of empirical science has been sought mainly in the theory of probability. One approach, due to John Maynard Keynes (1921) and Rudolf Carnap (1962), is to choose a particular conditional probability measure $p(T|E)$ whose value is to be interpreted as the degree of confirmation or partial entailment of theory T by evidential report E. The idea is that in a logically valid argument from premise E to conclusion T there can be no counterexample, whereas the strength of a partially valid or *inductive* argument depends on the *weight* or *proportion* of possible cases satisfying E that also satisfy T. Weight is interpreted as some probability measure $p(D)$ over possible state descriptions D of the universe as describable in the scientist's language. Then the conditional probability $p(T|E) = p(T\&E)/p(E)$ can be viewed as the degree to which premises E partially entail theory T, with full entailment implying $p(T|E) = 1$ and refutation implying $p(T|E) = 0$. The trick is to pick out some special choice of p. But which one? Carnap wanted to make sure that confirmation would allow for induction or learning from experience, which for him meant that more observations satisfying property Q confirm that the next observation will also satisfy Q:

$$p(Q \text{ tomorrow} \mid Q \text{ until today}) > p(Q \text{ today} \mid Q \text{ until yesterday}).$$

Traditionally, induction is traced to *uniformity of nature*, the view that the future will resemble the past. Uniform courses of nature are usually thought to be simpler than random or messy courses of nature, so uniformity of nature is an instance of Ockham's razor and, therefore, Carnap was really attempting to *explain* Ockham's razor in terms of confirmation or empirical argumentation.

An obvious idea is to assume that every possible state description of the universe carries equal probabilistic weight, but that idea makes successive observations probabilistically independent, so there is no induction in the sense just described. Instead, Carnap assigned equal probabilities to *structural* descriptions of the universe, where two states of the universe have the same structure if they are identical up to a permutation of individual names. Thus, if there is just one attribute, white vs. black, then each state description involving three objects corresponds to an assignment of colors to these objects:

$$\begin{aligned} p(\circ \circ \circ) &= 3/12 \\ p(\circ \circ \bullet) = p(\circ \bullet \circ) = p(\bullet \circ \circ) &= 1/12 \\ p(\circ \bullet \bullet) = p(\bullet \circ \bullet) = p(\bullet \bullet \circ) &= 1/12 \\ p(\bullet \bullet \bullet) &= 3/12. \end{aligned}$$

Carnap's proposal evidently imposes imposes a not-so-subtle prior probabilistic bias toward uniform or simple sequences of experience—a bias that becomes more pronounced as the number of individuals is increased. Given that a priori bias, it is hardly surprising that a run of black circles increases the probability of seeing another black circle.[1]

Carnap's explanation of induction has all the advantages of theft over honest toil. But even a thief needs to tell a consistent story, and Carnap's confirmation idea can't do so. Suppose that the predicate "blite" (Goodman 1995) is defined as black up to stage three and white thereafter and "whack" is defined as white up to stage three and black thereafter. To a speaker of the blite/whack language, the "uniform" worlds are "always blite" and "always whack", so the inductive bias becomes:

$$\begin{aligned} p(\circ \circ \bullet) &= 3/12 \\ p(\circ \circ \circ) = p(\circ \bullet \bullet) = p(\bullet \circ \bullet) &= 1/12 \\ p(\circ \bullet \circ) = p(\bullet \circ \circ) = p(\bullet \bullet \bullet) &= 1/12 \\ p(\bullet \bullet \circ) &= 3/12. \end{aligned}$$

Now learning from experience means expecting black at stage three after an seamless run of white. Or anything else, depending on the language and the bias adopted. The first response is that "blite" and "whack" are phoney, gerrymandered concepts

[1]The trick is that Carnap did not include time of appearance of an object as a predicate in the underlying language. If he did, then every ternary sequence would be in its own structural isomorphism class and once again there would be no learning! Carnap insisted that confirmation be assessed only in light of the total information available. Refusing to include time as a predicate is a tacit violation of that principle.

until one stops to reflect that to the blite/whack speaker white and black are equally gerrymandered, since they amount to "blite until stage three and whack thereafter" and "whack until stage three and white thereafter". So far as logical syntax is concerned (and that's all Carnap's logical approach had to fall back on) the situation is entirely symmetrical.

This radical dependence of scientific justification on linguistic formulation was not necessarily a bitter pill for Carnap, himself. Carnap had a penchant for progressive ideas and for solving the world's and philosophy's conflicts with language reform—he was an avid Esperanto instructor as a youth, for example and he developed his own shorthand for making notes—to the chagrin of contemporary archivists. In his philosophy, the laws of logic—deductive or inductive—are necessarily true not because they reflect some deep structure in the world but because they are an empty, conventional game; so it was a foregone conclusion to him that the concept of confirmation or partial entailment should have such a conventional element as well. The irony of grounding Ockham's razor on the choice of a free parameter in his philosophy seems to have been lost on Carnap.

But then it must be conceded that the reasons for using one logic as opposed to another are *pragmatic* or external to logic, itself. In an Orwellian terminological maneuver, Carnap included the truth-conduciveness of inquiry guided by such a logic among the merely "pragmatic" considerations, reserving the term "logic" for empty conventional considerations irrelevant to the fundamental aim of science. That amounts to a concession that Ockham's razor has no logical explanation.

Sometimes the best strategy is to fall back and to rally on more secure terrain. Frank Ramsey (1990), who discussed similar ideas with Keynes at Cambridge, proposed that there is no need to justify choice of a particular p because $p(A)$ is nothing other than some individual's degree of belief or willingness to pay for a bet that A is true. Whereas Carnap and Keynes viewed p as part of the structure of scientific justification, itself, Ramsey's view, now called *subjective Bayesianism*, was that p is more like a material *premise* reflecting the background beliefs and hunches of the individual. Then the *logic* of scientific justification, or of empirical *rationality* more broadly, consists only of the rules of probability, themselves, together with the familiar updating rule:

$$p_E(T) = p(T|E) = \frac{p(T\&E)}{p(E)},$$

which says that one's degree of belief in theory T upon learning that E should be one's prior degree of belief in T conditioned on E.

