2004

Justification as Truth-Finding Efficiency: How Ockham's Razor Works

Kevin T. Kelly
Carnegie Mellon University

Justification as Truth-Finding Efficiency: How Ockham's Razor Works

KEVIN T. KELLY
College of Humanities & Social Sciences, Department of Philosophy, Carnegie Mellon University, Baker Hall 135, Pittburgh, PA 15213, USA; E-mail: kk3n@andrew.cmu.edu

Abstract. I propose that empirical procedures, like computational procedures, are justified in terms of truth-finding efficiency. I contrast the idea with more standard philosophies of science and illustrate it by deriving Ockham's razor from the aim of minimizing dramatic changes of opinion *en route* to the truth.

Key words: confirmation, convergence, mind-changes, model-selection, naturalism, Ockham, learning, simplicity

1. Introduction

The philosophy of science divides, roughly, into two schools. *Confirmation theorists* seek to explicate the concept of empirical justification *a priori* by systematizing intuitive reactions to various historical episodes and case studies. *Epistemological naturalists* view scientific practice as an empirical subject in its own right and seek to determine by empirical means which sorts of methods are well-adapted to finding the truth. The two groups tend to view the topic of this special journal issue – the relationship between computer science and the philosophy of science – rather differently. According to confirmation theorists, the job of explicating the concept of confirmation falls primarily to philosophers; computer scientists are left to deal with the applied task of determining which theories are confirmed or of searching for highly confirmed theories. According to naturalists, computers can implement a variety of inductive methods, which can then be applied to lots of computer-generated problems in order to produce extensive evidence about the empirical effectiveness of different inductive strategies. Either way, computational ideas seem peripheral, or merely auxiliary, to the topic of scientific justification.

I advocate the opposite view, that justification is just truth-finding efficiency and that the philosophy of science should look to the theories of computability and computational complexity as models of how to study it (Kelly, 1996). On my view, computer science is not merely a helpful assistant to the philosophy of science; it is a valuable resource for deep and systematic ideas about scientific justification itself. In this note, I describe the general

Minds and Machine **00:** 1–21, 2004.

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☐ CP
PDF-OUTPUT

viewpoint just sketched and illustrate it by deriving Ockham's razor from a kind of truth-finding efficiency in both empirical and computational contexts, without appealing, in the usual way, to a prior bias toward simplicity or to a primitive concept of confirmation or empirical rationality.

2. Confirmation Theory

Scientific laws and theories have consequences that outrun the available evidence, so the question of their justification naturally arises. A familiar response to this skeptical challenge is confirmation theory: the view that evidence partially justifies full belief (or fully justifies partial belief). The focus, then, is to 'explicate' the underlying concept of 'confirmation', which amounts to finding a simple, general rule or definition that more or less agrees with our intuitions regarding evidential support in a wide range of examples (e.g. Carnap, 1950; Hempel, 1965; DeFinetti, 1972; Glymour, 1980).

Confirmation theory is not merely a philosophical phenomenon. For example, the DENDRAL program (Buchanan, 1974) optimized a kind of confirmation score over possible molecular hypotheses and contemporary machine learning programs optimize such scores as the Bayes' information criterion or the Akaike information criterion (cf. Wasserman, 2000). In such procedures, there is a clear division of labor between the scoring rule, which amounts to a proposed confirmation relation, and the subsidiary rules of search that sift among possible models. Such practice fits well with confirmation theory's conception of its relationship to computer science; for according to confirmation theorists, empirical justification is nothing but confirmation, so after it is known that a theory is confirmed, the theory is justified independently of how we came to know or compute this fact (Laudan, 1980). So although computational search procedures are useful for finding confirmed hypotheses, they have nothing *per se* to do with the justification of the hypotheses they output.

Confirmation theory has its appeal. The concept of confirmation seems to respond to Hume's challenge for a justification of fallible belief. It also cuts a scientific theory loose from the history by which it was obtained, for the theory is justified by its current confirmation, regardless how it was discovered. Finally, confirmation theory neatly divides labor between pure philosophers of science, who speculate about the meaning of empirical justification without regard for computational efficiency, and practical computer engineers, who seek efficient procedures for checking the confirmation of particular theories or for finding maximally confirmed theories.

My concern is that such a collaboration can be sophisticated and formally challenging without ever getting around to the two most obvious and

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

pressing questions in the philosophy of science: whether the overall procedure hobbled together out of search and confirmation principles is any good at finding the truth and whether alternative approaches would be equally good or better. Finding the truth is the issue and everything in theoretical computer science is directed toward finding it efficiently. So why put philosophers in charge and trust them to decide that a broadly computational concern with truth-finding efficiency is useless in the empirical domain?

3. Empirical Naturalism

Whereas confirmation theorists take scientific methods to be justified insofar as they produce confirmed conclusions, epistemic naturalists reverse the story, so that conclusions are justified insofar as they are produced by reliable or truth-tracking methods (e.g. Nozick, 1981; Goldman, 1986). Since reliability is an empirical property of a learning method or disposition, it is, itself, subject to empirical investigation, e.g. by running a machine learning program on different learning problems to see how it does. Hence, computer scientists can be useful lab assistants, since they can program different inductive strategies and run them on lots of simulated test cases, presumably leaving philosophers in a theoretical or management position (e.g. Laudan, 1996).

I like the new devotion to reliability and truth-finding, but I am skeptical of the narrowly empirical perspective. Some evidence that a method 'works' in a certain application is fine, but one would also like to know more general and explanatory things, such as what it is about the mathematical structure of a learning problem that makes finding the truth hard or easy, what it is about methods that make them work well in problems of a certain kind, whether a given method can be improved, whether there is any truth-finding rationale for our inductive prejudices (a.k.a. principles of confirmation), and so forth. Such questions are hard to answer by applying particular methods to problem sets, by scouring the history of science for examples, or by appealing to general evolutionary considerations for a biological warranty on our wiring. They sound, rather, like the questions routinely studied in the theories of computability and computational complexity.

4. Computational Naturalism

Does there exist a method for answering a certain formal question? Could the question be answered in a stronger sense or more efficiently? What structural features of a question make it hard or easy to solve and how are algorithms helped or hindered by such structural features? Questions like these have been addressed routinely by computability theorists for over a half century (e.g. Rogers, 1967; Hinman, 1978; Garey and Johnson, 1979). The

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

mathematical theory of computability is naturalistic in the sense that rational mechanics or the theory of dynamical systems are: if the formal description fits a mechanical device, then the theorems aptly describe what it can or cannot do. And yet it has normative import, for an algorithm is justified, hypothetically rather than categorically, insofar as it solves its appointed problem efficiently. Moreover, although the results of the theory govern what real systems that implement such algorithms can do, the reasoning involved is purely *a priori* and mathematical. In short, theoretical computer science provides a deep and impressive model of how naturalism is compatible with a mainly *a priori* and normative approach.

