Learning Theory and the Philosophy of Science

Kevin T. Kelly; Oliver Schulte; Cory Juhl

Philosophy of Science, Vol. 64, No. 2 (Jun., 1997), 245-267.

Stable URL:
http://links.jstor.org/sici?sici=0031-8248%28199706%2964%3A2%3C245%3ALTATPO%3E2.0.CO%3B2-0

Learning Theory and the Philosophy of Science*

Kevin T. Kelly†‡
Department of Philosophy, Carnegie Mellon University

Oliver Schulte
Department of Philosophy, University of Alberta

Cory Juhl
Department of Philosophy, University of Texas, Austin

This paper places formal learning theory in a broader philosophical context and provides a glimpse of what the philosophy of induction looks like from a learning-theoretic point of view. Formal learning theory is compared with other standard approaches to the philosophy of induction. Thereafter, we present some results and examples indicating its unique character and philosophical interest, with special attention to its unified perspective on inductive uncertainty and uncomputability.

1. Introduction. Epistemology begins with the problem of induction, the observation that drawing conclusions beyond the available evidence entails some possibility of error. The philosophy of induction has developed four basic responses. (I) We can seek "justification" by obeying rules motivated by considerations other than finding the truth and avoiding error (e.g., conformity with practice or intuition). (II) We can neglect possibilities of error if there aren't "too many" or if they are all too "remote." (III) Even significant possibilities of error are

*Received January 1997.

†We are indebted to Clark Glymour and Teddy Seidenfeld for substantial comments on earlier drafts.

‡Send reprint requests to the senior author, Department of Philosophy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; e-mail kk3n@andrew.cmu.edu

Philosophy of Science, 64 (June 1997) pp. 245–267. 0031-8248/97/6402-0003$2.00

forgivable if we do the best we can. And finally, (IV) there may be a way to avoid error in the limit.

Strategy (I) is characteristic of philosophical work on confirmation (e.g., Hempel 1966, Glymour 1980), moralizing historical studies of the sort now standard in the history and philosophy of science, purely axiomatic approaches to induction such as belief revision theory (Gärdenfors 1988), and Dutch book arguments for coherence and conditionalization (cf. Hellman's contribution to this symposium).

Strategy (II) is adopted by probabilistic (Goldman 1986) and counterfactual (Nozick 1981) accounts of reliability. The former dismiss "small" sets of mistakes and the latter dismiss "remote" mistakes. It is also a crucial feature of probabilistic convergence theorems (e.g., classical "consistency" results and Bayesian "almost sure" learning results: cf. Hellman's contribution to this symposium), which also illustrate strategy (IV). More recently, strategy (II) is the basis of a development in computer science known as PAC (Probably Approximately Correct) learning theory, which mixes computational complexity considerations with short-run convergence bounds derived from a random sampling setting (Kearns and Vazirani 1994).

Strategy (III) is exemplified by decision theoretic approaches to induction in which the scientist is viewed as choosing among a range of possible new beliefs carrying "epistemic utilities" that reflect both truth and content (Levi 1983). The scientist is then exonerated for error so long as she chooses the belief that carries the highest expected utility at the moment it is chosen. Another example is Neyman-Pearson hypothesis testing, since the probability of erroneous acceptance of the null hypothesis is minimized but is not necessarily small (cf. Mayo's contribution to this symposium).

In its pure form, strategy (IV) is a thematic feature of the philosophies of Peirce and Popper. It is the basis of Reichenbach's frequentist vindication of induction (Kelly 1991) and it is still paid perfunctory homage by frequentist statisticians. In the early 1960s, Hilary Putnam (1963) and E. M. Gold (1967) independently recognized and exploited deep affinities between strategy (IV) and the theory of computability. Since then, the approach has been developed by cognitive and computer scientists under the somewhat misleading rubric of "formal learning theory" (cf. Osherson et al. 1986). "Logical reliability theory" is a more accurate name, since the basic idea is to find methods that succeed in *every* possible world in a given range. We will use the terms interchangably.

Formal learning theory is very simple in outline. An inductive problem specifies a range of epistemically possible worlds over which to succeed and determines what sort of output would be *correct,* where correctness may embody both content and truth (or some analogous virtue like empirical adequacy). Each possible world produces an input stream which the inductive method processes sequentially, generating its own output stream, which may terminate (ending with a mark indicating this fact) or go on forever. A notion of success specifies how the method should converge to a correct output in each possible world. A method solves the problem (in a given sense) just in case the method succeeds (in the appropriate sense) in each of the possible worlds specified by the problem. We say that such a method is reliable. Of two non-solutions, one is *as* reliable as the other just in case it succeeds in all the worlds the other one succeeds in. That's all there is to it!

Of course, that's not really *all* there is to it. The interesting part begins when various senses of solution are defined and we begin to ask an increasingly general sequence of questions. What problem does a given method solve? Does a particular problem have a solution? What is the strictest sense in which it is solvable? Do the solvable problems all share a certain structure? Are there *complete* methods that solve all solvable problems? Do problems of one type reduce to problems of another type? What notions of solution are equivalent? How do the answers to these questions change when inductive methods are required to be computable or are subject to further cognitive restrictions? And so on.

The purpose of this paper is to place learning theory in a broader philosophical context and to provide a glimpse of what the philosophy of induction looks like from a logical reliabilist's point of view. In the next section we compare it to some other standard approaches. Thereafter, we present some results and applications indicating its unique character and philosophical interest.

2. Some Comparisons. Formal learning theory shares with options (II) and (III) the idea that a fundamental aim of inquiry is to find nontrivial truth or something like it. In this respect it differs sharply from confirmation theories, belief revision theories, and other methodologies motivated along the lines of (I). Such approaches strike us as both evasive and conservative: evasive because they do not make contact with the aims of finding truth and avoiding error and conservative because they elevate current intuitions and practices into regulative ideals. On our view, the rationality restrictions philosophers impose on inquiry should serve as a cog in a process reliably directed toward the truth, or at least not stand in the way. An important learning theoretic project is therefore to determine whether a proposed methodological norm prevents inquiry from being as reliable as it could have been. For example, one can ask whether every solvable inductive problem can be solved by some Bayesian updating agent. One may ask the same question concerning computable or otherwise bounded agents. Often, requirements that seem to make sense for the ideal agents of epistemology (e.g., maintaining consistency between theory and evidence) stand in the way of success for computable agents and may even preclude the reliability of agents whose cognitive abilities exceed those of computers to an infinite degree (cf. Section 6).

