Naturalism Logicized

Kevin T. Kelly
Department of Philosophy
Carnegie Mellon University

September 24, 2001

1 Introduction

The approach to scientific methodology developed in my recent book *The Logic of Reliable Inquiry* (LRI) shares many general features with that summarized in Larry Laudan's concurrently published collection of papers *Beyond Positivism and Relativism* (BPR). Nonetheless, this fact might not be apparent, as my own work emphasizes mathematical theorems, whereas Laudan's draws primarily upon historiography. It is, therefore, of some interest to discuss the extent of the agreement and the significance of the differences. More generally, the discussion will (I) provide a logical analysis of the instrumental significance of empirical meta-methodology and (II) redefine the role of logic in a post-positivistic, naturalized approach to epistemology and scientific method.

2 Normative Naturalism

First, some important points of agreement. (1) We both view methodological principles as hypothetical imperatives (i.e., methods are recommended as means to an end) (BPR pp. 132-33, LRI p. 3). (2) We both identify an empirical component in these hypothetical imperatives (BPR p. 133, LRI p. 5). (3) We agree that hypothetically normative epistemology is consistent with naturalized epistemology (BPR p. 133). (4) We agree that aims can be criticized for being unachievable (BPR p. 77, LRI pp. 158-160, p. 190). (5) We agree that methodological norms should in some sense explain scientific progress (BPR pp. 138-39). (6) We agree that contemporary norms need not be satisfied by exemplary historical practice. Laudan's apt term for the position just sketched is *normative naturalism.* So far as this description goes, I am also a normative naturalist.

Our agreement does not end there. (7) We agree that the historicist attack on normative epistemology is founded, to some extent, on persistent positivistic dogmas, (8) we both question the normative force of methodological intuitions (BPR pp. 137-38), and (9) we agree that progress is not necessarily a matter of accumulating information.

Broad agreement on normative naturalism leaves considerable room for fundamental differences in emphasis, however. Whether a rule advances or inhibits our interests depends on such substantive matters of fact as the circumstances in which it is applied, our ability to follow it correctly, the quality of the input, and so forth. But there is evidently a structural dimension as well, for the form of a methodological rule, like that of a computer program, has a great deal to do with what it does and, hence, with its success or failure in promoting our ends.

Laudan emphasizes the empirical dimension of means-ends claims. Given this emphasis, Laudan's guiding metaphor for naturalized epistemology is Baconian empirical science. Instead of deductively unpacking the formal structures of methodological rules prior to consulting experience, he treats the rules like black boxes and recommends that we empirically estimate the chance of success of the rule by consulting the results of historical practice.

I prefer to emphasize the analogy between methodological rules and computational procedures. My guiding metaphor is not Baconian inquiry, but theoretical computer science. Computer scientists are, after all, in the business of recommending rules and procedures based on their ability to achieve desirable goals. Of course, the means-ends relations investigated in algorithm analysis are to some extent empirical: the algorithm cannot be applied beyond its appropriate domain of application, the software has to be installed correctly, it has to be free of mistakes in its code, and so forth. But the explanatory core of such a recommendation is, nonetheless, an *a priori* analysis of what a rule with a given formal structure *would* do in various possible circumstances if it were correctly followed. In fact, this approach better reflects genuine practice in mature empirical sciences. Newton's genius was to fully unpack the geometry of orbital motion prior to consulting experience, so that, for example, null precession over the centuries provided an extremely accurate estimation of inverse square centripetal attraction. I propose that the theory of computability and computational complexity can serve to focus and to organize naturalistic methodology in much the way that geometry organized mechanics.

The *a priori* version of normative naturalism that I have just described is not new. It has been developed over the past forty years to a level of some sophistication by computation theorists under the heading of "formal learning theory".[1] The name of the subject is perhaps misleading, until one realizes that for computer scientists, "learning" is a matter of reliable convergence to a correct answer to an empirical question, so that a theory of learning is actually a general theory of the existence of feasible, reliable,

[1] The basic idea of providing a computational analysis of the problem of finding the truth was independently proposed by Hilary Putnam (1963) and E. M. Gold (1965). For book length presentations, cf. (LRI, Osherson et al. 1986, and Martin and Osherson 1998).

empirical methods.

3 Laudan's Program

In this section, I review Laudan's position in some detail, marking the points at which we differ. The discussion follows the outline of Laudan's programmatic paper "Progress or Rationality? The Prospects for Normative Naturalism" (BPR, pp. 125-141).

Laudan introduces normative naturalism as a response to recently popular nihilistic views about scientific method. According to Laudan, this nihilism arises from two assumptions. (1) Most great scientists have chosen rationally among alternative theories and (2) a methodology of science is an account of unconditional or categorical rationality. It follows that an account of scientific method must be satisfied by the practice of most great scientists.

Laudan rejects (2), responding that scientific rationality depends on the scientist's methodological aims and on her current beliefs about which acts are likely to further those aims. Our methodology should reflect our own aims and beliefs rather than those of historical figures. Laudan then distinguishes methodological "soundness" from "rationality". Presumably, a "sound" method really promotes our goals, whereas rationality reflects an individual's beliefs about what would further her own goals.

While I agree with Laudan in rejecting (2), I don't think this maneuver responds effectively to methodological nihilism. For example, Feyerabend's nihilism requires neither (1) nor (2). Rather, it is based on a straightforward means-ends argument with respect to an aim that seems plausible in the present day, namely, progress.

> We find ... that there is not a single rule, however plausible, and however firmly grounded in epistemology, that is not violated at some time or other. It becomes evident that such violations are not accidental events .... On the contrary, we see that they are necessary for progress (Feyerabend 1975, p. 23).

Since this argument is offered in the spirit of Laudan's empirical normative naturalism, it is hard to see how Laudan's position could respond to it, except by reinterpreting the verdicts of history. I prefer to criticize Feyerabend for claiming to have proved that no general methodological directives exist after discrediting a few proposed examples. In computer science, where impossibility results are routinely proved, pessimism based on the failure of a few, particularly simple, programming attempts is not taken seriously, and properly so. I recommend that naturalized epistemology reform itself in a similar direction.

Moreover, I am not as eager as Laudan and other historicists to trace methodological variation to divergent ends and beliefs. Even for scientists who share goals and beliefs (e.g., finding a correct answer to an empirical question), different scientific problems require very different means for their solution. For example, Bacon's methods of similarity and difference demonstrably lead to the truth when it is assumed in advance that the truth is a conjunction of monadic predicates. When disjunctions of such predicates are relevantly possible, more powerful methods are required (LRI, chapter 12). These strategies are very different from strategies for estimating limiting relative frequencies. Inferring conservation laws in particle physics suggests still other strategies exploiting the richer structure of linear spaces (cf. Schulte 1998, 1999a, 1999b). This is analogous to the situation in computer science. Some formal problems seem to require search, others succumb to recursive "divide and conquer" techniques, and still others are unsolvable unless we weaken our notions of success. If one's aim is to get the right answer as soon as possible, it is hard to see what sorts of interesting algorithmic principles would be suitable independently of the specific type of empirical problem one faces. It would be more plausible to discuss relational methodological principles that depend on the structure of the problem at hand. That is precisely the approach of formal learning theory.

Laudan then sketches normative naturalism, as described above. He first observes that methodological rules like "avoid *ad hoc hypotheses*" are really disguised hypothetical imperatives of the form "if you want to develop theories which are very risky, then you ought to avoid ad hoc hypotheses."[2] Such a conditional is "warranted", according to Laudan, if we "find" that following the recommendation is the best way we have yet thought of to promote the intended aim. Thus, hypothetical imperatives are subject to empirical investigation.

Laudan next addresses the obvious, skeptical charge that empirical justifications of empirical methods are circular. Faced with this problem, other epistemologists have advocated genuinely circular, coherentist epistemologies. Laudan opts for a methodological version of foundationalism in which a single, unobjectionable method is used to justify more sophisticated rules, which are in turn used to justify still more sophisticated rules, and so forth. The rule he chooses is something like maximization of expected (methodological) utility with respect to objective chances of success estimated using the straight rule of induction.[3]

> (R1) If actions of a particular sort, m, have consistently promoted certain cognitive ends, e, in the past, and rival actions, n, have failed to do so, then assume that future actions following the rule "if your aim is e, you ought to do m" are more likely to promote those ends than actions based on the rule "if your aim is e, you ought to do n" (BPR p. 135).

Laudan's proposal bears some resemblance to Hilbert's foundational program in mathematics, for both approaches propose the use of more elementary, uncontroversial means

[2] I think Popper would have more plausibly preferred "if you don't want to end up preserving a false hypothesis for eternity, then you ought to avoid ad hoc hypotheses".

[3] Laudan's position recalls Hans Reichenbach's familiar argument for using the straight rule of induction: if any other method works, the straight rule of induction will eventually lead us to follow that method.

(finitist arithmetic, the straight rule of induction) to vindicate the soundness of more controversial means (the infinitary methods of analysis, sophisticated scientific practice).

> I hasten to add that (R1) is neither a very sophisticated, nor a very interesting, rule for choosing between rival strategies of research. But then, we would be well advised to keep what we are taking for granted to be as rudimentary as possible. After all, the object of a formal theory of methodology is to develop and warrant more complex and more subtle criteria of evidential support (BPR p. 135).

