

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

3 KEVIN T. KELLY

4 *College of Humanities & Social Sciences, Department of Philosophy, Carnegie Mellon*
5 *University, Baker Hall 135, Pittsburgh, PA 15213, USA; E-mail: kk3n@andrew.cmu.edu*

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

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

12 **1. Introduction**

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

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



Minds and Machine 00: 1–21, 2004.

© 2004 Kluwer Academic Publishers. Printed in the Netherlands.

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

40 2. Confirmation Theory

41 Scientific laws and theories have consequences that outrun the available
 42 evidence, so the question of their justification naturally arises. A familiar
 43 response to this skeptical challenge is confirmation theory: the view that
 44 evidence partially justifies full belief (or fully justifies partial belief). The
 45 focus, then, is to 'explicate' the underlying concept of 'confirmation', which
 46 amounts to finding a simple, general rule or definition that more or less
 47 agrees with our intuitions regarding evidential support in a wide range of
 48 examples (e.g. Carnap, 1950; Hempel, 1965; DeFinetti, 1972; Glymour,
 49 1980).

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

65 Confirmation theory has its appeal. The concept of confirmation seems to
 66 respond to Hume's challenge for a justification of fallible belief. It also cuts a
 67 scientific theory loose from the history by which it was obtained, for the
 68 theory is justified by its current confirmation, regardless how it was discov-
 69 ered. Finally, confirmation theory neatly divides labor between pure phi-
 70 losophers of science, who speculate about the meaning of empirical
 71 justification without regard for computational efficiency, and practical
 72 computer engineers, who seek efficient procedures for checking the confir-
 73 mation of particular theories or for finding maximally confirmed theories.

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

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

83 3. Empirical Naturalism

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

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

107 4. Computational Naturalism

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

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

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

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

151 5. Counting Retractions

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

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

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

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

191 Computer scientists who study retractions are usually interested in clas-
 192 sifying problems – formal and empirical – according to the number of
 193 retractions required to solve them. My aim is a bit different: it is to address
 194 the 'short-run arbitrariness' objection to convergent methodology by show-
 195 ing that key intuitions regarding short-run theory preference follow deduc-
 196 tively from the fact that the method in question minimizes retractions *en*

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. Rissanen,
 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

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 $C = C_0 \vee C_i \vee \dots \vee C_n$. 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
 240 is 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,

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

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.

270 6.3. INSUFFICIENT ARGUMENT

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

279 7. How Ockham Helps

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

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

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

310 8. Counting Things

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

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

332 The U-turn argument for Ockham's razor is almost a retelling of the
 333 freeway story. Suppose you converge to the true answer in each possible
 334 world, but you disregard Ockham's advice after receiving finite sequence of
 335 experience σ including n observed marbles by concluding some number of
 336 marbles other than n . Then σ is compatible with the world w_n that presents σ
 337 followed by uniformly marble-free experience. On pain of failing to converge
 338 to the truth in w_n , you must retract, eventually, to ' n marbles' – say, by the
 339 time finite sequence of experience τ_n extending σ has been received. That is
 340 your extra, initial U-turn, for had you listened to Ockham, you would never
 341 have retracted after σ in w_n . Now there exists a world w_{n+1} compatible with σ
 342 in which one marble is presented after σ . Since you converge to the truth, you
 343 eventually retract ' n marbles' in favor of ' $n + 1$ marbles', say by the end of
 344 τ_{n+1} extending σ . The argument continues in this manner forever, so we have
 345 constructed an infinite sequence of worlds w_n, w_{n+1}, \dots such that you retract
 346 $k + 1$ times after σ in w_{n+k} , which satisfies ' $n + k$ marbles'. The obvious
 347 Ockham method that simply counts the marbles as they appear, on the other

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

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

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

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

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

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

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

417 9. Freeing Parameters

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

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

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

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

443 10. Bayesian Retractions

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

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

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

473 11. Ockham's Statistical Razor

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

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

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

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

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

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

549 12. Ockham and Statistical Model Selection

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

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

559 Furthermore, there exists, in sufficiently well-behaved problems, a strategy
 560 whose worst-case retraction bounds over answers in the overall problem
 561 cannot be improved, so that Ockham violations in the simplest worlds in the
 562 overall problem lead to dominance in the full problem. For a simple illus-
 563 tration, suppose that you have a joint normal distribution over two random
 564 variables X, Y , and the question is which coordinates of the joint mean vector
 565 are zero. In this case, a natural, *ad hoc* method is to adopt nice tests for the
 566 individual mean components with suitably corrected low significance levels so
 567 that the chance that neither test rejects in world $(0, 0)$ is high. Tune the
 568 significance level of each test downward as sample size increases by a slow
 569 schedule, so that power rises monotonically at each world in which the test's
 570 null hypothesis is false. On a given sample, perform both tests and free the
 571 parameter of each rejecting test. This method converges in probability to the
 572 right answer and retracts in probability at most k times in a world with k
 573 nonzero mean components. The approach generalizes naturally to higher
 574 dimensions⁷. So under the assumptions jointly required for both of the
 575 preceding results, *if you produce a needlessly complex answer in a simplest*
 576 *world w in the overall problem, you are dominated in worst-case retractions*
 577 *over the set of answers w , without encountering the measure-theoretic diffi-*
 578 *culties encountered in the definition of prior probabilities over models and*
 579 *parameters in an infinite space of models.*