On the other hand, computability concerns the crisp world of algorithms that halt with the right answer; empirical science is intrinsically fraught with uncertainty and guesswork. So isn't it obvious that the right mathematical framework for science is probability theory rather than computability? Not in the slightest. First, not all formal problems are computable and the ones that aren't seem quite analogous to Hume's problem of induction (e.g. 'will an arbitrary computation continue to run forever?') Moreover, some empirical questions can reward extreme diligence with infallible results (e.g. 'is there a needle in the haystack') and they resemble decidable formal problems. Compare apples with apples and oranges with oranges.[1] Second, it is far from obvious that assigning probabilities or intermediate degrees of belief to hypotheses respond in a relevant way to the issue of undecidability in either the empirical or the formal domain. Yes, if you ever become completely certain that a general law is true, you might encounter a counterexample tomorrow and have to completely reverse your opinion. But if your partial degrees of belief converge to unity on the basis of increasing numbers of positive instances, as is often claimed by advocates of partial degrees of belief, then you may have to reverse your opinion from $1 - \epsilon$ to zero anyway, where $\epsilon > 0$ is arbitrarily small. Either way, the real issue posed by undecidability is that any possible method guaranteed to converge to the truth must undergo a severe reversal of opinion in some possible worlds.

Putting the two points together, it is natural to concede that some questions – both formal and empirical – can only be solved by methods that may change their opinion quite sharply prior to arriving at the right answer, and to study the efficiency of such methods in a unified, broadly computational framework. The basic idea was first articulated by Putnam (1965) and has since been developed in detail in the literature of 'formal' or 'computational learning theory'.[2]

5. Counting Retractions

A familiar, philosophical objection to the viewpoint just described is that any finite variant of a convergent method is still convergent (e.g. Salmon, 1967),

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

so that strong, intuitive biases regarding theory choice in the short run are left unexplained. It is widely believed, therefore, that there must be some short-run relation of confirmation that justifies one such conclusion better than others, etc. (e.g. Earman, 1992).

But that argument neglects considerations of efficiency: although many routes lead to the truth, the best route may be unique. In standard computability theory, computations halt and computational complexity is measured in terms of computational steps or storage space expended prior to halting. In the case of convergent, or non-halting methods, there is another, natural measure of cognitive complexity: the number of times the method retracts its opinion prior to convergence to the right answer, where a retraction is a change in opinion in which the new view does not entail the old view, so that some loss of content occurs (Gäerdenfors, 1988).

There is a cultural divide on this issue. Retractions are intensively examined in computational learning theory, but when I mention the idea to philosophers, they ask why anyone would care. Whenever this happens, I can't help but imagine a philosopher who keeps running around the block, crying 'All I care about is being home.' When you express surprise, she responds: 'When did I say I cared about the number of times I run around the block?' There is such a thing as *too* pure a love of truth, for as Plato remarked, true lovers are seekers and gratuitously roundabout seekers are unworthy of the name. Computational learning theorists recognize that and instinctively recognize retractions as a cognitive or epistemic consideration analogous to halting. The halting of an algorithm with the right answer is what confers certainty on the output and is ultimately why proof systems yield certainty. Counting retractions measures a kind of epistemic distance from the ideal of halting and, hence, is not merely a practical matter. Indeed, there is a hierarchy of formal problems unsolvable in the usual sense that are solvable with bounded numbers of retractions, and there is a parallel hierarchy of empirical problems (cf. Kelly, 1996; Jain, 1999). Furthermore, incorrigibility is a traditional topic in philosophy and retractions measure the distance from that ideal as well. Finally, Kuhn (1970) caused a philosophical sensation by urging an explicitly diachronic perspective on science in which sharp breaks or revolutions in opinion constitute an inevitable and healthy part of the process. Surely, taming this bohemian notion within a broadly computational, normative theory of truth-finding efficiency should be of some philosophical interest.

Computer scientists who study retractions are usually interested in classifying problems – formal and empirical – according to the number of retractions required to solve them. My aim is a bit different: it is to address the 'short-run arbitrariness' objection to convergent methodology by showing that key intuitions regarding short-run theory preference follow deductively from the fact that the method in question minimizes retractions *en*

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

197 *route* to the truth. In particular, I will derive a version of Ockham's razor
198 entirely from retraction-efficient convergence to the truth.

199 6. Ockham's Razor

200 One of the deepest puzzles in the philosophy of science concerns simplicity.
201 When several, possible theories are compatible with experience, scientists
202 incline toward the 'simpler' one, where simple theories are somehow more
203 uniform, symmetrical, unified, or free from independently adjustable
204 parameters. The question is how a fixed bias toward simplicity helps science
205 find the true theory. The outlook for an answer seems bleak, on the face of it,
206 for a fixed bias toward simplicity can no more indicate the (possibly complex)
207 truth than a broken thermometer, whose reading never changes, can indicate
208 temperature. There are attempted responses, but they tend to be of three
209 disappointing types: circular, bait-and-switch, or insufficient – which is not at
210 all to say that they are trivial or easy to implement. On the contrary, the
211 mathematical formidability of some of these ideas is part of their allure: it
212 seems that so much mathematical trouble must somehow help us find the
213 truth.

214 6.1. CIRCULAR ARGUMENTS

215 (1) You can simply adopt prior probabilities biased toward simple theories
216 (Jeffreys, 1985; Salmon, 1990). But that merely presupposes the very bias to
217 be explained. Nor does it explain how adopting such a bias is better than
218 adopting a prior bias toward complex theories.
219 (2) According to the *minimum description length* approach (Solomonoff,
220 1964), sequences computed by smaller Turing machines are more probable
221 than sequences computed only by larger machines. But that is simply because
222 a prior bias toward small machines is assumed.
223 (3) According to the *minimum message length* approach (cf. Rissannen,
224 1983 and the discussion in Mitchell, 1997), an efficient coding scheme assigns
225 shorter codes to symbols that are more likely to be transmitted to some
226 recipient, in which case the expected message length is minimized. If 'chances
227 of transmission' are interpreted as prior probabilities on theories, then the
228 principle stipulates that more *a priori* probable theories should be assigned
229 shorter descriptions. The result is the same as before: a prior bias toward
230 possibilities with shorter descriptions.
231 (4) It is sometimes claimed that the data would be a miracle if a complex
232 theory were true. That finally sounds like a matter of objective likelihood,
233 rather than of prior opinion. Is it? Let S be a simple theory that entails e. Let