The pivotal notion of "consistently promoting" in the definition of (R1) is vague in a manner that masks difficult questions. What if one method succeeded the only time it was tried, while the other was tried thousands of times with a few failures? Also, what if the current application has a rare feature on which the most successful method always failed and on which an infrequently applied competitor always succeeded? Or even worse, what if the current application has a feature that one can see by computational analysis to defeat the rule even though the method has never been used in such circumstances in the past? So although (R1) is simple, its recommendations are hardly as uncontroversial as Laudan suggests.

Another objection, due to Robert Nola (1999), concerns Laudan's requirement that goal achievement be an observable variable in the historical record. Laudan's proposal to use method (R1) to determine the instrumentality of a method M may work for observable goals such as maintaining consistency with the current data. But it cannot work for such aims as truth, empirical adequacy, or even future problem solving effectiveness because they are not observable in the historical record, and hence cannot generate instances of the kind (R1) requires as input.[4] One might use another inductive method M' to determine whether such an (unobservable) goal G is, in fact, achieved, but then (R1) would not be able to vindicate the instrumentality of M' with respect to the goal of determining whether unobservable goal G is satisfied, and so forth, for chains of any finite length. So there is no way in which to "bootstrap" up from (R1) alone to methods vindicated with respect to aims like truth, empirical adequacy, or problem-solving effectiveness.

Perhaps the most serious objection to Laudan's proposed, meta-methodological program is that for all its emphasis on means and ends, it doesn't explain what would be *achieved* by a chain of meta-methods, each of which oversees the performance of its predecessor. Although his program holds out the hope of replacing intuition mongery with

[4]Laudan seems to miss this point. After dismissing "transcendent" goals like finding the truth as appropriate aims for science, Laudan writes: "My own proposal . . . is that the aim of science is to secure theories with a high problem-solving effectiveness. From this perspective, *science progresses just in case successive theories solve more problems than their predecessors*" (BPR p. 78). In this passage, Laudan plays loosely with modality and tense, both of which are crucial to any discussion of the problem of induction. How many problems a theory actually solved in the past is observable. How many problems it could solve given more ingenuity and time is not. But "effectiveness" concerns the latter, dispositional concept, not the former, empirical one.

objective means-ends relations, this standard of intelligibility is not applied reflexively to his own program.

In spite of these objections, the idea of using one inductive method to empirically justify another as a means to a goal raises interesting logical and epistemological issues when it is presented with sufficient generality and without the encumbrance of Laudan's empiricistic and foundational commitments. In Section 9 below, I employ learning theoretic techniques to establish *a priori* when reliable meta-methodological chains of various kinds are possible and what can be accomplished by them.

Laudan next observes that his naturalistic approach eliminates the need to base methodology on "methodological intuitions". I agree with that, but this feature of naturalism is independent of Laudan's strongly empirical approach to naturalistic methodology. The computationally informed naturalism I advocate is both instrumental and largely *a priori*, appealing not to historical data but to the respective formal structures of the particular empirical problem addressed and of the various methods that might be employed to solve it.

Although Laudan understands the primary aim of methodology to be the empirical justification of methodological rules as means for local, observable ends, he is also interested in explaining scientific progress as the result of repeatedly achieving such ends through time. I prefer a more direct approach, in which progress is viewed as an aim in its own right. Learning theory is directly concerned with such hypothetical imperatives as "if you want to converge to the truth (in a given sense) then use method M."[5] Laudan's conceptual detour through more proximate aims is thereby eliminated.

A key feature of Laudan's position is its emphasis on axiology, or the appropriateness of goals. This emphasis stems from Laudan's desire to rout methodological relativism, for he realizes that viewing methodological norms as hypothetical imperatives opens the door to relativism with respect to goals. Laudan's response is to claim that the appropriateness of scientific ends is itself an objective fact, since (a) appropriate ends must be feasible and (b) appropriate ends must have been reflected in the history of science (BPR pp. 157-58). This gives rise to a "reticulated" account of justification in which changing theories of feasibility lead to changes in aims which lead to changes in methods, which lead to changes in theories, etc. (Laudan 1984 pp. 79-80).

I agree strongly with Laudan's emphasis on feasibility of aims. Feasibility is a matter of problem solvability by agents of a given kind. Some empirical problems are unsolvable even by logically omniscient agents. Others are solvable by logically omniscient agents, but not by computable agents. Still others are solvable by computable agents, but not by any agent with a finite memory store, and so forth. Learning theorists are keenly interested in discerning the general features of empirical problems that make them solvable in one

[5] Its focus on diachronic utilities separates learning theoretic analysis from other a prioristic approaches to normative naturalism, such as Isaac Levi's methodology of maximizing expected true content (Levi 1983).

sense rather than another.

I also agree, to some extent, with Laudan's requirement that aims share some continuity with the past. Such sensitivity to practice is essential if the theory of computability is to yield relevant results. When computer scientists face such ill-defined problems as "planning" or "learning", they cannot begin to apply computability theory until they associate the informal problem with a spectrum of mathematically precise models of what "planning" or "learning" require. It is understood that this extra-theoretical process of explication must reflect, to some degree, actual planning and learning behavior. Actual behavior needn't turn out to be an optimal solution, but it should at least appear to have been directed toward a solution to some mathematically precise problem in the spectrum. It is always open, in a computation theoretic analysis that yields highly counterintuitive results, to question whether the formal problem addressed reflects what people actually want to accomplish.

But practice is not supreme. Computability analysis, by its very nature, forces one to turn a logical microscope on the problem under study, to an extent that intuitive, philosophical, or historical discussions rarely achieve. When practice and analysis disagree, it is possible that theory has unearthed structural possibilities that never would have come to light in the historical record because historical figures didn't notice it either. That is why I oppose Laudan's particular emphasis on history in the philosophy of science, an emphasis which has been the received view in the field for some decades. If the philosophy of science is to earn its keep, it should do more than report back to scientists what they actually do. It should, like science itself, open new and exciting possibilities. History may suggest plausible goals and methods, but these suggestions are merely suggestions.

4 Elements of Learning Theory

Although formal learning theory is sometimes thought to be rather forbidding in detail, it is refreshingly simple in outline.[6] For all the scientist knows (or cares) the actual world may be one of many relevantly possible worlds. Each relevantly possible world responds to the scientist's acts with inputs through time. The scientist is capable of responding to these inputs in different ways. If the scientist's task is to determine whether a given hypothesis is empirically adequate, she may respond to the inputs with successive test outcomes (accept, reject) or with successive assignments of degrees of belief or of confirmation to the hypothesis in question. Any task involving such responses about a given hypothesis will be referred to as a hypothesis *assessment* problem.

In other circumstances, the scientist starts not with a hypothesis but with a *question* to be answered. Some hypotheses will be *relevant* to this question and a question may for our purposes be identified with its potentially relevant answers. An answer to the question

[6]For book length expositions, cf. (LRI, Osherson et al. 1986, Martin and Osherson 1998).

is a potentially relevant answer that is also correct (e.g., true or empirically adequate). Such tasks are called hypothesis *generation* or *discovery* problems.

In either case, the scientist hopes to converge, in some sense, to a correct output; whether it be a correct assessment of a given hypothesis or a correct answer to a given question. Many scientific discoveries have resulted from happy accidents, but methodology is about guaranteed or *reliable* success, meaning success over a "broad" range of relevant possibilities. To summarize, learning theory concerns the ability of a method or strategy to converge to a correct output (test result or relevant hypothesis) over a specified range of relevant possibilities. An *empirical problem* is a specified range of possibilities, together with a hypothesis to assess or a question to answer. Thus, learning theory concerns solutions to and the solvability of various empirical problems.

5 Strategic Goals for Hypothesis Assessment

Much variation is possible within the vague framework just described. The Socratic spirit demands that such vague terms as "relevant possibility", "success", and "convergence" be provided with precise explications at the outset. Learning theory follows a different approach, providing a scale or spectrum of clear interpretations rather than a single one. This leads to a range of different types of scientific goals, each of which having a unique, epistemological character.

For example, consider the case of hypothesis assessment. Very ambitiously, one might hope for a method guaranteed to produce the truth value of the hypothesis by some time established in advance. But such ambitions usually cannot be achieved in science. More leniently, one might hope for a method that eventually halts with the truth value of the hypothesis. This is called *decision with certainty.* Decision with certainty is an empirical analogue of the computational concept of "decidability". But whereas many interesting formal problems are computationally decidable with certainty, the point of the classical problem of inductive generalization is that most general empirical hypotheses are not. At this point, the axiology of feasibility recommends moving to a weaker goal. Popper's original idea was that universal generalizations can nonetheless be refuted with certainty even though they cannot be verified with certainty. Similarly, purely existential hypotheses can be verified with certainty but not refuted with certainty.