The tactical retreat may seem like an utter rout, for science is no longer portrayed as the chaste operation of pure logic, but as the result of combining arbitrary, prior biases with data to arrive at new biases. Of course, advocates of personalism were keenly aware of that objection from the outset and, to some extent, it can be met (Howson and Urbach 1989). It follows immediately from the above definition that:

$$p(T|E) = \frac{p(E|T)P(T)}{p(E)},$$

a fact known as *Bayes' theorem.* Then $p(T)$ and $p(E)$ are called the *prior* probabilities of the theory and the evidence, respectively, and $p(E|T)$ is called the *likelihood* of the theory given E. The likelihood $p(E|T)$ corresponds, roughly, to the extent to which T explains or at least predicts E. The idea is that the theory T says what $p(E|T)$ is, so the only subjective elements remaining are $p(T)$ and $p(E)$. The prior probability $p(T)$ corresponds to the initial plausibility of T, and who can deny that initial plausibility has something to do with scientific justification? The prior probability $p(E)$ corresponds to how surprising or unexpected the evidence is. There are celebrated cases of strange predictions vindicated by nature (e.g., Fresnel's prediction of a bright spot in the center of a circular shadow) and it does seem that these have particular confirmatory value. After the connection of the formula to scientific method is clear, it does explain quite a bit of scientific method. Implausibility is bad for then $p(T) \approx 0$. Refutation is bad because $p(E|T) = 0$. Surprising predictions are good because then $p(E|T) = 1$ and $p(E) \approx 0$. Tediously repeated experiments have diminishing returns since then $p(E) \approx 1$. Those are just the sorts of vague maxims that come to mind when one thinks of scientific method and now they can be traced to the firm ground of probability theory and the updating rule, rather than to some unmotivated, particular choice of p. Also, conceding that p enters into science as a premise rather than as part of the logic, itself, allows one to integrate background knowledge into science in a smooth and motivated way. In Carnap's setup, one always faces the question whether one is talking about personal hunches or logical support by the data and it is not even clear what that distinction means in particular cases (Quine 1951).

For those reasons and others, personal Bayesianism has caught on like wildfire in philosophy, statistics, machine learning, economics, social science, advertising, insurance, management and, increasingly, in the natural sciences, themselves (Lee 1989). It is an example of a philosophical view about science that has become part of the very fabric of science.[2]

What about Ockham's razor? Recall that Carnap's account depended on a particular choice of p, which, on the Bayesian scheme, seems to make it an explicit, material *premise* rather than a structural feature of the logic of science, itself. But at a deeper level Bayesians do have something to say about simplicity. Consider dependent variable Y and independent variable X and consider the orders of complexity of polynomial relationships between them: zero, constant, linear-but-not-constant, quadratic-but-not-linear, etc. Think of all the polynomial equations as being nested together into one infinitely long equation:

$$Y = \epsilon + a_1 + a_1 X^1 + a_2 X^2 + \ldots,$$

where ϵ is a random variable that injects noise so that the plot of dependent variable Y is not exact for a given value of the independent variable X. Let hypothesis H_n say that exactly $a_1, \ldots, a_n$ are non-zero and assume that some hypothesis H_n is true. If

[2]Of course, that makes it rather self-serving to argue for Bayesianism on the basis of "case studies" that reveal Bayesian behavior—one ought at least to present alternative philosophies of induction (e.g., mine) and then take a survey.

the observations were perfect, then one observation would suffice to rule out H_0, two would suffice to rule out H_1, etc. (Popper 1959). But science isn't that easy because the random error term ϵ spreads out the observations, making the arbitrarily small bumps in an arbitrarily complex polynomial curve arbitrarily hard to discern at a given sample size.

The extra bumps allowed by H_i as i increases are paradigmatic of what scientists understand empirical complexity to be. Each bump in is governed by a *free parameter* a_i that can be twiddled to *accommodate* the data. Hence, simpler theories *explain* simple data better or less arbitrarily so they are *tested more severely* by the data. These are not explanations of Ockham's razor, however (contrary to what many seem to think)—they are just trivial restatements of it.

The Bayesian explanation of Ockham's razor now goes like this. Suppose that the data nearly all fall on a flat line around zero. That is more or less what H_0 or $Y = 0$ would predict, so $p(E \mid H_0)$ is not so small. Now consider H_1, which says that $Y = a_0$ for *some* $a_0 > 0$. If a_0 is tuned toward 0, $p(E \mid Y = a_0)$ converges to $p(E \mid Y = 0)$, which is the Bayesian way to say that the more complex theory H_1 has a *free parameter* a_1 that can be *tuned* to *accommodate* the data. But H_1 says only that there exists *some* value of $a_1 > 0$ such that $Y = a_1$, so the likelihood of sample E given H_2 is the weighted average:

$$p(E \mid H_1) = \int_z p(E \mid Y = z) \cdot p(Y = z \mid H_1)\ dz.$$

Since $p(E \mid Y = a)$ is at best a little smaller than $p(E \mid Y = 0)$ when $a \approx 0$ and drops off to zero elsewhere, the weighted average is far smaller than $p(E \mid H_0) = p(E \mid Y = 0)$, so $p(E \mid H_0) \gg p(E \mid H_1)$. Hence, by Bayes' theorem, the sharp explanation by the simple theory swamps almost any prior hunch in favor of the complex theory. It seems, then, that even Ockham's razor is enforced by the structural part of probability theory, so the retrenchment to defensible ground was a resounding success.