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

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

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

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

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

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

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

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

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

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

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

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

693 13. Ockham's Computational Razor

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

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

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

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

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

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

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

758 Notes

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

760 ² For surveys of the computational learning literature cf. (Osherson et al., 1986) and (Jain
761 et al., 1999). For systematic attempts to connect computational learning theory to the phi-
762 losophy of science, cf. (Kelly, 1996) and (Martin and Osherson, 1998).

- 763 ³ The following story builds upon the approach in Schulte (1999). That argument is quite
 764 similar, but did not yet extend to problems requiring unbounded retractions for their solution.
 765 ⁴ For an extended computational study of the inference of conservation laws in particle
 766 physics in a paradigm of the sort just described (cf. Schulte, 2001).
 767 ⁵ I am indebted to one of the referees for urging emphasis on this issue.
 768 ⁶ Proof. Given the hypothesis, the method must probably produce H in w by some sample size
 769 $n' > n$, which is the initial U-turn. So we have the base case for $k = 0$. Inductively, Suppose
 770 that your convergent method has retracted $k + 1$ times in some world w_k by sample size n_k , at
 771 which the method probably produces the right answer H_k true in w_k . Let H_{k+1} be an arbitrary
 772 answer that frees up exactly one parameter in H_k . By the topological assumption, for each
 773 sample size $i \leq n_k$, if your method probably (i.e. with chance $> 1 - \alpha$) produces some answer
 774 or other at sample size i in w_k , then there exists an open neighborhood S_i of w over which your
 775 method also probably produces the same answer. Let S denote the intersection of these open
 776 neighborhoods, which is still an open neighborhood of w_n . By the topological assumption, S
 777 contains a world w_{k+1} in H_{k+1} . Hence, your method probably performs the same k retractions
 778 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
 779 right answer in w_{k+1} , there exists a sample size $n_{k+1} > n_k$ at which it probably produces H_{k+1} ,
 780 which is one more retraction. That completes the induction.
 781 ⁷ E.g., the PC algorithm for causal graph search, recommended by Spirtes et al. (2000) uses
 782 conditional independence tests in this way to infer causal connections among variables.
 783 ⁸ More generally, there will be a subproblem separating an arbitrary closed set from an
 784 arbitrary world outside the set.
 785 ⁹ For let an arbitrary, convergent method be given for a multi-dimensional model selection
 786 problem. Hence, the method probably ($> 1 - \alpha$) produces the simplest model H_0 in the sim-
 787 plest world w_0 by some sample size n_0 . By continuity, there is also a nearby world w_1 that frees
 788 one parameter in which the method probably produces H_0 at n_0 . Now choose w' , whose
 789 subproblem Q at n_0 includes w_1 but not w_0 . By the U-turn argument in w_1 and the appropriate,
 790 pieced-together test method for Q , the given, convergent method is dominated in worst-case
 791 retractions over answers compatible with Q .

792 References

- 793 Buchanan, B. (1974), 'Scientific Theory Formation by Computer', in J. Simon, ed. *Computer*
 794 *Oriented Learning Processes*, Leyden: Noordhoff.
 795 Carnap, R. (1950), *The Logical Foundations of Probability*, Chicago: University of Chicago
 796 Press.
 797 DeFinetti, B. (1972), *Probability, Induction and Statistics*, New York: Wiley.
 798 Earman, J. (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*,
 799 Cambridge: MIT Press.
 800 Forster, M. and Sober, E. (1994), 'How to Tell When Simpler, More Unified, or Less ad hoc
 801 Theories will Provide More Accurate Predictions', *British Journal for the Philosophy of*
 802 *Science* 48, pp. 21–48.
 803 Friedman, M. (1974), 'Explanation and Scientific Understanding', *Journal of Philosophy* 71,
 804 pp. 5–19.
 805 Gärdenfors (1988), *Knowledge In Flux: Modeling the Dynamics of Epistemic States*, Cam-
 806 bridge: MIT Press.
 807 Garey, M. and Johnson, D. (1979), *Computers and Intractability*, New York: Freeman.
 808 Glymour, C. (1980), *Theory and Evidence*, Princeton: Princeton University Press.

- 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 Rissanen, 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.