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

234 C be a complex theory that has a free parameter i that can be twiddled among
235 n possible values to 'save' the data, so that $C = C_0 \vee C_i \vee \cdots \vee C_n$. Simple
236 theory S has no such parameter and faces the tribunal of experience with a stiff
237 Popperian lip. If the parameter i is 'free', we shouldn't have strong *a priori*
238 ideas about how it is actually fixed, so assume a uniform distribution. Suppose
239 that, as it happens, S entails e and C_0 entails e, but for all other i, C_i is
240 incompatible with e. Then we have $P(e|S) = 1$ but $P(e|C) = \sum_{i \leq n} P(C_i)$
241 $P(e|C_i) = P(C_0)P(e|C_0) = P(C_0) = P(C)/n$, which is very small when n is
242 very large. So it seems that e would be a miracle given C if n is sufficiently large
243 (Rosencrantz, 1983). Notice, however, that the miraculously low value
244 $P(C)/n$ is simply the agent's subjective prior probability for parameter value
245 C_0 which gets passed through the weighted average. Since prior probabilities
246 are assumed in the argument anyway, consider the ratio of posterior proba-
247 bilities. By Bayes' theorem,

$$\frac{P(S|e)}{P(C|e)} = \frac{P(e|S)P(S)}{P(e|C)P(C)} = \frac{P(S)}{P(C)/n \cdot P(C)} = \frac{n}{P(C)}\frac{P(S)}{P(C)}.$$

249 So as the number n of parameter values increases, one would have to be
250 increasingly 'unfair' to S in order for C to win. But notoriously, Bayesians
251 can't be fair in every comparison: fairness over 'blue versus non-blue' is bias
252 over 'blue versus green versus something else'. In this case, 'fairness' in the
253 contest of S against C induces a huge bias in favor of S compared to each C_i,
254 including C_0, and the miracle argument merely passes this bias along until S
255 is refuted and C wins for sure. So the miracle argument amounts to a circular,
256 prior bias in favor of simple worlds.

257 ### 6.2. BAIT-AND-SWITCH ARGUMENTS

258 (1) Simple theories have other virtues: explanatory power (Harman, 1965),
259 testability (Popper, 1959; Glymour, 1980), and symmetry or unity (Friedman,
260 1974; Kitcher, 1989). But absent a prior probabilistic bias toward simplicity,
261 none of these features indicates the truth. So *inferring* a theory on such
262 grounds is an instance of wishful thinking (VanFraassen, 1980).

263 (2) When one uses a model for purposes of prediction, one may do better
264 with an over-simplified model than with the true one if the true one has many
265 free parameters. That provides a kind of truth-directed motive for choosing
266 simpler models (Forster and Sober, 1994), but the motivation is not that
267 doing so helps you find the true model (cf. Wasserman, 2000). Still, I prefer
268 this story to the preceding ones: at least it is relevant, in a non-circular way,
269 to finding the truth about something.

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

6.3. INSUFFICIENT ARGUMENT

Nor does it suffice to observe that if the simplicity presumption is mistaken, future evidence will swamp it and set us right eventually (Sklar, 1977), for a mistaken presumption of complexity would also be set right, eventually, by an increasing sequence of 'null results' in which anticipated complexities persistently fail to appear. An asymmetry is required. But the idea is on the right track, and would be quite interesting if it could be strengthened by efficiency considerations to *single out* Ockham's razor as the right bias for efficient convergence to the truth. That is the approach I will now pursue.

7. How Ockham Helps

The puzzle at hand is to explain how fixed advice could possibly help you find something, regardless where it is, unless you already know where it is. Some magical connection seems to be required and, too often, the magical connection provided is a question-begging prior probability assignment. But the puzzle conceals a questionable assumption: that 'help' means 'immediately indicate or point to'. Usually you want – and get – a different kind of help. Suppose you become lost in a small town on your way home from another city. You ask a local resident for directions. She gives you street directions to the nearest freeway entrance ramp. Such advice doesn't necessarily indicate where your home is, since the entrance ramp could be in the opposite direction. And yet you ought to follow the advice, for suppose you disregard it. Then you eventually end up on the outskirts of town on winding, rural routes until you finally make a U-turn, retrace your route back to the helpful resident, and then follow her advice to get on the freeway, which is the most direct route home, wherever home is along the highway corridor. So your reward for disregarding her advice is an an extra, needless, U-turn before you even get properly started on the long freeway journey home. The same will be true at each future detour from the freeway: if you disregard the local advice, you add an extra U-turn to the remaining route home, from that point onward. So you should always follow the (fixed) local directions to the freeway entrance, no matter where you happen to be headed.

That is essentially how Ockham's razor helps you find the truth without indicating what the truth is: disregarding Ockham's advice opens you to a needless, extra U-turn or reversal in opinion prior to all the reversals that even the best of methods would have to perform if the same answer were true. So you ought to heed Ockham's advice. Simplicity doesn't indicate the truth, but it minimizes reversals along the way. That's enough to explain the unique connection between simplicity and truth, but it doesn't promise more than

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

can be delivered, namely, a philosophical warranty against more twists and bumps in the future.

8. Counting Things

Consider a very simple situation in which marbles (or new types of particles or empirical 'effects') are presented at irregular intervals from some source that is known to contain at most finitely many marbles.[3] The question is how many marbles will be presented. Ockham recommends positing no more marbles than have been seen so far. Several intuitive aspects of simplicity conspire toward this conclusion. First, that answer minimizes 'entities' (i.e., marbles). Second, it is most uniform (no more marble appearances). Third, it is most testable (it is crisply refuted by another marble if it is false). Fourth, it has the fewest 'adjustable parameters' (the times of appearance of the posited marbles).

The preceding glosses depend on the question asked. For example, suppose the question had asked how many 'tharbles' will be observed, where a 'tharble' is a non-marble prior to stage 1000 and a marble from stage 1000 onward (Goodman, 1983). Then the hypothesis 'no tharbles' is most uniform (in terms of tharbles) and has fewest adjustable parameters or auxiliary assumptions (concerning 'appearance of new tharbles'), but 'no tharbles' is inconsistent with 'no marbles'. Goodman's example was a scandal in confirmation theory, but it is just what a computer science undergraduate is trained to expect. Efficiency is always relative to an aim or problem, so if simplicity is to help us answer questions about nature, it must, somehow, conform itself to the contours of the question asked.