Not exactly. Suppose that there are two theories, that a marble in a box is blue and that the marble is not blue. Indifference there results in degree of belief 1/2 in both theories. Further indifference among the ways of being non-blue results in probabilities far less than 1/2 for each of these colors (assuming sufficiently fine distinctions in color). So if the question is to guess the color of the marble, whatever it is, your ignorance now favors blue. Alternatively, you could be indifferent about color, in which case you would be strongly biased against blue in the question of blue vs. non-blue. The familiar moral is that that it is impossible to avoid strong prior biases in all questions even if one is confessedly completely ignorant with respect to all of the questions—modeling ignorance as flat probability forces everyone to take an a priori stand on some questions. The Bayesian "explanation" of Ockham's razor is just a case in point: ignorance over the possible answers to the question of polynomial *degree* implies that the prior probability of the simple *curve* stands in an infinite ratio with the prior probability of each complex curve and that infinite prior bias is passed through Bayes' theorem. Alternatively, a flat prior probability over curves implies an infinite prior bias *against* simple polynomial degrees—i.e., a strongly anti-Ockham bias.

The biases under discussion are not really knowledge or information, as habitually loose Bayesian talk about "making use of one's available information" might suggest—they are in place before data collection even starts. But neither can such strong biases be considered genuine ignorance, which ought not to be biased. Better to refer to such unavoidable Bayesian biases as *ignoredge*, to pre-empt the irresistable but epistemologically fatal slide from "ignorance" to "knowledge". The underlying metaphor that p is a premise in an argument abets the illusion of getting something for nothing here. In real logic, you are free not to have a premise and then you really aren't biased at all. In the Bayesian logic of "rationality", on the other hand, starting with a prior probability p is *mandatory* and each choice of p implies some material bias, so genuine ignorance is impossible.

But what is pure rationality, after all, other than a quasi-sociological model of the reasoning of the people whose reasoning we find most compelling—i.e., *ours*? For example, Leonard Savage (1954) held that the only ultimate defense of his formal axiomatization of Bayesian theory is that when violations of the axioms are pointed out, people will take steps to bring themselves into conformity. But, as a matter of fact, most of us vote with our feet when Bayesian rationality forces us to replace our genuine, honest, ignorance with Bayesian ignoredge. The following example is due to Daniel Ellsberg (1961). There is an urn filled with red, yellow, and blue balls, all thoroughly mixed. The known proportion of balls is 30% red and 60% yellow or blue and the proportion of yellow and blue balls is unknown. You are offered a hundred dollar bet on red vs. black and a hundred dollar bet on red-or-yellow vs. blue-or-yellow. Most people side with red in the first bet and with blue-or-yellow in the second, which contradicts Bayesian decision theory.[3] Assuming that the agent's ignorance is really ignoredge, favoring red in the first bet implies $p(R) > p(B)$. Favoring blue-or-yellow in the second implies $p(B) + p(Y) > p(R) + p(Y)$ and, hence, $p(B) > p(R)$, contradicting $p(R) > p(B)$. But if you are just plain ignorant about the relative frequency of blue vs. yellow, $p(B)$ and $p(Y)$ should not carry precise values and there is no contradiction.

Bayesianism can be liberalized in various ways to allow for genuine ignorance — e.g., by representing the agent's mental state by a *set* of possible probability measures (Kadane et al. 1999). That approach is more in keeping with the original metaphor of scientific justification as an argument, for now the trivial set containing all possible probability measures corresponds to the empty set of premises. But if both the Ockham and the counter-Ockham biases are added to an agent's set of distributions, as our true ignorance would seem to require prior to the onset of inquiry, the Bayesian argument for Ockham's razor evaporates.

3 Simplicity and Inquiry

There is an alternative, more active and dynamic metaphor for scientific justification: that the aim of inquiry is to *discover* or to *learn* the truth and that scientific justification is a matter of *truth-conduciveness*, or pursuing the truth as effectively as possible.

[3] Note that the second pair of bets results from the first pair of bets by replacing prize 0 in case "blue" with prize 100 in case "blue".

Plato, for example, viewed philosophy as a mystical path to ultimate truth. His central metaphors for philosophy were, accordingly, learning, in the dialogue *Meno* (2007). The learning viewpoint was expounded by the American Pragmatists C. S. Peirce (1878) and W. James (1897) and, in Carnap's time, by the philosopher/physicist H. Reichenbach (1949). Classical (non-Bayesian) statistics is filled with *procedures* for testing and estimation rather than with proofs. At my university, there is now an academic department devoted entirely to machine learning (Mitchell 1997). Computational learning theory (Jain et al., 1999) is entirely devoted to the truth-finding effectiveness of computable scientific methods. Machine learning and data-mining are now seeing their way back into statistics, itself (Wasserman 2004) and philosophy (Kelly 1996).

As fundamental metaphors for scientific justification, proof and discovery are quite different and, perhaps, even opposite (Kelly 1987, 2000, 2004). On the preceding, logical metaphor, confirmation is a merely formal or conventional relation that screens off the relevance of all process and procedure from scientific justification. The discovery picture is just the opposite: what matters to justification is how well science *works* as an overall learning machine and formal relations like confirmation and rationality are nothing but gears and subroutines embedded in the overall machine that are justified only to the subtle extent that they contribute to the truth-conduciveness of the overall process. The vaunted "published research report" that allegedly screens off the relevance of all procedural considerations is just a token in the overall, ongoing, social machine of inquiry. Even the social rules for assigning credit and blame to individual scientists can be viewed as social programming that furthers or hinders the overall truth-conduciveness of science as a collective enterprise (Kitcher 1993). One of the things that keeps philosophy fresh and exciting is the way truisms reverse when fundamental metaphors are shifted.

Truth-conduciveness comes in various strengths or grades. At the strong end of the spectrum is *certifiable reliability*: the requirement that a scientific method produce answers in a manner that entails a low, a priori bound on the objective chance of *error*, or production of a false answer. This is the property guaranteed by a statistical test. The size or *significance level* α of a test is an a priori bound on the chance of rejecting the hypothesis under test if it is true (called, unrevealingly, error of *type I*). Suppose that we conclude that the null hypothesis is false if the test rejects it but follow the advice in the statistics textbooks and conclude nothing at all if the test fails to reject. Refusal to produce a conclusion cannot result in error and the chosen significance level ensures a low chance of concluding in error that the null hypothesis is false, so statistical testing is certifiably reliable—with α as the a priori bound on chance of error. Taking acceptance of the null hypothesis seriously would not be certifiably reliable in most applications because possibilities quite close to the null hypothesis would produce samples almost indistinguishable from samples produced if the null hypothesis were true. So suspension of judgment when the null hypothesis is accepted is the price paid for certifiable reliability.