When learning theoretic analysis shows that there is not even a limiting solution to a given problem, it is attractive to adopt strategy (III), trying at least to maximize the reliability of the method one selects. An interesting question is whether there even exists a maximally reliable method for an unsolvable problem. Typically, there is an infinite ascending chain of ever more reliable non-solutions (Kelly 1996). Another learning theoretic application of strategy (III) is to answer the standard concern that limiting reliability is compatible with any silly behavior in the short run. This may be true, but not if we request, quite naturally, that the method converge to the truth as fast as possible and with the fewest number of troublesome retractions prior to convergence,without compromising its reliability (Juhl 1994, Kelly 1996, Schulte 1997).

A more ambitious, "myopic" appeal to strategy (III) drops strategy (IV) altogether and recommends always choosing the best *belief* one can choose, rather than choosing a reliable method or strategy and sticking with it (Levi 1983).[1] These local decisions to believe are viewed as maximizing expected epistemic utility, a quantity reflecting both the truth value and the content of a candidate belief. Since the expected epistemic utility of weakening one's current beliefs cannot exceed that of keeping them (content is lost and no serious possibilities of error are eliminated from one's current point of view), a different story is required to motivate retractions of belief.

Myopic methodology is analogous to learning theory in its emphasis on truth and content as explicit aims of scientific method. But whereas myopic methodology focuses on individual decisions to believe, learning theory reflects on the overall structure of our dispositions to respond to evidence with new theories. From this broader perspective, myopic inquiry involves much more than the aim of finding the truth. Along with the explicitly represented epistemic utilities, there are personal probabilities, confirmational commitments for how to maintain these probabilities, and rules of hypothetical reasoning employed in the account of belief retraction. The learning theoretic question is whether this whole assemblage of values and principles of rationality comprises a reliable strategy for approaching the truth. Every component of in-

1. The account of scientific progress offered in Kitcher 1993 is also myopic in this sense.

ductive practice, including the alleged principles of rationality, is judged according to its contribution to the global aim of approaching the truth, rather than by appeals to practice or intuition along the lines of option (I). As has already been mentioned, insistence on such principles can actually restrict the overall reliability of inquiry even under very mild restrictions on the cognitive abilities of the scientist.

Learning theory is similar to classical statistics in its emphasis on procedures and their reliability. When chance hypotheses are rigorously assumed to entail limiting relative frequencies of outcomes, then statistical inference problems are directly subject to learning theoretic analysis (cf. Section 4). But there are also important differences. When chances are held not to entail limiting constraints on the input sequence, learning theoretic analysis simply says that the inference problem is hopeless, since such hypotheses are logically consistent with anything happening in the long run. We think this raises as many questions about the practical and empirical relevance of propensities as it does about logical reliability (Kelly 1996). Also, formal learning theory applies when the existence of probabilities is not presupposed. And we do not share the classical statistician's dismay at attaching probabilities to hypotheses, so long as these numbers can be shown to converge to the truth values of the hypotheses they attach to.

Bayesian "almost sure" convergence theorems (strategies II and IV) seem to fit right into our program, and in a sense this is true: the gambit is clearly to raise the probability of success by moving to a limiting notion of success. The only difference is that learning theory focuses on the precise set of possibilities over which a method succeeds rather than on the probability of this set. This might not seem to make much difference, but it makes a great deal of difference. When probability is spread out over all the possible input streams a scientific method might receive in the limit, it ends up spread so thin that the problem of induction gets lost in a set of zero measure (cf. Section 3). Attending only to probabilities does not distinguish cases in which the problem of induction is present from those in which it is absent. We would prefer to know whether the problem of induction is actually present in a problem or is merely being ignored in a set of zero measure. If it is present, then neglecting the possibilities of error it gives rise to may be vindicated by strategy (III), but we should still be aware that we are helping ourselves to such vindication rather than actually solving the inductive problem at hand. And if they are not, then errors should not be excused, since we could have used a more reliable method. What we like about the learning theoretic approach is that neglected possibilities are neglected explicitly rather than under a blanket entitlement that both obscures objective differences in the relative difficulties of inductive problems and forestalls the question whether a more reliable method is available.[2]

With a few exceptions, the philosophy of induction has drawn upon the theory of probability for its inspiration. Global probabilistic coherence at an instant does not fit naturally with the local, stepwise character of computation. This leads to a characteristic dualism between inductive and formal methodology. Idealized, coherentist methodologies are proposed as ideals. Computational agents are enjoined to accommodate them by seeking professional help (Levi 1991), by steadily repairing inconsistent probability judgments (Good 1983) and making local corrections that in some cases can never lead to full coherence, or by insisting on full coherence in a metalanguage and conditioning on formal facts (Garber 1983).

Formal learning theory, on the other hand, draws its inspiration from the theory of computability. Its stepwise, strategic, problem-solving perspective on inquiry meshes seamlessly and symmetrically with computational considerations. For a probabilistic coherentist, seeing a black raven is a completely different matter than seeing the thousandth digit of π emerge from a computer. The former is learning from experience. The latter is either recovery from a bout of irrationality or an instance of "semantic ascent." From the learning theoretic perspective, there is no interesting difference between the two cases, which is as it should be. From our point of view, uncomputability is itself a form of inductive underdetermination. The trouble with the halting problem is that a given program might halt just when our algorithm for deciding the halting problem becomes sure that it never will (Kelly and Schulte 1996). And just as in learning theory, uncomputable problems that cannot be solved in the short run may be solvable in the limit. In learning theoretic analyses of computable inquiry, the trick is often to determine whether limiting solutions to the formal difficulties posed by an inductive problem can be interwoven with limiting solutions to the purely empirical part of the problem to yield a limiting, computational solution to the whole problem. Results concerning computable inquiry are among the most interesting and suggestive learning theory has to offer (cf. Section 6).

3. The Structure of Underdetermination. Reliability and underdeter-

2. In this respect, our critique of Bayesian convergence results goes much farther than that in Hellman's contribution. There are Bayesian motives for criticizing "almost sure" convergence: (i) non-Archimedean decision theories countenancing infinitesimal degrees of belief (Fishburn 1981), (ii) Bayesian "consistency" results (Diaconis and Freedman 1986), and (iii) "robust" convergence among agents who disagree in their placement of zeros.

mination are flip-sides of the same coin: data determine the truth if there is some reliable way to use the data to find it.[3] Hence, there are as many concepts of underdetermination as there are concepts of reliability. A central question of formal learning theory (and we think, of epistemology) is to isolate the structures of these various concepts of underdetermination. In a fully successful answer, the structures isolated should exactly characterize when reliable success is or is not possible. In this section, we answer the question for problems of empirical hypothesis assessment. As a byproduct, we will construct a *complete architecture* or recipe for producing methods from problems that is guaranteed to yield a solution for every solvable problem.