The U-turn argument for Ockham's razor is almost a retelling of the freeway story. Suppose you converge to the true answer in each possible world, but you disregard Ockham's advice after receiving finite sequence of experience σ including n observed marbles by concluding some number of marbles other than n. Then σ is compatible with the world w_n that presents σ followed by uniformly marble-free experience. On pain of failing to converge to the truth in w_n, you must retract, eventually, to 'n marbles' – say, by the time finite sequence of experience τ_n extending σ has been received. That is your extra, initial U-turn, for had you listened to Ockham, you would never have retracted after σ in w_n. Now there exists a world w_{n+1} compatible with σ in which one marble is presented after σ. Since you converge to the truth, you eventually retract 'n marbles' in favor of '$n+1$ marbles', say by the end of τ_{n+1} extending σ. The argument continues in this manner forever, so we have constructed an infinite sequence of worlds $w_n, w_{n+1}, \ldots$ such that you retract $k+1$ times after σ in w_{n+k}, which satisfies '$n+k$ marbles'. The obvious Ockham method that simply counts the marbles as they appear, on the other

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

hand, retracts at most k times after σ in an arbitrary world satisfying '$n + k$ marbles'. So your method retracts at least one more time than the Ockham method's worst-case performance within each answer compatible with σ. That's all there is to it.

Say that the *subproblem* entered in σ consists of all worlds compatible with σ and the answers to the subproblem are just the answers to the original problem that are compatible with σ. Furthermore, think of inquiry in the subproblem as beginning at the end of σ, so don't count retractions occurring along σ in the subproblem. Finally, exclude from the subproblem all answers incompatible with it. Then it has been shown that *if you are guaranteed to converge to the truth, then in the subproblem entered at your violation of Ockham's razor, your worst-case retractions in each answer to the subproblem exceed the obvious counting method's by at least one*. This can be summarized by saying that the obvious Ockham method *strictly dominates* your method *in worst-case retractions over answers in the subproblem*.

The key to the argument is the concept of strict dominance in worst-case retractions over possible answers. One cannot argue that your method is strictly dominated with respect to retractions *simpliciter*, since violation of Ockham's razor may result in fewer retractions in some complex worlds (e.g. those in which the anticipated marbles appear very quickly, before the method's confidence collapses). Nor can one argue that your method's overall worst-case retraction bound exceeds that of the Ockham method, since there is no finite such bound for either method (any number of marbles might come later). Finally, one cannot argue that the expected number of retractions is higher for your method than for the Ockham method, since any such argument would appeal to a question-begging prior probability distribution biased toward simple worlds.

The reason for bringing up subproblems in the argument is this. Suppose that a method guesses 'no marbles' *a priori* and persists in this conclusion after seeing the first marble. That's a violation of Ockham's razor. True, the method can be forced back to the true hypothesis 'one marble' by withholding new marbles long enough, but in the *overall* problem, that still amounts to just one retraction, which even the Ockham method requires in the worst case if 'one marble' is true. It is only in the subproblem entered upon seeing the first marble that the Ockham method dominates the violator, for in that subproblem the violator requires one retraction at some world satisfying 'one marble' but the Ockham method requires none.

There is more to be said. *Weak dominance* in worst-case retractions over answers means that some other method's worst-case retractions over answers are as good in each answer and better in some answer. A similar argument to the one just given shows that there is no subproblem in which an alternative method even weakly dominates the obvious counting method in worst-case retractions. I say, then, that the obvious counting method is a

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

retraction-efficient solution to the full problem. So no violator of Ockham's razor is retraction-efficient, but the obvious Ockham method is. It follows that the only retraction-efficient methods are those that either agree with the obvious counting method or refrain from saying anything at all for some finite number of steps. Thus, *the substantive outputs of retraction-efficient methods must all agree* in this problem.

Still more can be said. Someone might retract the uniquely simplest theory on general, skeptical grounds (after reading Hume, say) before it is refuted by seeing yet another marble. In science, that general sort of skepticism is frowned upon: the simplest theory holds its ground until it gets into concrete, empirical trouble. Call the additional requirement that one should never drop the Ockham answer until it gets into concrete empirical trouble the *retention principle*. The retention principle also follows from a worst-case retraction dominance argument. For suppose you choose the answer '*n* marbles' in light of finite sequence of experience σ, which presents *n* marbles and then retract in light of experience τ extending σ that presents no new marbles even though '*n* marbles' is not yet refuted. So in the subproblem entered when σ is seen, one retraction occurs at τ, whereas a non-retracting Ockham method wouldn't retract after σ. So if no more marbles are seen, you use one more retraction than the method that hangs on to the Ockham hypothesis until it is refuted. That is an initial U-turn in the sub-problem entered upon experiencing σ. Now as before, nature can elicit at least one more retraction for each successive marble, so no matter which answer is true, you use more retractions than the method that hangs onto the Ockham answer until it is refuted. So you shouldn't have dropped the Ockham answer until it was refuted.

9. Freeing Parameters

Real science is a lot more than counting marbles; it involves constructing laws and models that explain empirical effects. But there is an important analogy. 'Effects' (e.g. causal influences, correlations, monomial coefficients in polynomial laws) may be small or arcane and may not show up right away, but when they do, they reveal that our current, simplistic models are wrong. In such circumstances, Ockham's razor is understood to favor waiting to add a parameter until the corresponding effect is verified. And if effects were as discrete as marbles, then arguments very similar to the preceding ones derive both Ockham's razor and the retention principle from retraction-efficiency.

Even in this more general setting, the preceding results concede just one degree of freedom to retraction-efficient methods in problems of the sort under discussion: *they can differ only in how much uniform experience after the last retraction they require before leaping to the next Ockham hypothesis.*

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

Furthermore, since the measurements that verify 'effects' can usually be presented in an arbitrary order, *whenever two retraction-efficient methods receive the same data set (regardless of the order in which it was collected) they produce the same answer if they produce any answer at all*. All of this is derived entirely from a concept of minimizing reversals of opinion *en route* to the truth – none of it was imposed in advance due to prior ideas about what 'rational' inquiry or 'confirmation' have to be like.[4]

Recall the objection that confirmation theory is necessary to explain short-run theory preferences and to screen scientific justification from the psychological accidents by which a theory was produced.[5] Both challenges are met in the preceding example, if retraction-efficiency is taken into account in addition to convergence to the truth.