Another standard, statistical method is to construct confidence intervals. A confidence interval is not really an interval, but a method that *generates* an interval around a theoretical quantity when provided with a sample. Confidence intervals have significance level α or *confidence level* $1 - \alpha$ just in case the chance is less than α that a sample will be received that results in an interval that fails to catch the true value of the parameter in question. Thus, confidence intervals with significance level α have an a priori bound α on chance of error (failure to include the true value of the parameter in question), so they are certifiably reliable. Again, certifiable reliability has a cost. As the bound α on chance of error is decreased, the intervals must get larger and, hence, less informative. As the sample size increases the intervals get smaller, but never collapse down completely to a unique parameter value.

Could Ockham's razor be certifiably reliable? Recall the polynomial degree problem discussed above. The simplest hypothesis H_0 implies that every parameter a_i is exactly zero, so it is a point in the infinite-dimensional space of possible parameter settings. Ockham's razor is a systematic bias toward choosing the simpler theory, so there should at least be a sufficiently large sample at which one's chance of choosing the simplest theory H_0 rises above α if H_0 is true. But even if the true polynomial has a hundred bumps, each bump could be tuned to be so flat that the data would appear indistinguishable from data produced by $Y = 0$ at the current sample size and then it would be at least reasonably probable that Ockham's razor still sides with H_0. Therefore, Ockham's razor cannot be certifiably reliable when the question is one of choosing the true theory.

Statistical tests are certifiably reliable if they suspend judgment when the null hypothesis is not rejected, but in that case they do not converge to the true answer when the true answer is the null hypothesis. If the null hypothesis *is* accepted (Jeffreys 1931, Spirtes et al. 2000), then the test is no longer certifiably reliable—the chance of mistakenly accepting the null hypothesis is arbitrarily high if the null hypothesis is not too far from the truth. Confidence intervals can't help but be certifiably reliable, but every confidence interval that includes parameter values at which H_0 is true also catches parameter values satisfying arbitrarily complex polynomial equations, so certifiably reliable confidence intervals never rule out complex theories—i.e., they never make the "inductive leap" required to arrive at the true theory. For these reasons, it is appropriate to say that classical statistical inference is not really *inductive*. The demand for certifiable reliability in classical statistics is, therefore, tantamount to *inductive skepticism*.

Due to its skeptical underpinnings, when classical statistics finally confronts the problem of theory choice, it does so under the rather pallid rubric of "model selection". The idea is to give up on finding true theories—as in money-laundering, it's better not even to ask which back-alleys they come from. In the staid halls of classical statistics, theories are only chosen or adopted as non-literal, predictive *instruments* (Akaike 1973, Forster and Sober 1994). It might seem that the best-predicting theory should be the true one, but not when observations are noisy and theories have free parameters. To see why, consider the following parable. The military has excellent sharpshooters, but no human is perfect. Researchers at Maliflirtin Corp. decide to address this critical problem by designing a device that increases overall accuracy. Maliflirtin assesses the accuracy of a marksman by measuring the distance of each bullet hole from the bull's eye, squaring them, and adding the results (the squaring keeps them all positive so there is no cancelation of errors). The design goal is to improve this score by means of advanced technology. The result is the Marksman-clamp. The Marksman-clamp prevents the marksman's rifle from moving at all in the horizontal dimension, canceling the marksman's jitters in that dimension entirely. Unfortunately, this is government work, so the Markman-clamp is typically installed crooked so that the marksman can't possibly hit the bull's eye. Not to worry—as long as the marksman's clamp doesn't force too large a miss, the reduction in width of the cloud of hits compensates for not being able to hit the mark at all. To translate this parable to model selection: the untrammeled marksman corresponds to an unconstrained empirical estimate of a theoretical quantity. The Marksman-clamp corresponds to estimating the quantity with a simple but false theory. The point is that overall mean squared error can be reduced by estimating a parameter using an over-simplified, false theory. Hence, using a theory to improve predictive accuracy is not the same as finding the true theory and the classical conception of model selection is a natural extension of the classical statistician's systematic commitment to inductive skepticism.

If Ockham's razor is to be truth-conducive, one must give up either on reliability or on certifiability. The second possibility is pursued in philosophy by *naturalistic epistemologists*. Naturalistic epistemology is, indeed, natural in applications like perceptual knowledge. The basic idea is that you have perceptual knowledge that a vase is present when your perceptual belief in vases tracks (Nozick 2006), indicates, is correlated with, is caused by (Goldman 1967), or is informative about (Dretske 1981) the presence of vases. It also works with direction signs, compasses, and meters, all of which probably point in the right direction even though the user may not have any idea why or how or how accurate they are. Dropping certifiability staves off skeptical doubts, since one can have knowledge with no idea why one has it or how to get more—a hidden force or channel of information suffices.

But how does the idea work for Ockham's razor and induction? Suppose that *actually*, the truth is the simplest theory H_0. Then the actual chance of error is low, since H_0 is the theory probably chosen by Ockham's razor, if H_0 is true, making Ockham's razor uncertifiably reliable. That is not very impressive, however, since a blind dogmatist who accepts H_0 without looking at the sample at all would (actually) be even *more* reliable by this argument! If reliability is to be more than mere luck—the core motive of reliabilist epistemology— it must, somehow, span alternative statistical possibilities. But reliability over all the statistical possibilities would be certifiable, which is incompatible with Ockham's razor, as has already been discussed.