Unfortunately for Popper's original idea, most scientific hypotheses are not really refutable with certainty either. Notoriously, a hypothesis can be saved from refutation by tinkering with the rest of the theory. And even in an idealized, empirical setting in which experimental outcomes are unproblematically theory-independent, probability estimates are logically consistent with any data in the short run, even if such an estimate is understood to imply a limiting relative frequency of outcomes in the future data. The same is true of the hypothesis that there are only finitely many types of elementary particles to be discovered, the hypothesis that a system is chaotic as opposed to orderly, and the hypothesis that a given sequence is produced by a Turing machine rather than by some uncomputable process.

Popper's response (1968) was to reconceive falsificationism as an injunction against coddling pet views rather than as a criterion of success. An alternative option is to weaken the criterion of success again, so that certainty is never required, whether the hypothesis is true or false. For example, one might require only that a method stabilize to the state of correctly rejecting or accepting the hypothesis under assessment without necessarily halting or providing a sign that it has done so. It is natural to call this standard *decision in the limit*. As Peirce and James emphasized, limiting convergence, unlike convergence with certainty, allows for an appealingly fallible sense of methodological success, according to which following the method is guaranteed, eventually, to reach a correct answer, but certainty is never forthcoming because there is never any guarantee that the method will not change its conjecture after seeing the next datum. Within the comfortable confines of a viable research paradigm, the possibility of a major crisis is a mere, philosophical curiosity. But from the outside looking in, the history of science is a history of broken certainties and no amount of "inductive support" or other holy incantation can ensure that the same will not happen again. At best, we can hope that inquiry is organized so as to eliminate surprises after some future time that will not be recognized as such.[7]

Hypotheses about limiting relative frequency, computability, or the finite divisibility of matter are not decidable in the limit either. This remains true of limiting relative frequencies even if we assume *a priori* that the limit of the observed frequencies exists. Feasibility demands yet weaker aims. It turns out that a hypothesis specifying a particular value for a limiting relative frequency is *refutable in the limit* given that the limit exists, where *limiting refutation* requires convergence to rejection just in case the hypothesis is true and *limiting verification* requires convergence to acceptance just in case the hypothesis is false. Computability and finite divisibility are verifiable in the limit. Also, if chances are understood to entail limiting relative frequencies, the existence of an "unbiased" statistical test is also equivalent to verifiability or refutability in the limit, depending on whether the "rejection" zone is defined with a strict or a non-strict inequality (LRI, Chapter 4).

Limiting verification and refutation are very weak, in the sense that the vacillations witnessing nonconvergence may come with arbitrary rarity. Surely, we would like to do better. But if there were an *a priori* bound on how long one must wait to see the next vacillation, if it occurs at all, this bound would allow one to construct a limiting decision procedure, which is impossible in the examples mentioned. So again, I agree with Laudan's feasibility condition: if limiting verification is possible and limiting decision is not, then don't demand an upper bound on the frequency of a limiting verifier's rejections when the hypothesis under test is false.

If it is not assumed that a limiting relative frequency exists, a hypothesis asserting

[7] I think of this as the core of truth in Popper's "deductivism". The same sentiment is reflected in Reichenbach's "pragmatic vindication" of the straight rule.

that it exists with a given value is not even verifiable in the limit.[8] This leads, by the axiology of feasibility, to even more attenuated notions of success. *Gradual decision* requires that the real values assigned to the hypothesis by the method approach the truth value of the hypothesis, possibly without ever actually reaching it. Gradual decidability is in fact equivalent to limiting decidability, because a gradual decision procedure can be converted into a limiting decision procedure by means of accepting or rejecting according to whether the gradual method's output exceeds or fails to exceed a cutoff (e.g., 0.5). The one-sided versions of gradual decision are strictly more lenient than their limiting analogues, however.[9] *Gradual verification* requires that the outputs approach unity just in case the hypothesis is true and *gradual refutation* requires that the outputs approach zero just in case the hypothsis is false. In fact, limiting relative frequency is gradually verifiable but not gradually refutable.

6 The Long Run in the Short Run

Limiting success occasions the natural objections (a) that the limit is too long to wait for and (b) that limiting correctness provides insufficient constraints on what to believe in the short run. These objections can be met, to some extent, by requiring that no reliable method converge *as fast* as our method in each relevant possibility and *faster* in some relevant possibility, in which case our method may be said to be *data minimal.* To demand the truth faster than a data-minimal method can provide it is to demand the impossible.

As Kuhn emphasized, it is both practically and cognitively costly to retool when a theory is retracted. Taking this concern seriously, we would prefer reliable methods that not only minimize convergence time, but retractions as well. Note that a single retraction could occur arbitrarily late, so convergence time and number of retractions are two different considerations.

It turns out to be too strict to require that no reliable method performs more retractions in any relevant possibility and fewer in some relevant possibility (Schulte 1999a), for this is only possible when the potential answers to an empirical question are all decidable with certainty, and hence there is no genuine problem of induction. Suppose, however, that there is an *a priori* bound on the number of retractions required prior to convergence. In the case of hypotheses that are refutable with certainty, at most one vacillation is required: start out accepting and then reject when the hypothesis is refuted. The hypothesis that exactly one star of a given mass exists is decidable in the limit with at most two retractions: reject until a star of that mass is encountered and then accept until another one is encountered. When such a bound exists, it is natural to insist on methods

[8] The same is true for any specification of a closed interval of such values.

[9] Thus, a gradual refuter and a gradual verifier cannot always be assembled into a gradual decision procedure (cf. LRI Chapter 4). This contrasts with the limiting case.

that decide the hypothesis in question in the limit, that minimax retractions and that are data minimal with respect to all limiting decision procedures.

In "The Will to Believe", William James (1948) remarked that finding the truth is different from avoiding error and that the two aims are usually in tension. Data minimality suggests the aim of finding the truth, since a method that refuses to conjecture a potential answer to the question at hand could not possibly have converged to the right answer yet, whereas a method that produces a potential answer consistent with the data *might* have already succeeded. Minimizing retractions suggests the aim of avoiding error, since a method that witholds judgment until the evidence is conclusive performs no retractions at all.

Reliability, data-minimality, and minimaxing mind changes can jointly impose strong requirements on methodology in the short run. To illustrate this point, suppose we know either that each stage will be green, or that at some finite stage n, green will give rise to blue forever after, in which case we may say that each stage is "grue(n)". If these are the only relevant possibilities, then the *unique* data-minimal, mind-change-minimaxing, limiting decision procedure is the one that conjectures that all stages are green until seeing a blue outcome (say at stage n), after which the method conjectures forever after that each stage is grue(n).[10] The same result obtains if we consider as relevant possibilities all hyper-grue predicates of the form grue($n_0, n_1, \ldots n_k$), where $k \leq m$, for some fixed m (Schulte 1999b). The predicate grue($n_0, n_1, \ldots n_k$) means green through stage n_0, blue from then through stage n_1, green from then through stage n_3, etc. If the fixed bound m is dropped, then success with bounded retractions is impossible.

One may think of decision with a bounded number of retractions as a criterion of success in its own right (Case and Smith 1983, LRI), where *decision with n retractions* requires that the method decide the hypothesis in the limit, vacillating between acceptance and rejection (or *vice versa*) at most n times. When retractions are being counted, it turns out to matter what one's leading conjecture is. For example, refutation with certainty is equivalent to decision with at most a single retraction, starting with acceptance, for a method that refutes with certainty starts out accepting the hypothesis and then, when trouble is encountered, retracts its former conjecture and replaces it with rejection. Similarly, verification with certainty is equivalent to decision with at most a single retraction, starting with rejection. What about decision with certainty? Since decision implies both verification and refutation, decision with certainty is equivalent to decidability with at most one retraction starting with an *arbitrary* conjecture (either acceptance or rejection). As we allow more retractions, we therefore arrive at generalized notions of verification, refutation, and decision. For example, decidability with at most two retractions starting with acceptance allows the method to begin with acceptance, change its mind thereafter

[10]This method performs at worst one retraction and is data-minimal since whatever it conjectures it possibly converges to, but failing to make a conjecture would fail to be data-minimal and conjecturing any grue(n) prior to the green hypothesis might require two retractions (one from grue(n) to green and another from green to some grue(n')) (Schulte 98).

to rejection, and finally switch back to acceptance. Once all the allowed retractions are used up, the method's output is certain.

7 Learning Theoretic Questions

Laudan's normative naturalism focuses on hypothetical imperatives for particular methodological principles and on the feasibility of particular aims. From a learning theoretic viewpoint, the former question concerns the relation "M solves problem P". The latter concerns the property "problem P is solvable", which is definable as the existence of a method M solving problem P.

Hypothetical imperatives and feasibility axiology work in tandem. Methodological understanding is obtained by formally solving for the strongest aim in the hierarchy of convergent goals that can be satisfied for a given empirical problem. Thus, the hierarchy of goals can be viewed as a kind of classification system for empirical problems. All of the problems within a given classification are in a precise sense "methodologically equivalent", giving rise to intuitively similar methodological difficulties and calling out for similar sorts of solutions.

The ability to formally isolate the strongest aim achievable for a given empirical problem addresses a weakness in Laudan's empirical version of normative naturalism. It is very difficult to show by means of empirical data that no stronger aim *could* have been realized for a given problem. Thus, Laudan adopts an empiricistic stance and asks history only if a *known* method was *observed* to do better. Learning theoretic negative results cover all possible methods, and hence allow one to show that no possible method could have done better. This is precisely the role that computability theory plays in computer science.