10. Bayesian Retractions

Although I have been rather hard on the Bayesian penchant for smug, circular explanations, Bayesian methods are quite another matter, and Bayesians who are sufficiently biased toward simple models in the typical way will tend to look good in terms of the preceding, non-circular argument. Let α be a fixed quantity strictly between zero and one half such that values lower than α are 'small' and values higher than $1 - \alpha$ are 'large'. A Bayesian agent can be said to retract when her posterior probability drops from a high to a low level on some answer to the question at hand. With this slight modification, the U-turn argument also applies to Bayesian agents whose posterior probabilities really converge to the truth (i.e. not just in the Bayesian's own mind: cf. Kelly, 1996 for more on this distinction). Moreover, Bayesians with a prior bias toward simple theories will tend to be retraction-efficient, since the high prior probability will remain if the simplest theory is true, will 'wash out' in favor of the next-to-simplest theory if that is true, and so forth, for a total of k retractions in the kth simplest answer. So Bayesians can have it both ways, as they do with their worst-case 'Dutch Book' arguments. Among the faithful, they can get by with their usual, circular appeals to their own simplicity biases. When seriously challenged, they can revert to the worst-case U-turn argument to get skeptics on board. After the skeptics are converted, the circular argument is good enough.

On the other hand, Bayesians and their biases have no monopoly on retraction-efficiency. Again, the issue is convergence to the truth with minimal retractions and hedging one's bets in the short run doesn't make them go away; for if the problem requires a certain number of retractions in each answer, then the Bayesian can be forced into that many retractions *no matter how small $\alpha > 0$ is chosen to be*. Moreover, the mathematical and computational difficulties frequently encountered when Bayesians attempt to define

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

appropriate prior probabilities can be skipped by simply adopting a method that chooses the uniquely simplest answer compatible with experience first.

11. Ockham's Statistical Razor

Scientific 'effects' are usually not as hard-edged as marbles. Randomness in nature and error in measurement introduce chance, so that the detection of an effect is no longer certain. Still, there is an intuitive sense in which small effects are probably missed at low sample sizes and are probably recognized when sample size increases sufficiently, with a grey interval in between in which the two chances are comparable.

A *statistical problem* is a partition over some set of statistically possible worlds, which may be thought of as possible models with their parameters fixed one way or another. I will focus on the special case in which all possible statistical parameters constitute a (potentially infinite) vector space and models result from setting all but finitely many of the parameters to zero. A *statistical method* is just a rule or strategy that (perhaps randomly) selects an answer to a statistical question in response to a sample of arbitrary size. Again, let α be a fixed parameter properly between zero and one half such that any probability less than α is 'small' and any probability exceeding $1 - \alpha$ is 'large'. Say that a statistical method *retracts in probability* in world w between sample sizes n and $n + k$ if the chance that it produces some answer drops from above $1 - \alpha$ at n to below α at $n + k$ in w. A Bayesian statistician retracts in probability in w between n and $n + k$ just in case she has a high chance of assigning a high degree of belief to an answer at n and a high chance of assigning a low degree of belief to the same answer at $n + k$, where 'high' and 'low' are understood in terms of α. To count overall retractions in a statistical world, tally the successive sample size intervals in which retractions occur. Finally, say that a method *solves a statistical problem in the limit* iff the chance that it produces the right answer converges to unity in each statistical world.

Statistical retractions are an unavoidable by-product of statistical inference. For consider a simple test of the point null hypothesis H_0 that the mean of a normal distribution with known variance is zero against the composite alternative H_1 that the mean has any value but zero. The usual approach to this problem is to adopt a test with low ($<\alpha$) significance of H_0. By tuning the significance level downward according to a sufficiently slow schedule as sample size increases, one can ensure that the chance of producing the null hypothesis if it is true rises monotonically to unity and that the power of the test also converges to unity at each alternative world as sample size increases. Hence, the sequence of tests is a statistical method that solves the binary decision problem in the limit. Moreover, the low significance level on H_0 assures that the method never retracts H_0 in probability as sample size

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

increases. On the other hand, the method probably produces H_0 in alternative worlds near to H_0 until the sample size rises to the point at which the corresponding test probably rejects H_0 in favor of H_1. That constitutes a retraction in probability. So the sequence of tests solves the problem in the limit with no retractions in H_0 and with at most one retraction in H_1. That sounds quite similar to the problem of counting at most one marble: either it appears or it doesn't, so the obvious counting method retracts either zero times or one.

Furthermore, no possible method does better. For let an arbitrary method that solves the problem in the limit be given. Since the method converges in probability to the right answer in each world, it does so in the zero-mean world w_0. So there exists a sample size n_0 at which the method probably (i.e. with chance $>1-\alpha$) produces the null hypothesis H_0 that the mean is zero. In this case (and in typical cases), the chance of a fixed sample event at a given sample size varies continuously with the parameter. Hence, there is a small, nonzero value r for the mean such that the method still probably produces H_0 in world w_r in which r is the true mean. But again, since the method converges in probability to the right answer in each world, there exists a sample size $n_1 > n_0$ at which it probably produces H_1. That is a retraction in probability in H_1, so no possible method achieves better worst-case performance in each answer in this problem than zero retractions in H_0 and one retraction in H_1.

So the test-based method minimizes retractions over the whole problem. It might also be said to follow a statistical version of Ockham's razor, for it favors the 'simple' hypothesis that the statistical parameter in question (the normal mean) is zero at the expense of the complex alternative which frees it. Now suppose that your method reverses this bias and at some stage probably produces the complex alternative H_1 in the simple world w_0, in which H_0 is true at some sample size n_0. Since your method converges in probability to the right answer, there is a larger sample size n_1 at which the method probably produces H_0 in w_0. That is a retraction in probability and represents your method's initial U-turn as a consequence of its violation of Ockham's razor. Again, by continuity there exists a small $r > 0$ such that both answers are also produced in w_r at the corresponding sample sizes n_0, n_1. Since your method converges to the right answer, there is a sample size $n_2 > n_1$ at which the method probably produces H_1. So your method probably retracts at least twice in some world in the alternative hypothesis. Hence, your method does worse than the test-based method's worst in in each answer to the problem.