That leaves an intriguing possibility: that some unknown causes or chances somehow correlate simplicity with truth *across* possible statistical worlds. G. Leibniz, the celebrated co-founder of the calculus, proposed something of the sort—that the Deity is an engineer who loves simplicity and who, therefore, judges the best possible world to be the simplest or most elegant (1714). To make the idea more explicit, suppose that the Deity is a sporting Fellow, who creates the universe according to a *chance* sampling distribution over possible universes—a distribution that looks, coincidentally, just like the standard, Bayesian prior bias toward simplicity. Then simplicity would be an objective indicator of truth—like a metaphysical compass needle—even though we cannot certify its significance level a priori. Indeed, all one needs are the Bayesian likelihoods for each theory. For if the Bayesisan likelihoods were objective chances, it would be the case that if the truth weren't as simple as we think, it probably wouldn't be complex in a way so contrived as to have fooled us on a sufficiently large sample, as those parameter settings are only a very small range of the possible parameter settings.[4] So then Ockham's razor would track the truth, not in terms of the overt sampling chances discussed posited by the theories themselves (that would imply certifiable reliability) but in virtue of the Deity's hidden, metaphysical Urn chances. Hope beyond hope, if something like that is true, then even if we never figure out whether or how or why, Ockham's razor is busy producing non-accidental knowledge as we speak and there is a point, after all, to scientific inquiry!

Hidden causes and chances may explain Ockham's razor in sufficiently low-level inquiries, in which the free parameters in the laws can reasonably be viewed as random variables in their own right (e.g., infection rate in an epidemiological model). But nowhere is Ockham's razor more indispensable than at the level of *fundamental* theory and then the naturalistic turn becomes problematic, as R. Koons (2000) has argued. Suppose that the theories under consideration are fundamental, so that their parameters govern all the natural chances in the universe. Then natural chance is a function of the parameters, so the chance, given the actual setting of the parameters, that they are set as they are must be unity, if it even makes sense to speak of such a chance:

$$p_{\theta=r}(\theta = r) = 1.$$

Hence, there is no possible universe in which the natural chance of the fundamental parameters being set in a given way is flat or even continuous, as in the Bayesian argument for Ockham's razor.

Koons concludes that the truth-conduciveness of fundamental science is incompatible with naturalism—the view that all causes and chances are natural causes and chances "in" the natural universe, leaving the reader to carry the argument to its Leibnizian conclusion. My diagnosis is different. Short-run reliability is the wrong explication of truth-conduciveness for understanding theory choice, induction, and Ockham's razor. Insistence upon short-run reliability results either in inductive skepticism or in a pseudo-scientific, hidden, causal super-structure for science that would make Ockham blush. Better to explicate scientific truth-conduciveness in a weaker but certifiable manner that avoids both skepticism and shamanism.

A traditional alternative to reliability is certifiable *convergence*. Thus, although Ockham's razor does not certifiably point at or indicate the right answer immediately, eventually all the false, overly-simple answers are refuted and one is left with the true answer. This argument works entirely without occult, Ouija-board forces—it is no more occult than shooting a row of tin cans off of a fence (Sklar 1977). Even the Bayesians, who put their faith in a formal theory of rationality into which one may insert arbitrary, personal hunches, invoke convergence theorems to maintain some

[4] For example, Roush's (2005) version of tracking uses such likelihoods and, therefore, yields at least a subjective explanation of Ockham's razor.

sort of connection with reality. Unlike certifiable reliability, certifiable convergence is sufficiently weak to apply to the problem of theory choice—a point often repeated in the literature on statistical model selection.

But if reliability is too strict to single out Ockham's razor, convergence is too weak (Salmon 1967). For any theory whatever could be brought to the front of the queue and entertained until the complex effects it predicts (e.g., bumps in the polynomial case) have failed to appear by a given time fixed in advance. This procedure *might* find the truth faster than Ockham's razor (e.g, if the favored theory is true and the effects it predicts happen to be sufficiently large). Indeed, it is part of Bayesian lore that arbitrary prior opinions are less of a problem than one might suppose because they *wash out* in the face of empirical evidence (Savage 1954), so that divergent starting points come to agreement. But then Ockham's razor is more like a *defeasible hindrance* in finding the truth than a *positive guide* to the truth—and that much could be said of a prior bias toward complexity. So, again, certifiable convergence fails to explain how a prior bias toward simplicity helps one find the truth better than other prior biases would.

4 The Freeway to the Truth

The crux of the simplicity puzzle is to explain, without circularity, how a prior bias toward simplicity is better for finding the truth than a prior bias toward complexity would be. Mere convergence in the limit does not single out simplicity as the right bias to start out with and simplicity points toward the truth in the short run only under the (circular) prior assumption that the truth is simple. Indeed, it seems that any explanation of Ockham's razor *must* be circular, for how could a prior bias help one find the truth *regardless* how complex the truth happens to be? That sounds as paradoxical as giving good directions for arriving at an unspecified destination: wherever one points, the destination could lie in the opposite direction.

But not only can such advice be given; it happens every day. If you are lost in a small town and ask for directions, any local resident who recognizes that you live elsewhere will send you to the freeway entrance ramp. It isn't necessary for the resident to know which major city you are headed to, since the best route to any major city is on the freeway system. The hardest case for Ockham's razor is the one in which the entrance ramp lies in the opposite direction from your ultimate destination. But it is *still*, in a sense, better to follow the local directions to the freeway, which is straighter and more direct overall, even though the entrance ramp is in the opposite direction from your ultimate destination. For suppose you disregard the advice and, as luck would have it, head on a compass course directly toward your destination. You either end up taking a much more circuitous route home over rural roads or you eventually turn around and head toward the entrance ramp, adding at least one gratuitous course reversal to your overall route. In this parable, the resident puts you on the most direct route to your goal without knowing where you were headed. He does it without occult, mind-reading signals or circular assumptions. Indeed, knowing where you were headed wouldn't have improved his advice. Perhaps, in some similar manner, Ockham's razor keeps one on the most direct route to the truth, even though that route is not without some unavoidable twists and bumps of its own.[5]

Convergence to the truth implies only that one's chance of producing the truth climbs ever closer to 1. That is compatible with arbitrarily many arbitrarily large vacillations or retractions of earlier answers along the way and no possible convergent method is guaranteed to avoid arbitrarily many such retractions. More precisely, say that a method retracts answer H (in chance) to extent r at sample size $n+1$ just in case the chance that the method produces H drops by r from sample size n to sample size $n+1$. The total retractions in chance can be obtained by summing over all sample sizes and all possible answers.[6] Retractions are a healthy but painful sign of scientific progress, so there is no question of eliminating them altogether. But surely, it would be better not to reverse the course of science needlessly.