Once one has seen a good number of solvability and non-solvability results, one wishes to know if there is an elegant structural characterization of solvability. That is, one desires a purely structural property Φ (making no explicit reference to methods or to success) such that an arbitrary problem P is solvable just in case it has Φ. Such results are called *characterization theorems.* Since they provide necessary and sufficient conditions for the possibility of reliable inquiry, they might be viewed as logically valid *transcendental deductions.*

Characterizations of the concepts of assessment introduced above are easily presented (cf. LRI, Chapter 4). Assume a given set of relevant possibilities to be specified. Assume, also, that the hypothesis is not *globally underdetermined*, in the sense that the same infinite input stream arises from worlds in which it is respectively true and false (else no possible method could find the truth value of the hypothesis in each relevant possibility). Such a hypothesis is verifiable with certainty just in case each relevant possibility satisfying the hypothesis eventually presents inputs whose occurrence entails that the hypothesis is true. A hypothesis is refutable with certainty just in case its complement is verifiable with certainty, and is decidable with certainty just in case it is both verifiable and refutable with certainty. At the next level, a hypothesis is verifiable in the limit just in case it is a countable disjunction of hypotheses that are refutable with certainty. A hypothesis is refutable in the limit just in case its complement is verifiable in the limit, and is decidable in the limit just in case it is both verifiable and refutable in the limit. A hypothesis is gradually verifiable just in case it is a countable conjunction of hypotheses that are verifiable in the limit. It is gradually refutable just in case its complement is gradually verifiable and is gradually decidable just in case it is decidable in the limit. Thus, each notion of reliable success corresponds to a structural recipe for building up all the hypotheses for which that sense of success is achievable.

These results illustrate the grain of truth in the positivists' attempt to relate "cognitive significance" to logical form. *If* hypotheses are expressed in a first-order language and *if* the input stream presents all the true, quantifier-free sentences in the language, and *if* each object is named by some constant in the language, *then* the quantifier prefix of the hypothesis determines the senses in which it can be reliably assessed. Specifically, quantifier-free hypotheses are decidable with certainty, existential hypotheses are verifiable with certainty, universal hypotheses are refutable with certainty, hypotheses with quantifier prefix $\exists\forall$ are verifiable in the limit, hypotheses with quantifier prefix $\forall\exists$ are refutable in the limit, finite, boolean combinations of existential and universal hypotheses are decidable with bounded retractions and hence in the limit, hypotheses with quantifier prefixes of form $\exists\forall\exists$ are gradually refutable and hypotheses with prefixes of form $\forall\exists\forall$ are gradually verifiable. But none of this is a function of logical form *per se*; nor is it a characterization of meaning. It is a *contingent* relationship between logical form and levels of achievable reliability, where the contingency relating the two is an assumption about the kind of data that would arise in a given relevantly possible world.

A simple structural characterization of decision with bounded retractions can also be given (LRI, Chapter 4). Hypotheses that are verifiable with certainty are decidable with one retraction starting with rejection and refutability with certainty characterizes one retraction starting with acceptance. Decision with $n+1$ retractions starting with rejection is possible exactly when the hypothesis under test can be expressed as the disjunction of a verifiable hypothesis with a hypothesis that is decidable with n retractions starting with acceptance. Dually, decision with $n + 1$ retractions starting with acceptance is possible when the denial of the hypothesis can be decided with the same number of retractions, starting with rejection.

8 Historicism Reconsidered

Verification and refutation with certainty can be understood in two very different ways. Refutation with certainty requires that the data logically contradict the hypothesis and verification with certainty requires that the data logically entail the hypothesis. Thus, refutation and verification are logical entailment relations. So when it is discovered that a hypothesis is neither verifiable nor refutable with certainty, one response is to look for a "generalized" entailment relation (degree of confirmation or inductive support) that does hold between the data and the hypothesis.

But as suggested above, refutation and verification with certainty may *also* be viewed as success criteria for empirical methods, just as they are viewed as success criteria for formal methods in mathematical logic and computability theory. The shift in type is important. Success criteria are goals (ends) rather than methods (means). So on this perspective, intuitive or historical arguments for the propriety of a fixed method (generalized entailment relation) give way to objective, computation theoretical arguments about achievability of the various goals. When it is discovered that a hypothesis is neither verifiable nor refutable, it is natural to move to weaker criteria of success that are achievable (e.g., limiting and gradual success).

Limiting goals, such as decision in the limit, are different from verification and refutation with certainty because it is no longer an option to view limiting success criteria as fixed logical relations that determine *when* to reject or to accept a given hypothesis. For example, one limiting method may converge to the truth on a given data stream faster than another such method does, and thereby converge more slowly on some other data stream, so the two methods disagree for some arbitrary length of time on both data streams. The logic of efficient, limiting convergence does not favor one solution over the other, leaving ample room for hunches, predilections, and scientific "bon sense", so long as they do not inhibit the strategic goal of finding a correct answer as soon as possible.

Viewing verification and refutation with certainty as success criteria, rather than as generalized logical relations, leads to a reconception of the debate between historicism and logic. First of all, one argues for a generalized notion of entailment by a process of explication or reflective equilibrium, which is a kind of spiral process of correcting the explication with practice and correcting practice with the explication. This leaves methodology open to the plausible charge that it is merely armchair sociology in logical dress. Since learning theory focuses on objective, computational questions about the solvability of empirical problems, it does not invite this objection.

Portraying scientific method as a fixed, generalized entailment relation also occasions the objection that following such recommendations would have precluded scientific progress when the social character and costs of inquiry are considered (Feyerabend 1975). Such arguments cannot be directed a logical approach to scientific method based on learning theory, because they *are* learning theoretic arguments. In fact, many of the results of learning theory can be viewed as *formally* grounded Feyerabendian critiques of particular methodological proposals (LRI, Osherson, et al. 1986, Martin and Osherson 1998). Such critiques have the form that some empirical problem would have been solvable had the recommendation not been insisted upon. A computational critique of the rule of rejecting a theory when it is logically contradicted by the data will be discussed in detail below.

Finally, the logical relation conception of methodology invited Kuhn's nihilism concerning the logic of scientific change. Kuhn's basic argument in *The Structure of Scientific Revolutions* (1962) is that in major episodes of scientific change, no logical relation or generalization thereof holding between theory and evidence *rationally compels* one to drop the theory *when it is dropped*, so the change is arbitrary. According to Kuhn, the momentous empirical question facing a scientist is not the correctness of a given hypothesis, but the viability of her paradigm. Viability is a vague matter, but it has something to do with the potential of the paradigm to generate puzzles and solutions to them. *Piece-meal* viability means something like: *for each* new anomaly that a competitor can handle at the time, *there exists* an articulation of the paradigm that resolves it. *Uniform* viability is more ambitious, requiring that the paradigm possess some as-yet unknown articution that will once for all absorb all new anomalies handled by competing theories (e.g., the "end of science" foretold by some advocates of the fundamental particle paradigm in physics).[11] Piece-meal viability is of $\forall\exists$ form, and is therefore refutable in the limit, whereas uniform viability is of $\exists\forall$ form and is therefore verifiable in the limit. Barring *a priori* bounds on how long it would take to find such an articulation, neither question is decidable in the limit (cf. Kelly et al. 1997). Recall that in order to verify a hypothesis H in the limit, a method must reject H infinitely often if H is false. So whereas it is not arbitrary that a limiting verifier perform these rejections at some times or other, it is up to the method rather than to logic when, exactly, they occur. Nonetheless, there are still normative recommendations to be made on a logical basis, for some methods will fail even to verify the hypothesis in the limit and others will converge more slowly than necessary. Thus, the absence of an objective compulsion to drop the paradigm at a particular time is explained by, rather than raising a difficulty for, the learning theoretic logic of the paradigm selection problem.

The ultimate, historicist argument is relativism. Relativism is a danger to the generalized logical relation conception of methodology for the obvious reason that others may reject the proposed relation in light of different, culturally informed intuitions. Since there is no further reason for following the relation, the discussion ends there. On the instrumental approach, there can at least be agreement about the possible circumstances in which various methods would work, even when there are differences in aim and in beliefs about what the actual circumstances are.

It might be thought that relativism poses a serious problem for learning theory nonetheless, for how can we converge to the truth, even if we want to, if meaning and truth change in incommensurable ways through time? But there are still intelligible, strategic goals that do not presuppose translatability across scientific revolutions, so long as we do not measure progress in terms of increasing *content* (which requires content comparisons

[11] Kuhn explicitly rejects uniform viability as a goal, since it would reduce the subject to trivial textbook exercises that could not be published. But as a cognitive, rather than a career goal, it would clearly be more desirable than piece-meal viability. For example, Hubert Dreyfus' (1979) objection to the strong A.I. program is that each bit of human behavior can be duplicated by a computer program, but no single program will ever duplicate all of human behavior due to scaling problems.

across incommensurable languages) or *verisimilitude* (which requires a fixed metric defined across incommensurable languages).[12] We might require, for example that science eventually stabilize the truth value of the hypothesis under investigation and then learn what it is. Or we might require, more weakly, that whatever the truth value is in the future, we eventually always know what it is. This raises another Feyerabendian sort of question: would it injure the power of inquiry to require that the truth value eventually stabilize? The answer is affirmative (LRI Chapter 14, Kelly and Glymour 1992), so learning theory provides a *logical* argument in *favor* of inducing incommensurable changes.