12. Ockham and Statistical Model Selection

The statistical U-turn construction is very general, requiring only the assumption that the chance of a particular sampling event varies

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

continuously with parameters and that each world has arbitrarily close worlds in which an arbitrary, finite set of parameters relaxed. In fact, under these weak, topological conditions, a general U-turn argument shows that: *if a convergent method probably produces the wrong answer in a given world w satisfying model H at sample size n, then for each answer H′ that frees* $k \geq 0$ *parameters in H, there exists a world near to w in H′ in which the method retracts at least* $k + 1$ *times*.[6]

Furthermore, there exists, in sufficiently well-behaved problems, a strategy whose worst-case retraction bounds over answers in the overall problem cannot be improved, so that Ockham violations in the simplest worlds in the overall problem lead to dominance in the full problem. For a simple illustration, suppose that you have a joint normal distribution over two random variables X, Y, and the question is which coordinates of the joint mean vector are zero. In this case, a natural, *ad hoc* method is to adopt nice tests for the individual mean components with suitably corrected low significance levels so that the chance that neither test rejects in world (0, 0) is high. Tune the significance level of each test downward as sample size increases by a slow schedule, so that power rises monotonically at each world in which the test's null hypothesis is false. On a given sample, perform both tests and free the parameter of each rejecting test. This method converges in probability to the right answer and retracts in probability at most k times in a world with k nonzero mean components. The approach generalizes naturally to higher dimensions[7], So under the assumptions jointly required for both of the preceding results, *if you produce a needlessly complex answer in a simplest world w in the overall problem, you are dominated in worst-case retractions over the set of answers w*, without encountering the measure-theoretic difficulties encountered in the definition of prior probabilities over models and parameters in an infinite space of models.

The extension of Ockham's razor to the remaining, less-than-simplest, worlds requires a statistical surrogate of the concept of compatibility with experience. In the marble example, consistency with experience is sharp and objective. In statistics, the statistical experience afforded 'in probability' by a statistical world at a sample size is just the sampling distribution determined by the world at that sample size, and sampling distributions usually vary continuously with parameters. Therefore, statistical consistency with experience is bound to be a matter of degree. One approach is to define the *statistical distance* between w and w' at sample size n to be the supremum over all measurable sample events E of the absolute difference between the probability assigned to E in w at n and the corresponding probability in w'. Now the set of all worlds *ϵ-compatible with experience* in w at n is just the set of all worlds whose statistical distance from w at n is less than r.

Ockham's razor can now be stated in terms of ϵ-compatibility in the following way: *do not probably produce a false answer in arbitrary world w at n*

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

unless you avoid producing the wrong answer in all the simplest worlds among the worlds ϵ-compatible with w at n. This formulation is equivalent to the usual principle in non-statistical problems, for then the method does the same thing in each world in a subproblem at the time the subproblem is entered. Moreover, the proposed principle absolutely forbids errors in simplest worlds in the overall problem, so it agrees with favoring the simple null hypothesis in a statistical test. Finally, it still implies an asymmetrical bias toward simplicity at less-than-simplest worlds, for probably producing the simple truth in a nearby, simple world excuses error in a complex world, but probably producing the complex truth in a nearby, complex world does not excuse error in a simple world.

Recall that the derivation of Ockham's razor in less-than-simplest worlds required a concept of subproblem in the non-statistical case. Subproblems are again defined in terms of compatibility with experience, and hence must be a matter of degree. Since α is the fixed notion of smallness, say that the *subproblem* entered in w at n is just the set of all worlds α-compatible with w at n. Such subproblems differ from those in non-statistical settings because they can overlap. Indeed, in sufficiently regular settings, continuity implies that *for each pair of distinct worlds w, w' and sample size n, there is a statistical subproblem entered in some world w'' at n that contains w but not w'*.[8]

The possibility of separating any two points with a statistical subproblem breaks the worst-case dominance argument for Ockham's razor in the following way: *every convergent method is dominated in worst-case retractions over answers consistent with the subproblem entered in some world at some sample size*.[9] So there is no question of a method avoiding worst-case retraction dominance in every subproblem. The best one can do is to attempt to minimize, at each sample size, the region of worlds in which the dominance argument applies, without compromising convergence in probability to the truth.

Indeed, Ockham still wins if one is interested in reducing the geometrical volume of worlds in which the method is dominated in worst-case retractions in the subproblem entered at sample size n. Call the set of all such worlds the *dominance argument zone* (DAZ) for the method at sample size n. Recall the simple, bivariate mean problem with known variance. Assume that all possible parameter values lie within a finite square, in order to keep areas finite. To aid visualization further, assume that the bivariate normal distributions are radially symmetric. Then the subproblem entered in w at n is an open disk of radius r centered on w, where r depends only on α and n. Similarly, Ockham's razor holds at sample size n to degree ϵ iff there exists s (depending on n and ϵ) such that for each disk of radius s, if the method probably produces an error at the center of the disk, then it doesn't produce an error in any simplest world in the disk.

To see why Ockham's razor is a good idea, consider a world w at which Ockham's razor fails dramatically (i.e. for a large radius s). If the failure is

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

dramatic enough, then no subproblem (disk of radius r) containing w catches simpler worlds in which the method doesn't probably make a mistake. So the dominance argument (using the appropriate test-based method and the general U-turn construction) works in each such disk that catches w. The midpoints of such disks are contained in an open disk of radius r about w. So a dramatic failure of Ockham's razor at w contributes a full disk of radius r centered on w to the DAZ. Now consider a very slight failure of Ockham's razor. In this case, w is very close to simpler worlds in which the method doesn't probably produce an error and is close to no worlds at least as simple as those in which the method probably produces an error. The dominance argument arises in subproblems that include w but not any of the simpler worlds in which the method avoids error. Hence, the dominance argument arises only in subproblems entered in worlds in a circle of radius r around w but not in any circle of radius r around the nearby worlds in which the method avoids error. So only a thin sliver of the former circle is contributed to the DAZ; the smaller the deviation from Ockham's razor, the thinner the sliver.

All such slivers can be made arbitrarily small, if the method is constructed of tests in the manner described earlier and the significance levels of the tests are dropped very slowly so that power can be maintained at a high level. That explains the importance of power even when the tests are being used in an 'unofficial' way to 'fish' for models. It explains, further, why boldly leaping to the simplest answer with high chance on small samples is not a good idea. Doing so forces more extreme violations of Ockham's razor on the side of failing to reject a statistically 'refuted' theory soon enough and these violations may contribute more to the DAZ than probably producing a complex hypothesis in a simple world. So much of the workaday business of statistics still makes good sense even if your only goal is to minimize retractions in the sense of minimizing the DAZ.