An empirical hypothesis is a hypothesis whose truth depends entirely on the input stream, so we may simply identify possible worlds with their input streams and identify an empirical hypothesis H with the set of possible input streams of which it is true.[4]

A hypothesis assessment method produces real numbers in the unit interval in response to the inputs received so far. It may also output a halting mark, indicating that inquiry has ended. For simplicity, we will consider all possible such methods. The parallel results for computable and finite state methods are presented in Kelly 1996.

To define reliability, we must define convergence to an output. Let s be the method's output stream on some input stream. s converges to v with certainty just in case the halting mark occurs exactly once and all the following outputs are v. s converges to v in the limit just in case there is a stage after which s is constantly v. s converges to v gradually just in case for each nonzero distance from v, there is a stage after which the entries in s eventually remain at least within that distance from v.

Success may also be defined in several ways. Decision requires convergence to the truth value of H, whatever that truth value is. Verification requires convergence to 1 if H is true and anything but convergence to 1 if H is false. Refutation requires convergence to 0 if H is false and anything but convergence to 0 if H is true. We may now entertain such notions of success as decision with certainty, verification in the limit, gradual refutation and so forth.

The empirical proposition corresponding to a finite sequence of in-

3. In his contribution, Hellman insists on a much stronger condition for underdetermination, requiring essentially that every data stream arising from a world making a given hypothesis true could have been produced by a world making the hypothesis false. For us, even failure of the condition *w.o.d.* discussed by Hellman is not necessary for underdetermination.

4. H *supervenes* on the data stream, in the terminology of Hellman's contribution.

puts is the set of all possible input streams extending this sequence.[5] Such a proposition looks like an infinite fan whose handle is the finite input sequence that all members of the fan share in common (Fig. 1).

The structural characterization of verification with certainty is quite simple: a hypothesis is verifiable with certainty just in case it can be represented as the union of a collection of fans (Fig. 2). For example the hypothesis that some raven is white is the (disjoint) union of the set of all fans whose handles have white ravens occurring only in the final position.

The hypotheses verifiable with certainty are the open sets of a topological space.[6] Hypotheses that are refutable with certainty are the complements of hypotheses verifiable with certainty and hence are closed sets. Hypotheses that are decidable with certainty are both open and closed (i.e., clopen).

In this topological space, a limit point of an empirical proposition is an input stream s such that for each position along s there is an input stream in the proposition that agrees with s up to that position. A

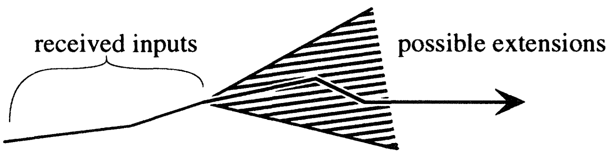

Figure 1. A fan of input streams.

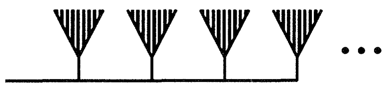

Figure 2. An open set.

5. Some readers may prefer to skip immediately to the examples in Section 4 before proceeding with what follows.

6. The empty hypothesis is verifiable with certainty by a method that always outputs 0. The space of all possible input streams is verifiable with certainty by a method that always outputs 1. To verify with certainty an arbitrary union (disjunction) of verifiable empirical hypotheses, return 1 when the verifier for any hypothesis in the union returns 1 and return 0 otherwise. To verify with certainty a finite intersection (conjunction) of hypotheses, return 1 only when the verifiers for each hypothesis return 1. (When the assumption of perfect memory is dropped, only closure under finite intersection fails.) The resulting space is the usual infinite product space over the set of possible inputs.

boundary point of an empirical proposition is a limit point both of the proposition and of its complement (Fig. 3).

The classical problem of induction arises only on the boundary points of hypotheses. Suppose that a hypothesis contains a boundary point *s*. Then although the hypothesis is correct for *s*, the evidence presented along *s* remains forever consistent with the incorrectness of the hypothesis. This is just the sort of situation that an inductive skeptic's wily demon can exploit. The demon feeds *s* to our hapless method until the method declares certainty in the hypothesis. Then the demon veers off of *s* onto an input stream for which the hypothesis is incorrect. If the method never declares certainty in the hypothesis along *s*, it loses. If it does declare certainty along *s*, then it still loses since it cannot take back its declaration of certainty after the demon veers. Since the demon has a winning strategy in the game of inquiry, the hypothesis is not verifiable with certainty. In fact, a hypothesis is verifiable with certainty just in case it contains none of its boundary points and a hypothesis is refutable with certainty just in case it contains all of its boundary points.

To characterize limiting verification, refutation, and decision, we require topological generalizations of the open, closed and clopen sets, respectively. Say that a set is limiting open just in case it is a countable union of closed sets, limiting closed just in case it is the complement of a limiting open set, and limiting clopen just in case it is both limiting open and limiting closed.[7]

Every limiting open hypothesis is verifiable in the limit. For suppose the hypothesis is limiting open. Then there exists a countable collection of closed sets such that the hypothesis is the union of all the sets in the collection. Place a moveable pointer at the beginning of the enumeration (Fig. 4). As each new input arrives, bump the pointer to the first non-refuted closed set in the enumeration (recall that the closed sets are refutable with certainty). Output 0 if the pointer bumps when the last datum is read; Output 1 otherwise. If the hypothesis is correct, so

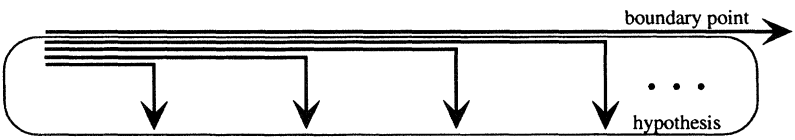

Figure 3. A boundary point in a hypothesis.

7. Officially, the limiting open sets are called Σ^0_2 Borel sets, the limiting closed sets are called Π^0_2 Borel sets, and the limiting clopen sets are called Δ^0_2 Borel sets.

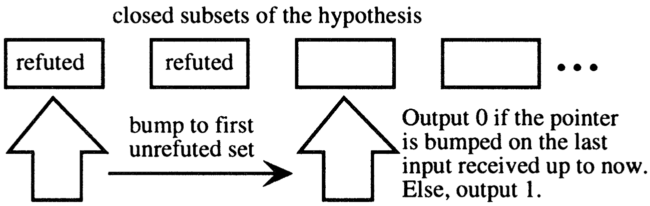

Figure 4. The bumping pointer method.

is some closed subset and the pointer can never bump past it so the method stabilizes to 1, as required. If the hypothesis is incorrect, the pointer is eventually bumped past each closed set in the enumeration, so infinitely many 0s are output, as required.