The historicist quarrel with logic is actually a quarrel with the "generalized logical relation" approach to methodology.[13] Reconceiving refutation as the first success criterion in a sequence of ever weaker criteria leads to a logical perspective on methodology that embraces the central premises of the historicist position without drawing the nihilistic conclusions.

9 What Empirical Naturalism Can and Cannot Do

The core of Laudan's epistemological program is the idea of using one empirical method to investigate the conditions under which another method will succeed in achieving a given goal. I criticized Laudan for depicting such inquiry as a search for empirical correlations between means and ends, since such ends as achieving empirical adequacy are not directly observable in the historical record. Furthermore, I objected that Laudan's correlational approach leaves no room for *a priori* computational analysis of the conditions under which a method would succeed. Finally, I objected that for all the emphasis on means and ends, Laudan did not say what could be accomplished by methods assessing methods assessing methods. But the general idea is of sufficient interest to warrant a fresh approach.

Suppose we are interested in finding out whether a given method will succeed. This question has several empirical dimensions. If we ignore the structure of the method and treat it as a black box, as Laudan seems to suggest, then it is an empirical question even to determine what the method would direct us to do in a given situation. But if we look at what the method is, and if a precise sense of convergent success is specified, then

[12]Indeed, Miller's (1974) counterexample shows that verisimilitude metrics cannot even be preserved under translation. The moral is that metrical concepts should be avoided in defining progress (Mormann 1988). Learning theoretic success criteria are topological rather than metrical, and hence are not subject to this objection.

[13]It is not easy to pin down contemporary Bayesianism on this issue. Decision theoretic analyses of method, along the lines of Levi (1983), are explicitly instrumental. "Bayesian confirmation theory", based exclusively on the concept of updating by conditionalization, is instrumental insofar as one takes the diachronic Dutch book argument or the "almost sure" convergence theorems seriously. But it is increasingly fashionable not to do so. Many advocates of conditioning view limiting convergence as a useless idealization, while others object that diachronic Dutch book arguments are invalid. Without such arguments, conditioning is recommended as an explication of practice.

in principle[14] it is an *a priori* matter to determine the set of seriously possible future trajectories along which the method succeeds (in the specified sense). Call this set the *presupposition* of the method (relative to the intended sense of success).[15] Once a method's presupposition has been determined *a priori*, the problem of empirical meta-methodology reduces to determining whether the presupposition of the method under investigation is actually true. In what follows, I will assume that methods are transparent to the meta-methods investigating them, so the meta-methods merely assess the presuppostions of the methods they investigate.

The pressing means-ends question raised by empirical, normative naturalism is, then, what one could *do* with methods that check the presuppositions of methods that check the presuppositions of methods.... Could one, for example, by looking only at what the meta-methods do, converge to a correct conjecture about the original hypothesis H? If not, then it is hard to see what the point of all the assessing is supposed to be. In such a case, one might say that the sequence represents a *vicious* empirical regress of the sort condemned by skeptics like Sextus and Hume. But if there is a strategy for assembling the conjectures of the meta-methods in the chain into a single conjecture that converges to the truth, then the chain can be used to achieve a cognitive goal and the regress may be exempted from the charge of pointlessness.

If the converse is also true (i.e., a single method that succeeds in a given sense can be turned into a given kind of meta-methodological chain of methdods), then we may say that the chain is informationally or methodologically *equivalent* to the single method. Methodological equivalence imposes some discipline on our epistemological hopes in much the way that the concept of energy imposed discipline on our hopes for perpetual motion machines. There is no question of a meta-methodological chain allowing us to do the impossible (i.e., to construct a single method that succeeds in an unachievable sense). But a meta-methodological chain could do far *worse* than to be methodologically equivalent to the best sort of solution that a given problem admits, just as a heat engine may fall far short of being perfectly efficient in terms of energy transfer. The degree of viciousness of an epistemic regress may be viewed as the extent to which the sense of success to which the chain is equivalent falls short of the best achievable sense of success. For example, if there exists a certain refutation procedure, then a chain equivalent to a limiting decision procedure is inefficient, but less so than a chain equivalent to a limiting refutation procedure.

For a simple example of a methodological equivalence, consider the following situation. We throw method M_1 at the problem of trying to refute H with certainty given background knowledge K. But M_1 works only when an empirical presupposition P_1 is satisfied. Meta-

[14]In practice, of course, the method may be too difficult to analyze using computation theoretic techniques, as in the mundane case of an ordinary word processing program with thousands of features. But the kinds of methodological principles that come up in philosophical discussions are much more amenable to formal analysis than the average word processor program is.

[15]In LRI and (Osherson et. al. 1986) the presupposition of a method is called its *scope*.

method M_2 is supposed to refute with certainty whether this presupposition is, indeed, satisfied. Meta-method M_2 actually does refute the presupposition P_1 of M_1 with certainty given K. Now suppose that we only observe what the two methods conjecture through time, without looking at the data they receive. What could we tell about H?

Without saying more, not much. For suppose M_1 is a crazy method that alternates forever between acceptance and rejection without ever looking at the data, and suppose that M_2 rejects no matter what, without looking at the data. Then M_1 fails on every data stream in K so P_1 is unsatisfiable. And M_2 is correct on every data stream in concluding that P_1 is false, so M_2 refutes P_1 with certainty given K, as required. But one could conclude nothing about H from watching these two methods, since they both say the same thing no matter what they observe, and therefore erase all of the information in the data. According to the above criterion, such a pair represents a vicious empirical regress.

The situation changes, however, if both methods *aspire* to refute with certainty, in the sense that they outwardly appear to be refuting their respective hypotheses with certainty even if they really aren't. More precisely, say that M_1 aspires to refute H with certainty given K just in case K entails that M_1 starts out accepting and retracts at most once. We may also speak of aspirations to verify with certainty given K, verify in the limit given K, etc. For example, M aspires to decide in the limit given K just in case on each data stream satisfying K, M converges either to acceptance or to rejection.

Now suppose that M_1 aspires to refute H with certainty given K and that meta-method M_2 refutes with certainty given K whether the aspirations of M_1 will actually be realized. Then we can construct a method M that decides H given K with two retractions starting with acceptance that succeeds just by watching what M_1 and M_2 do. Method M may be defined as follows. Make M start out accepting. Thereafter:

1. M accepts when M_1 agrees with M_2.

2. M rejects when M_1 disagrees with M_2.

To see that M works as claimed, suppose that K is satisfied. There are four easy cases to consider:

1. P_1 and H are satisfied: then M_1, M_2 always accept so M always does so as well, with no retractions.

2. P_1 is satisfied but H is not: then M accepts until M_1 rejects and continues to reject thereafter, using one retraction.

3. H is satisfied but P_1 is not: since M_1 is an aspiring refuter, M_1 starts with acceptance and can retract at most once, so since P_1 is not satisfied, it must be that M_1 eventually reverses its initial acceptance to a rejection. Meta-method M_2 correctly reverses its initial acceptance to a rejection as well. So M converges to acceptance after at most two retractions, starting with acceptance.

4. Neither H nor P_1 is satisfied: again, M_1 can only have failed by never reversing its initial acceptance to a rejection. Meta-method M_2 eventually reverses its initial acceptance to a rejection. So M converges to rejection after at most one retraction.

In each case, M converges to the right conjecture about H using at most two retractions, starting with acceptance.

What if M_1 aspires to verify H given K and M_2 verifies the presupposition of M_1 given K? Then exactly the same construction again implies that H is decidable with two retractions starting with acceptance. The situation is similar if we have a refuter of verification or a verifier of refutation. In either of these cases, the result is the same except that M starts with rejection rather than acceptance.

So for aspiring methods, refutation of refutation and verification of verification imply two retraction decidability starting with acceptance and refutation of verification and verification of refutation imply two retraction decidability starting with rejection. Methodological equivalence requires the converse implications as well. Let us consider whether they hold in the present example. Suppose that M decides H given K with at most two retractions starting with acceptance. Can we construct M_1, M_2, P_1, such that M_1 aspires to refute H given K and does so under presupposition P_1 and M_2 refutes P_1 given K? It is up to us to choose the both the presupposition P_1 and the methods M_1 and M_2.

Here is how to do it. Choose P_1 as the (naturalistic, methodological) proposition that M retracts at most once. Let M_1 be the aspiring refuter (given K) that watches M and accepts until M retracts once, rejecting thereafter whatever else M does. Let M_2 start with acceptance and then reject, with certainty, when M retracts for the second time. Evidently, M_2 refutes P_1 with certainty given K. Moreover, M_1 refutes H with certainty under presupposition P_1, because when P_1 is satisfied, M_1 converges correctly to whatever M converges to. If P_1 is not satisfied, then M uses its second retraction and converges to acceptance, but M_1 incorrectly converges to rejection. Thus, M_1 refutes H with certainty if and only if P_1 is satisfied. The converses of the claims for verification of verification, refutation of verification, and verification or refutation are similar. So we have arrived at a simple example of a meta-methodological equivalence theorem:

Proposition 1 *The following situations are methodologically equivalent:*

1. *There are two methods M_1, M_2 such that*

   (a) *M_1 aspires to verify [refute] H with certainty given K and does so under presupposition P_1, and*

   (b) *M_2 refutes [verifies] P_1 with certainty given K.*

2. *H is decidable given K with at most two retractions, starting with rejection.*

The analogous proposition in which M starts with rejection and the aspirations of M_1 and M_2 mismatch is also true.