Although retractions in chance are rarely discussed as such, they are a familiar feature of statistical inference. For example, consider the policy of choosing the simplest polynomial hypothesis H_0 if it is not rejected by a test at low significance α and choosing the complementary hypothesis $\neg H_0$ otherwise. Then the chance of accepting H_0 is also high when the truth is the flat line $Y = a_0$, for a_0 slightly greater than zero. But at a larger sample size the disparity between a_0 and zero becomes noticeable and the chance of selecting $Y = 0$ drops to zero while the chance of accepting the true, complex answer $Y = a_1$ rises to unity. This reversal of opinion is due to the increasing power of the test as sample size increases, a generally recognized and unavoidable feature of of statistical testing.

Suppose you were to "fish" for the true hypothesis H_n by performing tests of the "nested" hypotheses:

$$(H_0), (H_0 \text{ or } H_1), (H_0 \text{ or } H_1 \text{ or } H_2), \ldots,$$

and choosing H_i when the first test not to reject is (H_0 or $\ldots$ or H_i). Then by the preceding logic, it is possible that the chance of choosing H_0 approaches unity and then drops toward zero, followed by the chance of choosing H_1 rising to unity and dropping to zero, followed by $\ldots$, until the chance of choosing H_n rises toward unity. So the theory fishing expedition retracts its conclusions n times when the truth is H_n. It may seem that the same result would be achieved regardless of the ordering—the thing you seek is always in the last place you look. But here is the interesting part: *every* method that converges in probability to the truth can be forced by nature to produce H_0 with arbitrarily high chance followed by H_1 with arbitrarily high chance

[5]The idea of paying attention to retractions was proposed by H. Putnam (1965), a student of H. Reichenbach. Since then, the concept has been extensively studied by computational learning theorists as a concept of problem complexity (Jain et al. 1999). O. Schulte (1999, 2000) and I proposed retraction minimization as a way to recover short-run constraints on scientific method and my own recent work (2002, 2007, 2007a, 2007b, 2008, Kelly and Glymour 2004) has applied that approach to the derivation of Ockham's razor.

[6]It is not hard to show that total retractions in chance are a lower bound on the expected number of total retractions.

followed by H_2 with arbitrarily high chance, etc.—the sequence corresponding to the intuitive ranking by simplicity. For let ϵ be non-zero but arbitrarily small. On data presented by $X = 0$, method M always produces H_0 with chance greater than $1 - \epsilon$ after the sample reaches some size n_0. Since sampling chance depends continuously on the form of the underlying equation, there exists a sufficiently small constant $a_0 > 0$ such that M has chance greater than $1 - \epsilon$ of producing H_0 even though $Y = a_0$ is true. But since M converges to the truth, there is a larger sample size n_1 after which the chance that M produces H_1 exceeds $1 - \epsilon$. So M first probably produces H_0 and then probably produces H_1. Note that ϵ can be chosen arbitrarily small, so the total retractions can be driven arbitrarily close to 1. This argument can be iterated to force retractions in chance arbitrarily close to n when the truth is H_n, using just convergence and arbitrarily good approximations of simple curves by means of arbitrarily complex curves with arbitrarily flat bumps. Since no convergent discovery strategy can avoid the retractions performed by Ockham's razor, one can say that Ockham's razor is *efficient* or *optimal* in terms of following the certifiably straightest or most efficient route to the truth.

The only escape from the retraction argument is to fail to find the truth at all, a remedy worse than the disease. Even Bayesian rationality, with its degrees of belief apportioned appropriately to the evidence, cannot escape the argument—*because* it converges to the truth. Say that a Bayesian retracts H to degree r if her degree of belief in H is above r with chance r at one point and is below $1 - r$ with chance r at a later point. Now use convergence and approximation just as before to force the Bayesian into retractions in chance arbitrarily close to n when H_n is true.

Since the argument is, perhaps, unfamiliar, it may be helpful to actually see the retractions in chance of a real method for statistical theory choice employed in statistics, machine learning, and data-mining, namely, the Bayes' Information Criterion or BIC for short (Schwartz 1978). One of the standardly cited advantages of BIC is that it converges to the right answer, because the BIC score of a theory converges to the degree of belief of a simplicity-biased Bayesian. The BIC score for a model T with k free parameters on sample E of size n is just:

$$BIC(T, E) = \log(\text{highest chance } E \text{ could have under } T) - \frac{k}{2}\log(n),$$

and the strategy is to choose the theory that maximizes the BIC score. The left-hand-term rewards models that can be "fit" closely to the data (i.e., that make the data very probable) while the right-hand-term penalizes the number k of free parameters in the theory—an obvious nod to Ockham.

Suppose that the truth is a bivariate normal distribution over two variables X, Y and H_0 says that both variables have zero mean, H_1 says that one variable has zero mean, and H_2 says that neither variable has zero mean. As in the polynomial case, H_0 is zero dimensional, H_1 is uni-dimensional, etc. so, statistically the problems are quite analogous although this one is easier to plot and to visualize. The Mathematica plots depicted in figure 1 illustrate the retractions in chance by BIC as the sample size n increases. The strangely shaped white zone covers the points in sample mean space at