Conversely, only limiting open hypotheses are verifiable in the limit, so the bumping pointer construction is a complete architecture for limiting verification.[8] Hence, a hypothesis is verifiable in the limit just in case it is limiting open. By similar arguments, a hypothesis is refutable in the limit just in case it is limiting closed and is decidable in the limit just in case it is limiting clopen. As a corollary to our argument, the bumping pointer method is complete in the sense that every hypothesis that is verifiable, (refutable, decidable) in the limit is verifiable (refutable, decidable) in the limit by some implementation of the bumping pointer method.

Limiting decidability and gradual decidability are coextensive in terms of problem solvability,[9] but the latter criterion is more natural for methods like Bayesian updating that treat 1 and 0 as incorrigible marks of certainty. Gradual verification and refutation extend the scope of inquiry beyond their limiting counterparts, however. Say that a set is gradually open just in case it is a countable union of limiting closed sets and is gradually closed just in case it is the complement of a gradually open set. Then it is easily shown that precisely the gradually *closed* sets are gradually verifiable (Kelly 1996). The various characterization results are summarized in Figure 5. The figure illustrates how

8. Suppose a method verifies a hypothesis in the limit. The hypothesis may be represented as the countable union of all empirical propositions of the form "from now on, the method outputs only 1's", which are all refutable with certainty (wait for the method to produce an output other than 1). Hence, the hypothesis is limiting open. Oddly enough, this fact cannot be proved in standard set theory by showing that there is a winning strategy for the demon whenever the hypothesis is not limiting open (Juhl 1995, Kelly 1996, Ch. 5).

9. A gradual decider can be converted into a limiting decider: just output 1 when the gradual decider makes an output greater than .5 and output 0 otherwise.

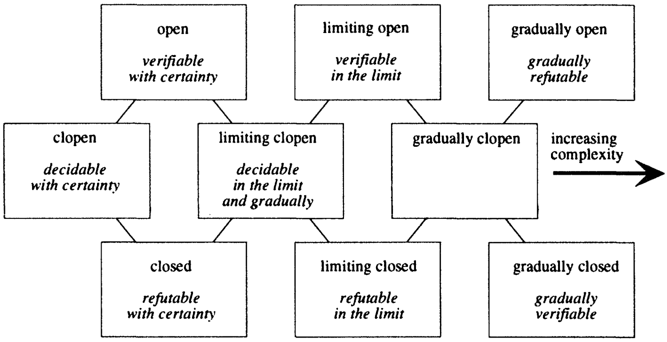

Figure 5. The reliability hierarchy.

the different senses of convergence serve as a precise scale of inductive underdetermination that corresponds exactly to a parallel scale of topological complexity. This scale provides an objective comparison of the intrinsic difficulties of inductive problems arising in very different areas of the philosophy of science, as we shall now illustrate with a few examples.

4. Some Examples. The various levels of topological complexity introduced in the preceding section correspond intuitively to numbers of alternations of quantifiers over the natural numbers. Open sets correspond to existential hypotheses and closed sets correspond to universal hypotheses. Limiting open sets correspond to hypotheses of form $\exists\forall$ and limiting closed sets correspond to hypotheses of form $\forall\exists$. Gradually open sets correspond to hypotheses of form $\exists\forall\exists$, and so forth. The correspondence is exact if it is assumed that the world under study is a countable relational structure for the hypothesis language, each individual in the structure is named by a constant in the language, the input stream consists of a complete enumeration of the quantifier-free sentences true of the structure, and each such enumeration is possible (Kelly and Glymour 1990). If any of these assumptions is dropped, then the logical form of the hypothesis may not correspond to its topological complexity and it is the topological complexity rather than syntactic form that ultimately determines the sense in which the hypothesis can be reliably investigated. Still, the rough and idealized correspondence can be useful as a first approximation when dealing with examples.

Kant's second antinomy of pure reason concerns the thesis that matter is infinitely divisible. Kant claimed that this hypothesis goes "beyond all possible experience." In fact, the hypothesis is refutable but not verifiable in the limit: *for each* piece of matter *there exists* a sufficiently high energy at which it can be split.[10] Limiting verifiability fails because it may always take much more energy than we expected to split the particles we have encountered so far, so that each time a new and larger "energy desert" is encountered, we are lulled into confidence that we have found the ultimate constituents of matter. In practice, this difficulty does not seem to arise because scientists usually focus on a few competing theories at a time so that it seems as though "eliminative induction" should suffice to determine what the elementary structure of the universe is like. But it is one thing to focus on the available theories and quite another to insist that one of them must be true.

There is something unsettling about conservation laws. When the conserved quantities do not seem to balance in a reaction, it may be because the missing quantities have assumed a new and unexpected form that we were not looking for, as in the case of conservation of energy (Poincare 1952, 166; Feynman 1994, 69). When the equation does seem to balance, unexpected forms of the quantity that overturn the equation may be discovered later. If we admit that we may be surprised by new forms of the quantity at any time, we can express the empirical content of a conservation law for quantity q as follows: *for each* reaction, *there is* a time such that *for each* later time, we discover no more q going into or out of the reaction and the q accounted for as going in exactly balances the q accounted for as going out. By its form, this hypothesis is gradually closed and hence is gradually verifiable. That is a pretty weak sense of success. But we have not introduced any background assumptions yet. It is usually assumed that only finitely much of the quantity q enters into or out of any given reaction. It is also sometimes assumed that q is quantized, in the sense of coming in chunks of some fixed size. Finally, it may sometimes be supposed that the lab never spuriously reports nonexistent amounts of q, even though it may fail for a while to recognize q that is present.

Relative to these assumptions, the conservation law is refutable in the limit, since its empirical content can be redefined as follows: *for each* reaction *for each* time *there is* a later time at which the observed q going into the reaction balances the observed q going out.[11]

10. All of Kant's antinomies have this character. The first antinomy: *for each* time, *there is* an earlier time. The third antinomy: *for each* state, *there is* an earlier state that causes it. The fourth antinomy: *for each* entity, *there is* another entity on which it is contingent.

11. Given the assumptions, the only way the equation can be observed to balance infinitely often without balancing all but finitely often is if the q going in and out continues to rise forever, first balancing and then violating the equation. But since q is quantized, the total quantity would have to rise without bound, so in the limit the total

A central thesis of cognitive psychology is that we are computable. Without further background assumptions, the thesis is verifiable but not refutable in the limit. The problem, intuitively, is that each finite sequence of behavior of a system is consistent with some computer program that "memorizes" exactly that sequence in a lookup table. So this problem has the same complexity as the complements of the Kantian and the conservation examples. The result is the same if we add the plausible background assumption that humans have finitely bounded memories but no particular bound is assumed. An assumed upper bound on memory size would make the problem decidable with certainty.