Let us now generalize the preceding analysis along two dimensions at once. We will move from a single meta-method to an arbitrary, finite chain of meta-methods, each of which second-guesses the presuppositions of its predecssor. And we may as well also allow each method in the chain to succeed under a fixed bound on retractions. When is such an attenuated meta-methodological situation possible? Just when there is a single method that uses the sum of the retractions of all the methods in the chain and whose first conjecture depends in a systematic manner on what the methods in the chain achieve. The exact statement of the equivalence is as follows. For convenience of notation in dealing with chains, let P_0 henceforth denote the original hypothesis H under investigation.

Proposition 2 *The following situations are methodologically equivalent:*

1. *There exists a finite chain $M_1, \ldots, M_k$ of methods such that*
   (a) *for each $i < k$, method M_{i+1} aspires to decide P_i given K with n_{i+1} retractions starting with c_{i+1}, and does so under presupposition P_{i+1}, and*
   (b) *K entails P_k, the presupposition of the final method in the chain.*
2. *There exists a single method M that decides P_0 with $n_1 + \ldots + n_k$ retractions given K, starting with conjecture c, where c is "reject" if an odd number of the c_i are "reject", and c is "accept" otherwise.*

The general proof of this proposition, and of all those that follow, is given in the Appendix. By way of illustration, consider a situation in which M_1 decides H with two retractions starting with acceptance, M_2 decides the presupposition of M_1 with three retractions starting with rejection and M_3 decides the presupposition of M_2 with one retraction starting with acceptance without presuppositions. The result tells us that this is possible exactly when there is a single, presuppositionless method M that uses $2 + 3 + 1 = 6$ retractions. Method M uses at most six retractions because it retracts once each time one of the component methods retracts. Since an odd number of the three methods start out rejecting, so does M.

The preceding analysis provides a clear motivation for empirical meta-methodology. It does not give us something for nothing (nothing could). Rather, adding more empirical meta-methods to the chain amounts to an even epistemological trade in which the sense of success is weakened (more retractions are countenanced) in exchange for weaker methodological presuppositions. Although it does not show up in the statement of the proposition, another such trade-off concerns time to convergence, for it will typically take longer for the single method constructed from the chain to converge than it would have taken M_1 to converge when its narrower presupposition is satisfied. Whether this trade is rational will depend upon the plausibility of the presuppositions and on the costs of retractions and delayed convergence. This is where history, individual psychology, and external cirumstances figure in. Logic presents the possible options and the systematic trade-offs among them.

The result just presented assumes that each meta-method in the chain succeeds with bounded retractions. What could we do with a finite, meta-methodological sequence of limiting decision procedures for which no such bounds exist? Assuming that the methods in the chain are all guaranteed to converge to acceptance or to rejection, we could turn the whole sequence into a single method that also decides the original hypothesis in the limit. One may describe the situation by saying that limiting decidability is preserved or closed under finite meta-methodological regresses.

Proposition 3 *The following situations are methodologically equivalent:*

1. *There exists a finite chain $M_1, \ldots M_k$ of methods such that*
   - (a) *for each $i < k$, method M_{i+1} aspires to decide P_i in the limit given K and does so under presupposition P_{i+1} and*
   - (b) *P_k, the presupposition of the final method in the chain, entails K.*
2. *There exists a single method M that decides P_0 in the limit given K.*

Before, we saw that bounded retraction meta-methodology trades retractions for weaker presuppositions and delayed convergence. The same is true here, except that the increase cannot be measured by a uniform bound as in the bounded retraction case (proposition 2).

10 Empirical Naturalism Without Foundations

In the finite meta-methodological chains considered in the preceding section, the method at the end of the chain serves as an anchor or foundation for the entire chain, since it is required to succeed in every serious possibility. This is reminiscent of Laudan's idea of picking a single method to anchor the process of empirically investigating what other empirical methods can do. But what if there is no foundation for the chain? What if every method in the chain has empirical presuppositions and more methods can always be added, on demand, to assess them? Theln there is nothing to science but assessments of assessements of assessments, without end. That is not to say that scientists ever use infinitely many meta-methods all at once. The relevant infinity is potential rather than actual: in the face of yet another challenge to her reliability, the scientist is disposed to respond with yet another meta-method to test the presuppositions of the method challenged.

The instrumental question, once again, is what one could *do* with a potentially infinite chain of methods, each of which investigates the presupposition of its predecessor. Suppose, then, that there is a (potentially) infinite sequence of meta-methods, each of which, say, refutes the presupposition of its predecessor with certainty under some presupposition. Moreover, suppose that none of the methods in the chain works without empirical presuppositions. Inquiry floats on an infinite abyss of presuppositions.

It is natural to assume that although no method works without presuppositions, the presuppositions tend to get weaker, so that for each $i \geq 1$, P_i entails P_{i+1}. Call such a sequence *increasingly reliable*. In other words, even though there is no "foundational" method, the infinite sequence is nonetheless *directed* in the sense that each successive meta-method is at least as reliable as the method it assesses.

When is an infinite, foundationless, increasingly reliable chain of aspiring refutation meta-methods possible? Whenever H is refutable with certainty by a single, "super-method" that succeeds over the disjunction of all the nested presuppositions. Thus, we can say that refutation with certainty is closed under infinite, increasingly reliable regresses.

Proposition 4 *The following situations are methodologically equivalent:*

1. *There exists an infinite chain $M_1, \ldots M_k, \ldots$ of methods such that*
   (a) *for each $i \geq 0$, method M_{i+1} aspires to refute P_i with certainty given K and does so under presupposition P_{i+1},*
   (b) *for each $i \geq 0$, P_i entails P_{i+1}, and*
   (c) *K entails $(P_1 \vee \ldots \vee P_n \vee \ldots)$.*
2. *H is refutable with certainty given K.*

A result of this kind is double-edged. On the one hand, an infinite, meta-methodological chain of refuting methods is not pointless, since it is equivalent to a single refutation method that has weaker presuppositions than any method in the chain. But in another sense it may seem pointless, since we could have used the equivalent, single method to begin with! There is, however, increasing interest in the philosophy of science these days in "local" or "piece-meal" methodology (e.g., Mayo 1996). The preceding result says that adding more and more "local" refuting methods when challenged can add up to performance methodologically equivalent to having a single refuting method, without committing one's self to a single method handling all possible contingencies from the outset. Since scientists do not really commit themselves to fixed methods for eterminty, the applicability of learning theoretic analysis is thereby greatly enhanced.

What could we do with an infinite, nested, sequence of verification methods? One might well expect a similar closure result, to the effect that an infinite, increasingly reliable chain of verifiers adds up to a verifier. But this is far from being the case, for an infinite, meta-methodological chain of certain verifiers is equivalent to an infinite chain of *limiting refutation* methods! It also turns out that an infinite chain of limiting refutation methods is equivalent to a single limiting refutation method, so limiting refutation, like certain refutation, is closed under infinite, increasingly reliable meta-methodological chains. Since the power of limiting decision lies between that of certain verification and of limiting refutation, it should come as little surprise that infinite chains of limiting deciders are also equivalent to having a single limiting refuter. Thus, limiting refutatability is a fairly robust necessary condition for the existence of increasingly reliable meta-methodological regresses. These results cannot be improved to equivalence with an infinite, increasingly reliable chain of limiting verifiers, since even a single limiting verifier can succeed on hypotheses that are not refutable in the limit (LRI).

Proposition 5 *The following situations are methodologically equivalent:*

1. *There exists an infinite chain* $M_1, \ldots M_k, \ldots$ *of methods such that*
   - (a) *for each* $i \geq 0$*, method* M_{i+1} *aspires to verify* P_i *with certainty given* K *and does so under presupposition* P_{i+1}*,*
   - (b) *for each* $i \geq 0$*,* P_i *entails* P_{i+1}*, and*
   - (c) K *entails* $(P_1 \vee \ldots \vee P_n \vee \ldots)$.
2. *Situation 1, with decision in the limit replacing refutation in the limit.*
3. *Situation 1, with certain verification replacing refutation in the limit.*
4. P_0 *is refutable in the limit given* K*.*

Corollary: in conditions (1-3), P_i *can be chosen to be of form* $H \vee R_i$*, where* R_i *is refutable with certainty given* K*, so all of the* M_i *can be chosen to converge to the truth given that* H *is true.*

The preceding result assumes that $P_0 = H$ entails P_1, so that M_1 converges to acceptance when H is true. If this condition is dropped, (1) and (2) become equivalent both to the existence of a limiting refuter of a certain verifier and to the existence of a certain verifier of a limiting refutation procedure. These "mixed" chains do not collapse into anything more elementary, and may be viewed as criteria of success in their own right.[16]

Infinite, nested chains of certain refuters are equivalent to certain refutability and infinite, nested chains of certain verifiers are equivalent to limiting refutation. Is there some kind of infinite, meta-methodological chain that characterizes limiting decision? Here is one example of such a constraint. Say that M *converges as fast as* M' given K

[16] The proof of this generalization is left to the reader. Hypotheses assessable by "mixed" chains of this sort are Boolean combinations of hypotheses that are verifiable in the limit. The complexities of such hypotheses are characterizable in the finite difference hierarchy over Σ^0_2.

just in case for each data stream e satisfying K, for each stage k of inquiry, if M' has converged by k, so has M.