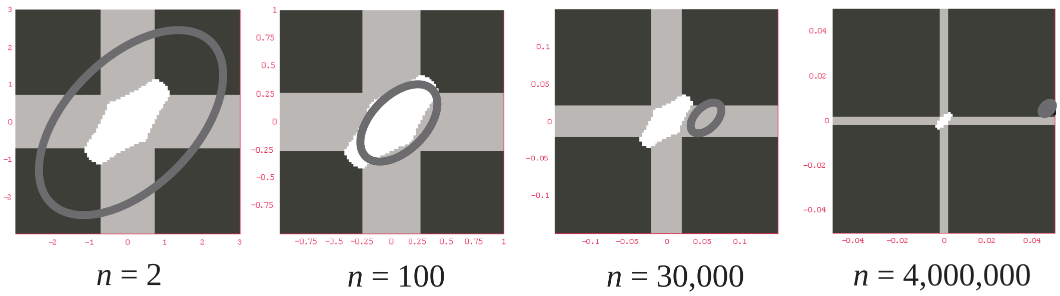

Figure 1: Retractions in chance by BIC

which BIC selects the simplest answer (that both coordinates of the population mean are zero).[7] The cross-shaped grey zone covers points at which the next simplest answer is selected (that one coordinate of the population mean is zero) and the black region is where the most complex answer is selected. Graphically, Ockham's razor corresponds to the fact that the white zone is "on top of" the grey zone which is "on top of" the black zone. The unfilled oval line is the boundary of the 95% quantile or footprint of the true sampling distribution, which is in this case has mean vector $(.05, .005)$. The BIC strategy first "takes the bait" for the simplest answer at around $n = 100$ (note that the 95% quantile is nearly contained within the acceptance zone for the simplest answer). BIC "notices" that the first component of the mean vector is nonzero at $n = 30,000$ (the 95% quantile is now nearly contained in the acceptance zone for the next-simplest answer) and then notices that the second is nonzero at around $n = 4,000,000$, for approximately two retractions in chance—the theoretical optimum in this case. In spite of the suggestive form of the BIC scoring rule, BIC falls short of implementing Ockham's razor fully due to the black and grey regions inside the oval at $n = 2$. In consequence, the exclusion of these regions at $n = 100$ amounts to an extra retraction that could have been avoided by expanding the white region or by returning no answer at all.

5 Unique Optimality

So Ockham's razor is certifiably optimal, but is it uniquely certifiably optimal? Recall the most striking feature of the freeway metaphor—the entrance ramp is the most direct route home even if it is a bit out of the way, so one ought to take the local resident's advice. Something similar is true of Ockham's razor. Think of a child following her

[7]The strange shape is a consequence optimizing a score expressed as the sum of a fit-rewarding term and a complexity penalizing term. The origin scores better than the background in a naturally shaped oval, but the origin scores better than the axes in a caustic star shape that extends far into the sectors and pinches in along the axes. The zone in which the origin does better than both is the intersection of the caustic star with the oval, as depicted in the figure. The Akaike information criterion or AIC yields a zone of the same shape but a different size (Akaike 1973). Contemplation of the shape may make one think twice about using either criterion, in spite of the cuteness of the formulas and the nod to Ockham's razor.

mother from one errand to another. If the child stays with or behind her mother, she follows the same path to the ultimate destination as her mother, which is the best she can do. If she ever runs ahead to the next store before her mother leaves the current one, the current store may have been her mother's last destination, so if her mother waits there long enough she has to backtrack to avoid permanent separation. After she returns to her mother's side, her mother is free to go to the next shop and the next. Now the daughter's path involves at least one extra U-turn compared to her mother's. Her "pursuit curve" is inefficient.

More literally, in the polynomial example, consider the world state in which the true parameter values are (a_1, a_2), where $a_1, a_2 > 0$ and a_2 is quite small but a_1 is large enough for H_0 to be rejected at low significance. So one might say that H_1 is the simplest answer compatible with experience and, hence, is the answer Ockham's razor should favor. Suppose, instead, that your strategy has a high chance of producing the *true* answer H_2 at sample size n (any answer other than H_1 would suffice, but this is the most interesting case, since it pits Ockham against short-run truth). Then since possibility $(a_1, 0)$ has a sampling distribution similar to that of (a_1, a_2), there is a non-negligible chance β that you accept H_2 at $(a_1, 0)$. Since you converge to the truth at $(a_1, 0)$, the chance of producing H_1 at $(a_1, 0)$ rises as high as you please as sample size increases, so the chance of producing H_2 drops from β to as near to zero as you please. Now the preceding argument forces nearly another unit retraction at some nearby world (a_1, a_3) satisfying H_2, for a total of $1 + \beta$ retractions after sample size n in the true answer H_2. Favoring the Ockham answer would have resulted in retractions certifiably near 1 in answer H_2 after rejection of H_0 at sample size n. So it is better, in terms of certifiably direct approach to the truth, to follow Ockham's razor—even if the violation happens to be true. Hence, there is no need to know in advance whether the truth is simple or complex, so there is no need for circles or Ouija boards to explain why Ockham's razor is the best possible truth-finding policy.

The retraction optimality argument does *not* establish that Ockham's razor is reliable. Since simple samples can be approximated by complex means, there is no way to certifiably avoid retractions in chance in theory choice. That is crucial to the asymmetry between simplicity and complexity: the retractions can be *forced* toward theories of higher in complexity but not in reverse. Pretense to reliability implies that nature cannot force retractions and, hence, makes it impossible to explain Ockham's razor without the usual recourse to wishful thinking or circles.

The argument is based on certifiable bounds. An Ockham violator can get lucky in terms of retractions if nature is kind enough to reveal all the anticipated complexities immediately, so it is not claimed that Ockham's razor does better in every possibility. Also, whether or not Ockham's razor does better in terms of expected retractions depends upon which question-begging prior probability distribution one adopts. It is not claimed that the above argument takes precedence over genuine plausibility considerations. Rather, it explains and justifies the usual selection of an Ockham ignoredge profile when one is actually ignorant.

The forcibility of retractions toward ever more complex theories also sheds light upon Goodman's whack/blite puzzle. Simplicity is not a matter of mere notation that can be coded away. Rather, the empirical complexity of a theory T, relative to a theory choice problem, is the number of retractions nature could force an arbitrary, convergent scientist to perform prior to convergence to T (Kelly 2007a, 2007b). That ordering determines the freeway route to the truth and Ockham's razor directs inquiry along the straightest path to any destination along the route.