Thomas Kuhn (1970) proposed that research is organized around a "paradigm" that can be "articulated" in various different directions. The function of a paradigm is to generate puzzles and to admit of articulations that resolve them. This "historicist" proposal seems to have little to do with limiting reliabilism, but in fact, it provides a number of interesting applications of our approach. Learning theoretic analysis applies not only to truth but to any virtue of theories or paradigms that goes beyond the available evidence. One such virtue of a paradigm is that it have *some* possible articulation that is capable of handling *all* future problems. This "meta-hypothesis" about the paradigm is verifiable in the limit. To verify this hypothesis in the limit, one must eventually reject it if it is false, even though there may be no intersubjectively compelling reason forcing the rejection at any particular time. This spontaneous rejection in the absence of any local reason is one of the main features of scientific revolutions stressed by Kuhn.[12]

The preceding account makes the completion of science the goal of science, for convergence would occur when no further articulations of the paradigm are required. Kuhn remarks that such static paradigms (e.g., geometrical optics) die even more surely than paradigms in an interesting state of crisis: nobody gets famous solving textbook exercises. What is desired, on this view, is not an anomaly-proof articulation of a paradigm, but a heuristically exciting paradigm that never stops generating interesting problems that it can later absorb. In other words, *for each* problem, there should *exist* an articulation that solves this problem along with most of the preceding ones.[13] This meta-

q involved in the reaction would be infinite. Feynman (1994, 69) emphasizes the epistemic role of quantization in the investigation of conservation laws.

12. We do not intend to take issue here with those who have attempted to provide rational reconstructions of revolutionary episodes: e.g., Kitcher 1993.

13. There remains a question whether the limit of the articulations is itself an articulation or an infinite sequence of patches that never fit together. Both proposals are interesting.

hypothesis is refutable in the limit but may fail to be decidable in the limit.[14] An even more "cynical" proposal is that it does not really matter whether the paradigm is capable of solving a problem if one's colleagues are so pessimistic that they dump the paradigm before the solution is found. This sort of thinking suggests that the paradigm must actually yield problem solutions in sufficient time to prevent the paradigm from being dropped prematurely. This bound on the search time for finding the solution would make the Kuhnian meta-hypothesis refutable with certainty. One important advantage of the strategic perspective of limiting reliabilism over confirmation theory is that standards of problem solution are widely held to be paradigm-relative, whereas our various notions of paradigm correctness adopt a strategic perspective that takes the relativity into account without begging questions about commensurability between problem solution standards across paradigms.[15]

The frequentist interpretation of probability assigns a determinate empirical proposition to each objective probability claim, and hence embeds probabilistic questions into the limiting reliabilist framework we have presented.[16] If frequentists were ever to admit the possibility that no limiting relative frequency exists for a given outcome, hypotheses asserting point probabilities would be gradually verifiable but not gradually refutable. But as a matter of fact, classical statistical procedures usually proceed under the assumption that probabilities exist, else their probabilistic error bounds would be undefined. Given this assumption, the empirical hypothesis that a certain outcome occurs with a given limiting relative frequency is refutable (but not verifiable) in the limit.[17] Surely that cannot be! Aren't statistical hypotheses refuted in statistical tests all the time? Not on the limiting relative fre-

14. This analysis assumes that problem solution can be recognized with certainty. If it is conceded that a solution can be overturned as a rounding error, faulty calculation, or experimental artifact, then we have to add an extra universal quantifier to the effect that the solution will never be overturned and the problem is no longer even gradually refutable.

15. The same point can be made concerning theory laden data and theoretical meaning variance: cf. (Kelly and Glymour 1992; Kelly, Juhl and Glymour 1994; Kelly 1996, Ch. 15).

16. For our purposes, this is the crucial feature of the view, so we will not distinguish frequentism from versions of the propensity interpretation that take propensities to logically entail limiting relative frequencies (e.g., Howson and Urbach 1990).

17. Interval hypotheses are often of interest in statistics. Absent any background assumptions, the hypothesis that the limiting relative frequency exists in a given interval (open or closed) is only gradually verifiable. Given that the limiting relative frequency exists, its membership in an open interval is verifiable in the limit and its membership in a closed interval is refutable in the limit (Kelly 1996, Ch. 3).

quency account: the test is merely guaranteed to have a bounded limiting relative frequency of erroneous rejections (and, preferably, a minimal probability of erroneous acceptances when it is repeated infinitely often). Then at least the procedure will eventually correct itself in the sense that if we were to repeat the test a few times we would see that the rejection was mistaken?[18] Again, not on the limiting relative frequency account: all that follows is that there is an even lower limiting relative frequency for proportions of rejections in runs of some fixed length *n*. But *this* probability statement just repeats the ∀∃∀ quantifier pattern. What is required to improve the sense of convergence is to actually bound the existential quantifier by saying which finite runs of outcomes are logically inconsistent with one's background beliefs and the statistical hypothesis in question. But most probabilists are unwilling to provide any such bound: the sample might simply have been very unlucky. Either way, the limiting reliability analysis yields a precise result: if we are willing to bound the existential quantifier in the ∀∃∀ pattern, then frequentist probability statements are refutable with certainty. If we are not, they are only refutable in the limit (even assuming that a limit exists).

In each of these examples, the background assumptions reduce the topological complexity of the hypothesis in question so as to improve the sense of reliable convergence attainable. If we are on the right track, we should expect scientists to exercise their ingenuity to make their questions appear to have the lowest possible topological complexities. Newton's investigation of universal gravitation illustrates this pattern. In the *Principia,* the three laws of motion are assumed as relatively stable background information. Kepler's laws are taken to summarize the astronomical data. One of the crowning achievements of Newton's *Principia* is Cor. I, Prop. XLV, of Book I. This result states that given the laws of motion and given that the force on a planet is centripetally directed and varies from the source by a power law, the inverse square law is uniquely characterized by the absence of orbital precession. So if the inverse square law were false, orbital precession would eventually be observed. Newton's geometrical analysis amounts to the construction of a method for refuting the inverse square law with certainty, assuming the laws of motion and that the force law is a power law.[19] Newton's careful exploitation of geometrical structure to exhibit a re-

18. ". . . if we are wrong to reject H_0 . . . we would find we were rarely able to get so statistically significant a result to recur, and in this way we would discover our original error" (Mayo 1996, 427).