Proposition 6 *The following situations are methodologically equivalent:*

1. *There exists an infinite chain $M_1, \ldots M_k, \ldots$ of methods such that*
   (a) *for each $i \geq 0$, method M_{i+1} aspires to decide P_i in the limit and does so under presuppositon P_{i+1},*
   (b) *For each $i \geq 1$, P_i entails P_{i+1},*
   (c) *K entails $(P_2 \vee \ldots \vee P_n \ldots)$, and*
   (d) *for each $i \geq 1$, M_{i+1} converges as fast as M_i given K.*
2. *H is decidable in the limit given K.*

This result may be understood, intuitively, as follows. Increased reliability requires a method to cope with more possibilities, which delays convergence. So requiring the successive meta-methods in the sequence to have nondecreasing reliability and *also* nonincreasing convergence time implies that reliability eventually stops increasing after some point in the sequence. The tail of the sequence provides no further essential information after that point leaving us with what is essentially a finite sequence of limiting decision methods (proposition 3).

11 Empirical Naturalism without Foundations or Direction

The preceding section provided an analysis of unfounded epistemic regresses. But it still assumed that the infinite, empirical regress is at least *directed*, in the sense that the meta-methods are increasingly reliable. What if we drop that assumption as well? Is foundationless, directionless meta-methodology necessarily pointless, in the sense that one could not turn the conjectures of the methods into a recognizable notion of convergence to the truth?

If no further conditions are added, then the answer is affirmative, since every hypothesis whatsoever possesses such a chain; so such a chain cannot be equivalent to methods succeeding in any of the convergent senses defined above.

Proposition 7 *Every H has an infinite chain $M_1, \ldots M_k, \ldots$ of methods like the one described in proposition 4, except that the nesting condition is dropped.*

The argument is simple: every method succeeds over some set (possibly empty) of relevant possibilities. So every infinite sequence of methods starting with acceptance and using at most one retraction witnesses the preceding proposition. Such chains are, therefore, extreme examples of vicious or pointless empirical regresses so far as reliable convergence to the right answer is concerned.

Are there further conditions we can impose on the methods in the undirected chain in order to end up with a condition equivalent to limiting decidability? Consider the following two properties. (1) A meta-method is *positively [negatively] reliable* given K just in case it never converges to acceptance [rejection] incorrectly on any data stream satisfying K. Let A_{i+1} denote the proposition that M_{i+1} eventually stabilizes to acceptance and let R_{i+1} denote the proposition that M_{i+1} eventually stabilizes to rejection. Then positive [negative] reliability requires that A_{i+1} [R_{i+1}] entail P_i [$\neg P_i$]. A meta-methodological chain is positively [negatively] reliable just in case each meta-method occurring in it is. (2) Another, possible property of infinite meta-methodological chains is *positive [negative] covering*. A chain positively [negatively] covers K just in case K entails $(A_2 \vee A_n \vee \ldots)$ $[(R_2 \vee R_n \vee \ldots)]$.

The positive [negative] covering and reliability conditions do not imply objective directedness in the sense that methods get more reliable farther out in the chain. Convergence to acceptance [rejection] implies truth, and every data stream is accepted [rejected] in the limit by some meta-method in the sequence, but that implies neither that later meta-methods accept more than earlier ones nor that later methods commit fewer errors than earlier ones. Indeed, a later method might reject every data stream. So these properties steer clear both of foundations and of directedness (i.e., increasing reliability).

Nonetheless, such a non-directed, foundationless chain exists precisely when the hypothesis is decidable in the limit by a single method! And the same is true even if the methods in the chain merely decide in the limit rather than refuting with certainty. So under suitable conditions, even unfounded, undirected meta-methodology can have an appealing point. [17]

Proposition 8 *The following situations are methodologically equivalent:*

1. *There exists an infinite chain $M_1, \ldots M_k, \ldots$ of methods such that*
   (a) *for each $i \geq 0$, method M_{i+1} aspires to refute [verify] P_i with certainty given K and does so under presuppositon P_{i+1},*
   (b) *the chain is positively [negatively] reliable, and*
   (c) *the chain positively [negatively] covers K.*
2. *There exists an infinite chain $M_1, \ldots M_k, \ldots$ of methods such that*

[17] The same result also holds for an infinite meta-methodological sequence of *verifiers* that are *negatively* reliable and whose eternal *rejection* propositions cover K. Simply follow the proof for the refutation case, substituting dual notions where appropriate.

(a) *for each $i \geq 0$, method M_{i+1} aspires to decide P_i in the limit given K and does so under the presuppositon P_{i+1},*

(b) *the chain is positively [negatively] reliable, and*

(c) *the chain positively [negatively] covers K.*

3. *P_0 is decidable in the limit given K.*

12 Refutations of Refutations and the Logic of Discovery

The preceding, meta-methodological characterizations of limiting decidability relate in an interesting way to the problem of discovery. Recall that a "discovery method" outputs propositions in response to new data and that an empirical question specifies a range of possible answers. Say that a method *answers* such a question *in the limit* just in case after some finite time it always produces a true conjecture that entails a correct possible answer to the question.

When the potential answers partition the relevant possibilities, it is well known that the question is answerable in the limit if and only if each potential answer is decidable in the limit (LRI, Chapter 9). Combining this fact with the preceding results yields

Proposition 9 *A question is answerable in the limit given K just in case each potential answer has a meta-methodological sequence S satisfying one of the following conditions:*

1. *S is a finite sequence of limiting decision procedures, the last of which has a presupposition covering K.*

2. *S is an infinite, nested sequence of limiting decision procedures whose presuppositions cover K, such that each later method is guaranteed to converge at least as fast as any preceding method, given K.*

3. *S is an infinite sequence of positively [negatively] reliable refuting [verifying] methods whose acceptance [rejection] sets jointly cover K.*

4. *S is an infinite sequence of positively [negatively] reliable limiting decision procedures whose acceptance [rejection] sets jointly cover K.*

Thus, even foundationless, undirected meta-methodology suffices for (and indeed is equivalent to) the existence of a method that answers the question in the limit. This result provides some logical vindication of Popper's otherwise perplexing faith that refutations of refutations of refutations without end ultimately add up to convergence to the truth in the limit.

13 Computable Methodology

The constructions occurring in the proofs of the above results are all computable (cf. the Appendix). So if the infinite meta-methodological sequences they operate upon are effectively presented, the result of composing the construction with the effectively presented meta-methodological sequence is a single, computable method that succeeds in the required sense. More precisely, say that meta-methodological sequence $M_1, \ldots, M_n, \ldots$ is computable just in case there exists a computable function C such that for each i and for each finite data sequence ϵ,

$$C(i, \epsilon) = M_i(\epsilon).$$

Thus, all of the above propositions continue to hold when the meta-methodological sequences and the methods equavalent to them are required to be computable.

This situation is not untypical. Since both computability theory and learning theory study similar criteria of problem solution, it often happens that results holding for "ideal" or "logically omniscient" methods can easily be transformed into closely analogous results concerning computable inquiry.

The same cannot be said for the alternative tradition in methodology, which models scientific method as a generalized entailment relation reflecting "confirmation" or "empirical support". According to that view, methodological norms are not based on computationally achievable aims, but on the maintenance of logical relations that computational agents cannot maintain, so that computational strategies are judged normatively deficient.

This tendency to model methodological norms using uncomputable logical relations is one of the most persistent features of the positivistic legacy. The *sine qua non* of logical positivism was a sharp distinction between questions depending on matters of fact and on mere relations of ideas. In methodology, this translates into a sharp distinction between empirical methods, which face inductive skeptical arguments, and formal methods, which do not. It has therefore seemed acceptable to deal with the problem of induction in its own right, reserving formal considerations like computability as an afterthought.

But once again, learning theory invites a very different viewpoint, despite its strongly logical character. If Hume had an excuse for thinking that all formal problems should be decidable *a priori* (since relations of ideas fall under the gaze of the mind's eye), we, as heirs to Gödel's legacy, do not. The message of Gödel's logical revolution is that from the viewpoint of a computational agent, formal problems are for all intents and purposes empirical, since a computer can no more see to the end of its computational process than a scientist can gaze at the indefinite future of her discipline. Indeed, when the formal problem appears to the computational agent to pose the problem of induction, the result is uncomputability! The easiest example of an undecidable formal problem is the *halting problem*, which requires one to determine whether a given Turing machine halts on a given input. The epistemic dimension of the problem is obvious: no matter how long the program has refused to return an output, it may nonetheless do so at the very next moment. Although the unsolvability of the halting problem is not usually proved by means of this skeptical argument, it can be (cf. LRI and Kelly and Schulte 1995a), and such a proof does much to clarify the structural analogy between the problem of induction and uncomputability.