In statistics, Ockham's razor is often viewed as a delicate balance between fit and simplicity. One might suppose that if confirmation theory is needed to explain anything, it is needed here. But the preceding story accounts for the balance entirely in terms of minimizing retractions on the way to the truth, without any circular appeal to a prior simplicity bias. First of all, there is an asymmetrical preference for simpicity, for probably producing a false, complex answer in a simple word will tend to add a fixed area to the DAZ even if there are nearby, complex worlds in which you probably produce the truth, whereas probably producing a false, simple answer in a complex world contributes only a sliver to the DAZ if there is a nearby simple world in which the answer is true. Hence, it is better to moderately err on the side of simplicity. But extreme commitment to simplicity should be avoided. If the chance of producing the simple answer in a simple world is too great at low sample sizes, there will be distant, complex worlds in which you probably produce the simple answer, and their contribution to the DAZ will be as

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

large, or larger, than if you had favored a complex answer in the simple world. In particular problems, the optimal tradeoff may have an analytic solution depending only on the sampling model.

It might be objected that the appeal to areas in the preceding argument reintroduces the Bayesian circle by weighting the relative importance of worlds. But the reverse is true. The Bayesian circle explains our prior bias toward simple worlds by presupposing it. In the preceding argument, each simple theory is a subspace of zero area, whereas in each bounded open set, the area of the set of complex worlds is identical with the area of the set. That hardly counts as a bias toward simple worlds and is best described as a strong prior bias *against* simple models. Nonetheless, Ockham wins in the nuanced manner just described.

13. Ockham's Computational Razor

I began with the promise of a philosophy of science inspired by theoretical computer science and the theory of computability. The ensuing discussion of Ockham's razor may seem to have strayed from its alleged motivation. I close, therefore, by showing how similar arguments apply in the crisp, *a priori* domain of purely computational problems (cf. Kelly, forthcoming).

Suppose you are given the Gödel number of a Turing machine and all that is known in advance is that the machine with that index halts on at most finitely many inputs. The computational problem is to determine how many inputs the Turing machine halts on. No possible, effective method can halt with the right answer to this question, but we can still ask, as in the empirical case, for a procedure that converges to the truth. And there is an obvious one: empirically simulate the machine with the given code number on different inputs for ever longer chunks of time and always guess (in accordance with Ockham) that the only inputs the machine halts on are the ones observed so far to have halted. Eventually, no more computations halt and the procedure converges to the right answer in the characteristically empirical sense that it has no idea when it has found the right answer.

Suppose that the preceding strategy is altered so as to violate Ockham's razor at some stage by guessing more (or fewer) halting computations than the n halting computations it has seen so far when studying some Turing index i_0. Then *the altered procedure M uses one more retraction than the Ockham method in each answer to the question compatible with experience so far*. For each k, construct a partial recursive function with index d_k that is defined by the following procedure. It feeds its own index d_k to M (via Kleene's fixed point theorem) and passes control to the program with index i_0 until M violates Ockham's razor. Then the procedure halts on no further inputs until M retracts to n. Thereafter, it halts on no further inputs until M retracts to $n+1$, etc., up to $n+k$. Hence, M retracts $n+k+1$ times on input

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

d_k, whereas the Ockham strategy retracts at most $n + k$ times on the index of an arbitrary function that halts on exactly $n + k$ inputs.

One may object that both the violator and the Ockham method adopt an empirical approach to the formal problem at hand by electing to simulate the input Turing index to see what it does. Surely some algorithm that performs a formal, *a priori* analysis of the program indexed by an input could reduce retractions. But not so: *the empirical Ockham strategy is computationally retraction-efficient with respect to all possible computational methods*. For let an arbitrary, effective procedure M' that solves the problem in the limit by whatever clever means be given. Using Kleene's fixed-point theorem again, construct a procedure that feeds its own index d_k to M'. The method halts on no inputs until M' outputs 0, halts on just input 0 until M' outputs 1, and so forth, up to k. Since M' is guaranteed to converge to the right answer, and the right answer for d_k will be greater than whatever M' converges to unless it retracts k times, no effective procedure succeeds with fewer than k retractions in in the worst case in answer k.

What about arbitrary methods that violate Ockham's razor? Here the analogy to the empirical case is weaker, since some Turing machines possess 'lookup tables' that produce the right answer to the problem by rote for various inputs. Such a machine could violate Ockham's razor gratuitously for an input on the list without being forced into extra retractions. So, at best, each possible procedure has to respect Ockham's razor on some inputs. Suppose that your procedure violates Ockham's razor on the input d_k constructed for it according to the preceding recipe by starting with an answer other than zero. Then your procedure ends up retracting $k + 1$ times rather than the k times the obvious Ockham strategy would require, so your method is not retraction-efficient even with respect to purely empirical strategies.

Only straightforward computational results and ideas are employed in the preceding arguments, so the proposed account of Ockham's razor does, after all, bear a strong resemblance to computational thinking. That is no accident, for there is an objective structure to efficient truth-finding that transcends the differences between formal and empirical problems. That structure is obscured when truth-finding gives way to confirmation, rationality, and evidential support as fundamental metaphors in the philosophy of science. It is equally obscured by a narrowly naturalistic perspective on methods as short-run truth-indicators, for the road to the truth may twist and turn any number of times in the future.

Notes

[1] For an extended development of this analogy, cf. (Kelly, 1996; forthcoming).

[2] For surveys of the computational learning literature cf. (Osherson et al., 1986) and (Jain et al., 1999). For systematic attempts to connect computational learning theory to the philosophy of science, cf. (Kelly, 1996) and (Martin and Osherson, 1998).

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

[3] The following story builds upon the approach in Schulte (1999). That argument is quite similar, but did not yet extend to problems requiring unbounded retractions for their solution.

[4] For an extended computational study of the inference of conservation laws in particle physics in a paradigm of the sort just described (cf. Schulte, 2001).

[5] I am indebted to one of the referees for urging emphasis on this issue.

[6] Proof. Given the hypothesis, the method must probably produce H in w by some sample size $n' > n$, which is the initial U-turn. So we have the base case for $k = 0$. Inductively, Suppose that your convergent method has retracted $k + 1$ times in some world w_k by sample size n_k, at which the method probably produces the right answer H_k true in w_k. Let H_{k+1} be an arbitrary answer that frees up exactly one parameter in H_k. By the topological assumption, for each sample size $i \leq n_k$, if your method probably (i.e. with chance $> 1 - \alpha$) produces some answer or other at sample size i in w_k, then there exists an open neighborhood S_i of w over which your method also probably produces the same answer. Let S denote the intersection of these open neighborhoods, which is still an open neighborhood of w_n. By the topological assumption, S contains a world w_{k+1} in H_{k+1}. Hence, your method probably performs the same k retractions in w_{k+1} up to n_k and probably produces H_k in w_{k+1} at n_k. Since your method converges to the right answer in w_{k+1}, there exists a sample size $n_{k+1} > n_k$ at which it probably produces H_{k+1}, which is one more retraction. That completes the induction.