19. Newton had to first subtract off an estimate of the precession due to interplanetary attractions.

liable connection between theory and evidence helps to explain his insistence that he "deduced" the theory from observation rather than merely "feigning" hypotheses.[20]

As the above examples illustrate, the interesting applications of limiting reliability analysis are those in which the hypothesis determines an empirical proposition relative to background assumptions and the problem involves unbounded quantifiers leading to unbounded searches in the input stream. When such topological complexity arises, it leads to a kind of methodological unease, as in the case of conservation laws and hypotheses that can always be protected with new auxiliaries to account for new anomalies. However, in many concrete cases it is hard to find such unbounded quantifiers. Sometimes this is because there is a great deal of relevant background theory and strong evidence. Possibilities may also be ignored due to the more questionable practice of assuming that one of the theories under consideration must be true. And complexity can come to be ignored through convention or habit. Physicists seem to have some skeptical worries about high energies and very small scales, but they exhibit little concern with the projection of regularities across unexceptional regions of spacetime. Similarly, packaging the three quantifiers of frequentist probability in the notion of a statistical test has fostered the impression that probabilistic hypotheses can really be refuted. Since the probabilities often arise at the "outside" of an analysis (e.g., as a theory of error tacked onto a deterministic theory) they end up being routinely ignored, focusing attention on quantifier alternations in the deterministic hypotheses themselves.

5. Some Remarks on Bayesian Updating. A great deal of philosophical attention has been devoted to "pure" Bayesianism, the attempt to account for scientific intuitions and practice in terms of personal probabilities and updating by conditionalization, without recourse to utilities (e.g., Earman 1992, Howson and Urbach 1990, and Hellman's contribution). With some of our cards on the table, we would like to relate our position to this one in a bit more detail.

We have mentioned that, unlike classical statisticians, we have no special objection to the assignment of probabilities to hypotheses or to the introduction of personal prior probabilities. In fact, the bumping pointer method described in Section 3, which is motivated as a complete inductive architecture rather than by any coherence considerations, bears an intriguing, qualitative resemblance to Bayesian updat-

20. Cf. Harper and DiSalle 1986 for a treatment of the case from the perspective of alternative (I).

ing. The enumeration of closed sets serves as a kind of "plausibility ranking" of different "ways" in which the hypothesis might be correct. Like the Bayesian's prior probability distribution, the plausibility ordering is normatively arbitrary but determines which hypotheses are taken more seriously by the method. The bumping of the pointer corresponds roughly to the redistribution of prior probability mass over the remaining hypotheses when a closed set of data streams is refuted. Reliability considerations alone therefore lead to a qualitative factoring of inductive practice into an arbitrary a priori part and a data-driven part.[21]

Some of the victories claimed for Bayesian updating require only this much apparatus to explain. One allegedly Bayesian virtue of varied evidence is that it knocks out more various rival hypotheses, thereby building up more mass on the true hypothesis (Earman 1992, Hellman's contribution). In the bumping pointer construction the story is similar: more varied evidence (for a particular implementation of the method) can knock out more closed sets in the enumeration, possibly speeding convergence of the pointer.

There are other Bayesian success stories that we do not count as such. For example, it is claimed that envisaging Bayesian updating as a scientific method solves Duhem's problem because some assignments of numbers favor the hypothesis and others favor the auxiliaries (Dorling 1979, Howson and Urbach 1990, Earman 1992). It will come as no surprise that we expect a proper solution to Duhem's problem to reliably identify the false hypotheses. The problem may be solvable in the short run if some patience and careful experimentation decisively identify the offending hypotheses (cf. Mayo's contribution). To correctly assign blame to the false hypotheses of a refuted theory in the limit requires only that each hypothesis in the refuted theory be individually decidable in the limit.

Folklore informs us that so long as alternative hypotheses receive nonzero prior probabilities, the initial, prior probabilities will be "washed out" by the evidence through the agency of the changing likelihoods. Helman's paper discusses a result by Gaifman and Snir (1982). In our setting, the following theorem is more immediately relevant:

For each empirical hypothesis (for which degrees of belief are de-

21. The idea of eliminative induction has been enjoying a recent resurgence in popularity (Earman 1992, Kitcher 1993, Hawthorne 1993). The bumping pointer method may be thought of as a genereralization of the eliminative inductivist idea. A standard objection to eliminative inductivism is that knocking out finitely many theories still leaves infinitely many at each stage. Our response is that knocking out finitely many possibilities at each stage can lead to nontrivial convergent success.

fined), each countably additive Bayesian agent with sufficiently comprehensive degrees of belief[22] must believe with unit probability that updating her current degrees of belief by Bayesian conditioning will gradually decide the hypothesis (Halmos 74, section 49, theorem B).[23]

This sort of result contrasts sharply with the hierarchy of underdetermination developed in the preceding section, which shows that gradual decision is possible only for limiting clopen sets. It cannot be that Bayesian updating has special powers learning theory overlooks: it is one of the methods whose powers are bounded by our characterization results. What is really going on is that the theorem grants background assumptions of unit probability to the Bayesian, and the hypothesis is decidable in the limit (and hence limiting clopen) with respect to these assumptions. With respect to the same assumptions, other methods would converge to the truth as well: the bumping pointer method for example. So the real moral of such results is that arbitrarily severe skeptical arguments (i.e., arbitrarily high topological complexity) can fit into an arbitrarily "small" set, if "smallness" is judged by a countably additive probability measure assigning probabilities to a sufficiently broad range of possible hypotheses. This suggests an intriguing mathematical explanation of the perennial debate between realists and anti-realists concerning the underdetermination of theory by evidence. Anti-realists look at an inductive problem and see high levels of topological complexity. Realists look at the same problem and see that the topological complexity is confined to a "small" set of possibilities. Both observations may be true.

One restriction on the probability measures considered in the result is countable additivity (the probability of a countable union of disjoint hypotheses is the sum of the probabilities). In fact, this hypothesis is necessary, since one can construct a finitely additive probability measure and a hypothesis that is verifiable in the limit such that *no possible method,* Bayesian updating or otherwise, can gradually decide the hypothesis in the limit with unit probability (Kelly 1996). So when countable additivity is dropped, the skeptical, topological structures focused on by formal learning theorists can give rise to nontrivial probabilities of error. Moreover, a finitely additive Bayesian can be "almost sure" that she will fail to solve a *solvable* inductive problem.[24]

22. i.e., degrees of belief are assigned to each set in the σ-field generated by the open (verifiable) empirical hypotheses.

23. For a detailed discussion of Halmos' result, cf. Schervish et al. 1990.

24. Adapting an example due to DeFinetti (1972, 87) along lines suggested to us by Teddy Seidenfeld, consider a Bayesian who thinks the data are generated by fair coin

But even when the conditions of the "almost sure" convergence theorems are met, learning theoretic considerations are relevant. By looking at the exact range of possible worlds over which Bayesian updating succeeds, we can determine whether a given Bayesian agent who succeeds "almost surely" solves the problem completely, or whether tinkering with the agent's initial distribution could make her more reliable still.[25]

More generally, one may ask whether every hypothesis that is gradually decidable in the limit is gradually decided by some Bayesian agent in the learning theoretic sense. The answer is affirmative,[26] but success may require significant tinkering with one's joint initial distribution. This should come as welcome news to those Bayesian confirmation theorists who would prefer to adopt stronger rationality constraints on the Bayesian's initial distribution (e.g., Maher 1996).