6 Conclusion

This essay began with the puzzle that Ockham's razor seems to require a metaphysical commitment to the simplicity of nature, pre-judging the future course of scientific inquiry. As avenues for addressing the puzzle, two fundamental metaphors for scientific reasoning were considered; argument and inquiry. The argumentative viewpoint leads ineluctably to a circle. The perspective of inquiry, on the other hand, explains Ockham's razor, without presupposing that the truth is simple, as the straightest possible path to the truth.

7 Acknowledgements

This work was generously supported by NSF grant number 0740681. The author is indebted to discussions with Sherri Roush, Jonathan Vogel, Conor Mayo-Wilson and Steve Fancsali, which led to the refinement and clarification of some of the preceding arguments.

8 Bibliography

Akaike, H. (1973) "Information theory and an extension of the maximum likelihood principle," *Second International Symposium on Information Theory.* pp. 267-281.

Aristotle (2007) *Posterior Analytics*, G.R.G. Mure, trans., eBooks@Adelaide, University of Adelaide, http://etext.library.adelaide.edu.au/a/aristotle/a8poa.

Bonjour, L. (1980) "Externalist Theories of Empirical Knowledge," *Midwest Studies in Philosophy* 5: 53-73.

Carnap, R. (1962) *Logical Foundations of Probability*, Second Edition. Chicago: University of Chicago Press.

Cartwright, N. (1999) *The Dappled World*, Cambridge: Cambridge University Press.

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge: M.I.T. Press.

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

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

Forster, M. and Sober, E. (1994) How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science* 45: 1 - 35.

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

Goldman, A. (1967) “A Causal Theory of Knowing,” *Journal of Philosophy*, 73: 771-91.

Goodman, N. (1955) *Fact, Fiction, and Forecast.* 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., Oshersion, D., Royer, J., and Sharma, A. (1999) *Systems that Learn.* Cambridge: M.I.T. Press.

James, W. (1897) *The will to believe, and other essays in popular philosophy.* New York: Longmans, Green and Company.

Jeffreys, H. (1931) *Scientific Inference.* Cambrige: Cambridge University Press.

Kadane, J., Schervish M., and Seidenfeld, S. (1999) *Rethinking the Foundations of Statistics*, Cambridge: Cambridge Universitiy Press.

Kant, I. (1987) *The Critique of Judgment*, W. Pluhar, trans, Indianapolis: Hackett.

Kelly, K. (1987) “The Logic of Discovery,” *Philosophy of Science*, 54, 1987, pp. 435-452.

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

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

Kelly, K. (2002)“Efficient Convergence Implies Ockham’s Razor,” *Proceedings of the 2002 International Workshop on Computational Models of Scientific Reasoning and Applications*, Las Vegas, USA.

Kelly, K. (2007) “How Simplicity Helps You Find the Truth Without Pointing at it,” V. Harazinov, M. Friend, and N. Goethe, eds. *Philosophy of Mathematics and Induction*, Dordrecht: Springer.

Kelly, K. (2007a) "Ockham's Razor, Empirical Complexity, and Truth-finding Efficiency," *Theoretical Computer Science*, 383: 270-289

Kelly, K. (2007b)"A New Solution to the Puzzle of Simplicity", *Philosophy of Science* 74: 561-573.

Kelly, K. (2008)"Ockham's Razor, Truth, and Information", in *Philosophy of Information*, Van Benthem, J. Adriaans, P. eds. Dordrecht:*Elsevier* pp. ***-***.

Kelly, K. and Glymour, C. (2004) "Why Probability Does Not Capture the Logic of Scientific Justification," C. Hitchcock, ed., *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell, pp. 94-114.

Kitcher, P. (1993) *The Advancement of Science*, Oxford: Oxford University Press.

Koons, R. (2000) "The Incompatibility of Naturalism and Scientific Realism," in *Naturalism: A Critical Analysis*, J. Moreland and W. Craig eds., London: Routledge.

Keynes, J. A (1921) *Treatise on Probability.* London: Macmillan.

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

Leibniz, G. W. (1714) *Monadologie*, in *Die Philosophischen Schriften von G. W. Leibniz*, vol. IV. Berlin: C. J. Gerhardt, 1875, pp. 607-23.

Mitchell, T. (1997) *Machine Learning.* New York: McGraw-Hill.

Nozick, R. (2006) *Philosophical Explanations*, New York: Belknap Press.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Plato (2007a) *Meno*, B. Jowett, trans. *Internet Classics Archive*, http://classics.mit.edu/Plato/meno.htm

Peirce, C. (1878) "How to Make our Ideas Clear," *Popular Science Monthly* 12: 286-302.

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.

Quine, W. (1951) "Two Dogmas of Empiricism," *The Philosophical Review* 60: 20-43.

Ramsey, F. (1990) *Philosophical Papers* Cambridge: Cambridge University Press.

Reichenbach, H. (1949) *The Theory of Probability*, London: Cambridge University Press.

Roush, S. (2005) *Tracking the Truth*, Oxford: Oxford University Press.

Salmon, W. (1967) *The Foundations of Scientific Inference*, Pittsburgh: University of Pittsburgh Press.

Savage, L. (1954) *The Foundations of Statistics*, New York: John Wiley and Sons.

Schulte, O. (1999) "Means-Ends Epistemology," *The British Journal for the Philosophy of Science*, 50: 1-31.

Schulte, O. (2000) "Inferring Conservation Principles in Particle Physics: A Case Study in the Problem of Induction," *The British Journal for the Philosophy of Science* , 51: 771-806.

Schwarz, G. (1978) "Estimating the dimension of a model," *Annals of Statistics* 6: 461464.

Sklar, L. (1977) *Space, Time, and Spacetime*, Berkeley: University of California Press.

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

Jonathan Vogel (2000): "Reliabilism Leveled," *Journal of Philosophy*, 97, pp. 602-23.

Wasserman, L. (2004) *All of Statistics: A Concise Course in Statistical Inference.* New York: Springer.