[7] E.g., the PC algorithm for causal graph search, recommended by Spirtes et al. (2000) uses conditional independence tests in this way to infer causal connections among variables.

[8] More generally, there will be a subproblem separating an arbitrary closed set from an arbitrary world outside the set.

[9] For let an arbitrary, convergent method be given for a multi-dimensional model selection problem. Hence, the method probably ($> 1 - \alpha$) produces the simplest model H_0 in the simplest world w_0 by some sample size n_0. By continuity, there is also a nearby world w_1 that frees one parameter in which the method probably produces H_0 at n_0. Now choose w', whose subproblem Q at n_0 includes w_1 but not w_0. By the U-turn argument in w_1 and the appropriate, pieced-together test method for Q, the given, convergent method is dominated in worst-case retractions over answers compatible with Q.

References

Buchanan, B. (1974), 'Scientific Theory Formation by Computer', in J. Simon, ed. *Computer Oriented Learning Processes*, Leyden: Noordhoff.

Carnap, R. (1950), *The Logical Foundations of Probability*, Chicago: University of Chicago Press.

DeFinetti, B. (1972), *Probability, Induction and Statistics*, New York: Wiley.

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

Forster, M. and Sober, E. (1994), 'How to Tell When Simpler, More Unified, or Less ad hoc Theories will Provide More Accurate Predictions', *British Journal for the Philosophy of Science* 48, pp. 21–48.

Friedman, M. (1974), 'Explanation and Scientific Understanding', *Journal of Philosophy* 71, pp. 5–19.

Gärdenfors (1988), *Knowledge In Flux: Modeling the Dynamics of Epistemic States*, Cambridge: MIT Press.

Garey, M. and Johnson, D. (1979), *Computers and Intractability*, New York: Freeman.

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

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK

809 Goldman, A. (1986), *Epistemology and Cognition*, Cambridge: Harvard University Press.
810 Goodman, N. (1983), *Fact, Fiction, and Forecast*, 4th edition, Cambridge: Harvard University
811 Press.
812 Harman, G. (1965), 'Inference to the Best Explanation', *Philosophical Review* 74, pp. ■ – ■.
813 Hempel, C.G. (1965), 'Studies in the Logic of Confirmation', in *Aspects of Scientific Expla-*
814 *nation*, New York: The Free Press, pp. 3–51.
815 Hinman, P.G. (1978), *Recursion Theoretic Hierarchies*, New York: Springer.
816 Jain, S., Osherson, D., Royer, J. and Sharma, A. (1999), *Systems that Learn*, 2nd edition,
817 Cambridge: MIT Press.
818 Jeffreys, H. (1985), *Theory of Probability*, 3rd edition, Oxford: Clarendon Press.
819 Kelly, K. (1996), *The Logic of Reliable Inquiry*, New York: Oxford.
820 Kelly, K. (forthcoming), 'Uncomputability: The Problem of Induction Internalized', *Theo-*
821 *retical Computer Science*.
822 Kitcher, P. (1989), 'Explanatory Unification and the Causal Structure of the World', in P.
823 Kitcher and W. Salmon, eds. *Scientific Explanation*, Minneapolis: University of Minne-
824 sota, pp. 410–505.
825 Kuhn, T. (1970), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.
826 Laudan, L. (1980), 'Why Abandon the Logic of Discovery?', in T. Nickles, ed., *Scientific*
827 *Discovery, Logic, and Rationality*, Boston: D. Reidel.
828 Laudan, L. (1996), *Beyond Positivism and Relativism*, Boulder, CO: Westview Press.
829 Martin, E. and Osherson, D. (1998), *Elements of Scientific Inquiry*, Cambridge: MIT Press.
830 Mitchell, T. (1997), *Machine Learning*, New York: McGraw Hill.
831 Nozick, R. (1981), *Philosophical Explanations*, Cambridge: Harvard University Press.
832 Osherson, D., Stob, M. and Weinstein, S. (1986), *Systems that Learn*, Cambridge: MIT Press.
833 Popper, K. (1959), *The Logic of Scientific Discovery*, New York: Harper.
834 Putnam, H. (1965), 'Trial-and-error Predicates and a Solution to a Problem of Mostowki',
835 *Journal of Symbolic Logic* 30, pp. 49–57.
836 Rissannen, J. (1983), 'A Universal Prior for Integers and Estimation by Minimum Description
837 Length', *The Annals of Statistics* 11, pp. 415–431.
838 Rogers, H. (1967), *The Theory of Recursive Functions and Effective Computability*, New York:
839 McGraw-Hill.
840 Rosencrantz, R. (1983), 'Why Glymour is a Bayesian', in John Earman, ed. *Testing Scientific*
841 *Theories*, Minneapolis: University of Minnesota Press.
842 Salmon, W. (1967), *The Foundations of Scientific Inference*, Pittsburgh: University of Pitts-
843 burgh Press.
844 Salmon, W. (1990), 'Tom Kuhn Meets Tom Bayes', in *Scientific Theories*, Minneapolis:
845 University of Minnesota Press, pp. 175–204.
846 Schulte, O. (1999), 'Means-Ends Epistemology', *The British Journal for the Philosophy of*
847 *Science* 5, pp. 1–31.
848 Schulte, O. (2001), 'Inferring Conservation Laws in Particle Physics: A Case Study in the
849 Problem of Induction', *The British Journal for the Philosophy of Science* 51, pp. 771–806.
850 Sklar, L. (1977), *Space, Time, and Spacetime*, Berkeley: University of California Press.
851 Solomonoff, R. (1964), 'A Formal Theory of Inductive Inference, Part 1', *Information and*
852 *Control* 7, pp. 1–22.
853 Spirtes, P., Glymour, C. and Scheines, R. (2000), *Causation, Prediction, and Search*, Cam-
854 bridge: MIT Press.
855 VanFraassen, B. (1980), *The Scientific Image*, Oxford: Clarendon Press.
856 Wasserman, L. (2000), 'Bayesian Model Selection and Model Averaging', *Journal of Mathe-*
857 *matical Psychology* 44, pp. 92–107.

Journal : MIND SPS Article No. : DO00000708 Dispatch : 28-6-2004 Pages : 21
PIPS No. : D000000708 ☐ LE ☐ TYPESET
MS Code : MIND MAM-4 ☑ CP ☑ DISK