The result that every solvable problem is solvable by a Bayesian updater grants to the Bayesian the considerable idealization of logical omniscience. The situation changes markedly when this idealization is dropped in favor of a uniform approach to induction and computability, as will now be seen.

6. Computable Inquiry. We have emphasized the analogy between empirical reliability and computability. This analogy facilitates a smooth transition between the computational and empirical aspects of inductive problems. Recall that the hypotheses ideally verifiable with certainty are precisely those that can be expressed as a union of fans. A hypothesis is effectively verifiable with certainty just in case it can be

---

flips up to some stage n, where n is randomly sampled from a finitely additive distribution in which each particular number is picked with probability zero. Thereafter, the result of the last fair flip is repeated forever. The hypothesis is that the data stream will eventually stabilize to 1. On each input stream, this Bayesian fails ever to update its prior probability (.5) on the hypothesis. But the obvious method that conjectures 1 when the last datum seen is 1 and 0 otherwise decides the hypothesis in the limit given the background assumption that the input stream converges to 0 or to 1 (the Bayesian assigns unit probability to this assumption). By similar means, one can construct a finitely additive Bayesian who fails to "almost surely" gradually decide a hypothesis that is refutable with certainty (Kelly 1996). A decision theoretic Bayesian with utilities on convergent success would prefer our recommendations to Bayesian updating in these cases.

25. The philosophy of science has tended to ignore these considerations entirely. Diaconis and Freedman (1986) raise issues of Bayesian unreliability in a classical statistical setting. Such results may be thought of as lying half way between Bayesian almost sure convergence results and learning theoretic analyses.

26. This claim is conjectured by Earman 1992. Versions of it are proved by Juhl 1993, 1997.

expressed as a union of fans whose handles form a set that is mechanically listable by a Turing machine. Such hypotheses are said to be recursively enumerable or r.e. Mechanical listability corresponds to formal verifiability: given an item to verify for set membership, crank out the enumeration of the set and wait for it to appear. Hence, computable empirical verifiability may be factored into two analogous parts: empirical verifiability in light of the input stream and formal verifiability in light of an internally generated enumeration of the set of possible finite input sequences that would logically verify the hypothesis under study. Formal learning theory provides a symmetrical treatment of the two enumerations. Just as there is no guarantee when the data relevant to the empirical verification of the hypothesis will appear, there is no guarantee when the formal verification *that* the inputs verify the hypothesis will appear. In fact, the formal verification may arrive in the "internal", formal enumeration much later than the stage at which the inputs empirically verify the hypothesis. The same may be said of refutation. It may, therefore, be possible for computable methods to be reliable verifiers or refuters even when it is *not* possible for them to satisfy the norms of logical consistency and entailment. There are some who would say that the computable agents should try harder to meet their deductive commitments (Levi 1991, 46). Our view is that "commitments" to logical omniscience make about as much sense as commitments to empirical clairvoyance. If we are going to take the scientist's bounded perspective on the universe seriously enough to propose specific norms governing empirical uncertainty, we ought to take a computable agent's bounded perspective on infinite formal enumerations equally seriously.[27] But that may mean trading some logical consistency for inductive reliability. Coherence and consistency are not unimpeachable arbiters of inductive rationality.

These reflections suggest a fundamental question illustrating the power of the unified, learning theoretic approach to formal and empirical problems: how uncomputable can the predictions entailed by a hypothesis be if the hypothesis is to be reliably tested by a computable method? Philosophical tradition suggests that the predictions ought to be formally derivable from the hypothesis: deduction works first; then the derived predictions are compared against the input data and the hypothesis is rejected if a mismatch is detected. But is derivability of

27. Some may object that the idealizing assumptions of Turing computability (e.g., unbounded memory) are equally unrealistic (Kitcher 1993, 66). In fact, learning theoretic analysis can be extended to examine what machines with bounded memory can determine in the limit. For a characterization of the limiting reliability of finite state automata, cf. Kelly 1996, Ch. 8.

the predictions necessary? Or does tradition inadvertently restrict the full power of computable inquiry by insisting on this division of labor between formal and empirical reasoning? The answer is quite striking. One can construct a hypothesis that is computably refutable with certainty but whose predictions are in a precise sense *infinitely* impossible to derive.[28] Whatever the mechanical test in this example does, it cannot amount to waiting for formal derivations of predictions and then checking them against the data. That would be infinitely impossible for a computer to do. What the method does is to use future empirical data as a formal "oracle" to help it determine whether past empirical data have already refuted the hypothesis; and this appeal to future data is unavoidable for any agent of even an infinitely non-computable sort (hyperarithmetically definable). In fact, it can be shown that no hyperarithmetically definable method that rejects the hypothesis as soon as it is refuted can even gradually refute or gradually verify the hypothesis! So insistence on the short-run norm of consistency makes even highly idealized agents held to a much weaker standard of success less reliable than mechanical agents who violate the norm.[29] Bayesian agents are obligated to adjust their degrees of belief in refuted hypotheses immediately to 0, so the same result applies to the requirement that updating proceed by conditionalization. Hence, commonly endorsed "rationality" restrictions on inductive methods may undermine rather than contribute to the global reliability of empirical inquiry; even for idealized agents endowed with infinitely non-computable powers.[30] In such cases, reliability considerations should be at least as weighty as the rationality constraints philosophers are accustomed to recommend.

28. i.e., they are not hyperarithmetically definable (Kelly and Schulte 1995, 1996; Kelly 1996). A similar result in a different setting is presented in Gaifman and Snir 1982.

29. This result is much stronger than the one in Putnam 1963, reported in Earman 1992. What Putnam showed is that there is an inductive problem no computable method can solve but that some uncomputable method can solve. Carnap's methods fail to solve it because they are computable: hardly a decisive embarassment. We have shown that there is a problem that no hyperarithmetically definable method that rejects a hypothesis as soon as it is refuted can even gradually decide in the limit that can nonetheless be refuted with certainty by a computable method. That is, even infinitely idealized computational powers (hyperarithmetical definability) and a much weaker standard of success (gradual decidability) cannot compensate for the debilitating effects on reliability of insisting on consistency with the input data. Kelly and Schulte 1995 also answers all other questions of this kind (e.g., how uncomputable can the predictions of a theory be if the theory is to be effectively verifiable in the limit, etc).

30. Osherson et al. (1986) present a range of examples showing how methodological principles can restrict the reliability of computable agents. For an alternative restrictiveness result for Bayesians, cf. Osherson and Weinstein 1988.

7. Conclusion. For reasons of brevity, we have had to neglect some of our favorite topics, including learning theoretic approaches to prediction, the logic of discovery, the problem of new ideas, optimal reliability, finite state methodology, experiment and causal inference, meaning variance, theory-laden data, belief revision theory (Osherson and Martin 1997, Kelly et al. 1996), the minimization of convergence time, the minimization of retractions, the inference of quantum mechanical conservation laws (Schulte 1997), and Goodman's problem (Schulte 1997).[31] Nonetheless, we hope to have conveyed something of the novelty, power, and interest of the approach; particularly concerning its detailed attention to the structure of inductive underdetermination, its potential for criticizing short run "rationality" constraints that restrict reliability, and its unified, symmetrical treatment of computability and empirical considerations.

REFERENCES

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

Diaconis, P. and D. Freedman (1986), "On the Consistency of Bayes Estimates", *Annals of Statistics* 14: 1–26.

Dorling, J. (1979), "Bayesian Personalism, the Methodology of Scientific Research Programs, and Duhem's Problem", *Studies in the History and Philosophy of Science* 10: 177–187.

Earman, J. (1992), *Bayes or Bust?* Cambridge, MA: MIT Press.

Feynman, R. (1994), *The Character of Physical Law.* Cambridge, MA: MIT Press.

Fishburn, P. C. (1981), "Subjective Expected Utility: A Review of Normative Theories", *Theory and Decision* 13: 139–199.

Gaifman, H. and M. Snir (1982), "Probabilities over Rich Languages, Testing and Randomness", *Journal of Symbolic Logic* 47: 495–548.

Garber, D. (1983), "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory", in J. Earman (ed.), *Testing Scientific Theories: Minnesota Studies in the Philosophy of Science,* 10. Minneapolis: University of Minnesota Press, pp. 99–133.

Gärdenfors, P. (1988), *Knowledge in Flux.* Cambridge, MA: MIT Press.

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

Gold, E. M. (1967), "Language Identification in the Limit", *Information and Control* 10: 447–474.

Goldman, A. (1986), *Epistemology and Cognition.* Cambridge, MA: Harvard University Press.

Good, I. J. (1983), *Good Thinking.* Minneapolis: University of Minnesota Press.

Halmos, P. (1974), *Measure Theory.* New York: Springer.

Harper, W. and R. DiSalle (1986), "Inferences from Phenomena in Gravitational Physics", in L. Darden (ed.), *Proceedings of the 1996 Biennial Meeting of the Philosophy of Science Association.* Chicago: University of Chicago Press, pp. S46–S54.

Hawthorne, J. (1993), "Bayesian Induction *is* Eliminative Induction", Unpublished manuscript.

Hempel, C. G. (1966), *Philosophy of Natural Science.* New York: Prentice Hall.

Howson, C. and P. Urbach (1990), *Scientific Reasoning: The Bayesian Approach.* LaSalle: Open Court.

Juhl, C. (1993), "Bayesianism and Reliable Scientific Inquiry", *Philosophy of Science* 60: 302–319.

31. The topics without references are all discussed with further citations in Kelly 1996.

———. (1994), "The Speed-Optimality of Reichenbach's Straight Rule of Induction", *British Journal for the Philosophy of Science* 45: 857–863.
———. (1995), "Is Gold-Putnam Diagonalization Complete?", *Journal of Philosophical Logic* 24: 117–138.
———. (1997), "Objectively Reliable Subjective Probabilities", forthcoming in *Synthese.*
Kearns, M. and U. Vazirani (1994), *An Introduction to Computational Learning Theory.* Cambridge, MA: MIT Press.
Kelly, K. (1991), "Reichenbach, Induction and Discovery", *Erkenntnis* 35: 123–149.
———. (1996), *The Logic of Reliable Inquiry.* New York: Oxford University Press.
Kelly, K. and C. Glymour (1990), "Theory Discovery from Data with Mixed Quantifiers", *Journal of Philosophical Logic* 19: 1–33.
———. (1992), "Inductive Inference from Theory-Laden Data", *Journal of Philosophical Logic* 21: 391–444.
Kelly, K., C. Juhl, and C. Glymour (1994), "Reliability, Realism and Relativism", in P. Clark and B. Hale (eds.), *Reading Putnam.* London: Blackwell, pp. 98–161.
Kelly, K. and O. Schulte (1996), "The Computable Testability of Theories Making Uncomputable Predictions", *Erkenntnis* 42: 29–66.
———. (1996), "Church's Thesis and Hume's Problem", forthcoming in *Logic, Methodology, and Philosophy of Science X,* Dordrecht: Kluwer.
Kelly, K., O. Schulte, and V. Hendricks (1996), "Reliable Belief Revision", forthcoming in *Logic, Methodology, and Philosophy of Science X,* Dordrecht: Kluwer.
Kitcher, P. (1993), *The Advancement of Science.* New York: Oxford.
Kuhn, T. (1970), *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.
Levi, I. (1983), *The Enterprise of Knowledge.* Cambridge, MA: MIT Press.
———. (1991), *The Fixation of Belief and its Undoing.* Cambridge: Cambridge University Press.
Maher, P. (1996), "Subjective and Objective Confirmation", *Philosophy of Science* 63: 149–174.
Mayo, D. (1996), *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.
Nozick, R. (1981), *Philosophical Explanations.* Cambridge, MA: Harvard University Press.
Osherson, D. and E. Martin (1997), "Scientific Discovery Based on Belief Revision", *Journal of Symbolic Logic,* forthcoming.
Osherson, D., M. Stob, and S. Weinstein (1986), *Systems that Learn.* Cambridge, MA: MIT Press.
Osherson, D. and S. Weinstein (1988), "Mechanical Learners Pay a Price for Bayesianism", *Journal of Symbolic Logic* 53: 1245–1252.
Poincare, H. (1952), *Science and Hypothesis.* New York: Dover.
Putnam, H. (1963), "Degree of Confirmation and Inductive Logic", in A. Schilpp (ed.), *The Philosophy of Rudolph Carnap.* LaSalle: Open Court, pp. 761–783.
Schervish, M., T. Seidenfeld, and J. Kadane (1990), "An Approach to Consensus and Certainty with Increasing Evidence", *Journal of Statistical Planning and Inference* 25: 401–414.
Schulte, O. (1996), *Hard Choices in Scientific Inquiry.* Doctoral thesis, Department of Philosophy, Carnegie Mellon University.