Moreover, one may entertain limiting notions of success when a formal problem is not computably decidable. For example, the halting problem is computably verifiable with certainty and its complement is computably refutable with certainty. Similarly, non-halting is refutable with certainty even though it is not verifiable with certainty. The limiting concepts of success are represented as well. For example, determining whether a given Turing program computes a total function is computably refutable in the limit but is not computably verifiable in the limit. Moreover, solvability in each of these attenuated senses can be characterized in terms of alternating quantifiers, in just the manner indicated above for empirical problems.[18]

Since learning theory's treatment of induction is parallel to the approach to formal problems in the theory of computability, it should come as little surprise that learning theory leads to a precise account of the power of computable inquiry. On this approach, computable inquiry may be viewed as posing a two-fold problem of induction, an external one, reflecting the degree of interleaving of the data streams for and against the hypothesis through time, and an internal one, corresponding to the interleaving of epistemically possible future trajectories of one's own internal computations (i.e., to uncomputability). Accordingly, the respective characterizations of empirical problem solvability and of computability in terms of quantifier alternations can be neatly assembled into a characterization of computable problem solvablility (LRI Chapter 7). Methodological approaches based on Bayesian updating, on the other hand, assume an ideal account of probabilistic coherence that is uncomputable over sufficiently rich formal languages. It is impossible to integrate computability considerations into such an account without compromising the required sense of coherence.

I claimed, above, that learning theory's uniformly instrumental perspective on formal and empirical methodology yields logical arguments for Feyerabendian conclusions. One such critique concerns the proposal, shared by Bayesians and Popperians alike, that a hypothesis should be rejected when it becomes inconsistent with the data. There are familiar historicist objections to this principle based on Duhem's thesis that no hypothesis is ever really refuted. But let us suppose for the sake of argument that the data are perfectly reliable and that the hypothesis really is ideally refutable with certainty: if it is false, the data will eventually say so. Wouldn't the consistency principle be rationally mandated in this case? After all, hanging onto a refuted hypothesis delays convergence to the truth, so a method obeying the norm would weakly dominate a method that did not in convergence time.

[18] Cf. also (Putnam 1965) and (Hajek 1978).

But what if the consistency problem is uncomputable, so that it is impossible for a computer to verify whether the current data are consistent with it? One might suppose that we are are rescued from such cases by feasibility: we cannot be required to do the impossible. But escape isn't that easy. The rule requires that we never hold onto a hypothesis that is refuted, not that we decide consistency. Even if the full consistency problem is effectively unverifiable, we can still effectively satisfy the rule by erring on the side of caution and rejecting unrefuted hypotheses. But another goal is finding the truth. The question is whether there are problems in which the two goals clash for computable methods in the following sense: either can be computably achieved by itself, but no computable method achieves both. In that case, the consistency condition would no longer accelerate convergence to the truth: it would *prevent* convergence to the truth. So insisting on the rule would have effects entirely contrary to its intended function.

The answer is resoundingly affirmative: one can construct an empirical hypothesis that a computer can refute with certainty, but such that no method that always maintains consistency with the data succeeds even gradually; even if the method is, in a precise sense (i.e., hyperarithmetically definable) infinitely more powerful than a computable method (LRI Chapter 7, Kelly and Schulte 1995b).

An interesting corollary of this result is that any such method whose (hyperarithmetically definable) subroutine for detecting logical consistency is insulated from the empirical data as a separate "subroutine" fails to achieve anything close to what a mere Turing machine can do, namely, refute the hypothesis with certainty. So the very idea that theorem proving can be functionally isolated from empirical information radically restricts the potential power of computable science! This last vestige of the traditional, methodological dichotomy between matters of fact and relations is mistaken.

This is the kind of result I had in mind when I distinguished methodological discoveries from reflections on existing practice. The proof is based on a construction involving mathematical concepts that historical scientists had no idea about, since the theory of comutability hadn't been invented yet. Combing through the history of science will never yield such an insight (unless scientists do the logical work themselves so that historians can read about it in the historical record). Whether such logical possibilities will be realized in future scientific inquiry is hard to say. But here I agree entirely with Laudan: it is the objective, means-ends relations that matter. Whether or not the relationship has arisen in practice has to do with evidence for the relation (which in this case is established *a priori*) rather than with its normative force.

14 Conclusion

This paper provides some idea of the similarities and differences between two divergent images of normative naturalism. The first emphasizes structural analysis of the conditions under which a method *would* succeed, whereas the second focuses on historical surveys of apparently successful applications of a method in the past. I explained how the *a priori* approach provides a compelling role for logic in post-positivistic, naturalized methodology that embraces, rather than resists, much of the historicist critique of positivism and that avoids the inherent conservatism of historical surveys. I also presented a new, logical framework in which to distinguish useful empirical regresses from "vicious" ones. An important feature of this analysis is that useful empirical regresses can be unfounded and undirected, in the sense that later methods may fail to be more reliable than earlier ones. Finally, I illustrated how logic can be used to provide computational critiques of methodological principles whose instrumentality is obvious when computability is ignored.

Logic is not all there is to normative naturalism. But neither is history. As history, itself, suggests, scientific success demands detailed attention to the mathematical implications of the structures under investigation. Learning theory is just normative naturalism informed by this lesson.

15 Acknowledgement

Oliver Schulte's careful and detailed corrections led to major improvements in a late draft of the paper. Clark Glymour also provided prompt comments on an early draft.

16 Appendix

Proof of Proposition 2: (1) $\Rightarrow$ (2): Suppose we are given a finite meta-methodological sequence $((M_1, P_1), \ldots, (M_k, P_k))$ such that each M_i is an aspiring n_i retraction decision procedure given K, M_1 decides H with n_1 retractions starting with c_1 under presupposition P_1 and for each i from 1 to $k-1$, M_{i+1} decides P_i with n_{i+1} retractions starting with c_{i+1} under presupposition P_{i+1}. Moreover, let P_k include all serious possibilities in K.

We must construct a single method M that decides H with $n_1 + \ldots + n_k$ retractions over K starting with c, where $c =$ "accept" if the proposition that no method in the sequence ever rejects entails that H is correct, and is $c =$ "reject" otherwise.

The construction of M is as follows. M simulates all the methods in the sequence on the finite sequence of data input so far. Then M calculates its current conjecture by setting $b :=$ the number of methods among $M_1, \ldots, M_k$ that currently reject. If b is even, M accepts, and M rejects otherwise.

Let data stream e satisfying K be given. Since each method in the sequence uses at most a finite number of retractions and there are only finitely many such methods, there is a stage m after which each method M_i has stabilized to its ultimate conjecture u_i. We may now reason by "backward induction" as follows. Since M_k's presupposition is trivially satisfied, u_k is correct. So if u_k is rejection, and M_{k-1} is an aspiring n_{k-1} retraction decision procedure, M_{k-1} converges to the wrong answer, so we may reverse

u_{k-1} and we agree with u_{k-1} otherwise. Call this corrected conjecture v_{k-1}. Now v_{k-1} is correct, so if it is reject we reverse u_{k-2} and otherwise agree with u_{k-2} to obtain the corrected conjecture v_{k-2}. Proceeding in this way, we ultimately obtain the corrected v_1, which correctly indicates whether e satisfies H. Since one reversal occurs for each "rejection" occurring in $(u_1, \ldots, u_k)$ and since two reversals cancel, the correct conjecture u_1 is "accept" just in case an even number of "reject" conjectures occur in $(u_1, \ldots, u_k)$. So M converges to the correct answer.

Observe that M retracts only when the number of rejecting methods in the sequence changes from even to odd. In the worst case, each method in the sequence retracts at a different time (if two retract simultaneously, M does not retract). So at worst, M retracts $n_1 + \ldots + n_k$ times.

M starts with the leading conjecture c_1 of M_1 if there are an even number of retractions among the initial conjectures $c_2, \ldots, c_n$ of the meta-methods and starts out with the reversal of c_1 otherwise. But by the backward induction argument, this is the correct conjecture about H if all of the conjectures $c_1, \ldots, c_k$ are correct for data stream e.

(2) $\Rightarrow$ (1): Induction on the binary examples provided in proposition 1. To see how to generalize the construction, consult the proof of proposition 8. $\dashv$

Proof of Proposition 3: (1) $\Rightarrow$ (2): Given the chain, the backwards induction construction used to prove proposition 2 works here as well.

(2) $\Rightarrow$ (1): In the other direction, suppose M decides H in the limit given K. Then extend M with k meta-methods, all of which accept no matter what.

Proof of Proposition 4: (1) $\Rightarrow$ (2): Suppose we are given an infinite sequence $((M_1, P_1), \ldots, (M_i, P_i), \ldots)$ of methods, such that each M_i is an aspiring refuter given K and for each $i > 0, M_{i+1}$ refutes P_i with certainty under presupposition P_{i+1} (where $H = P_0$). Also, suppose that for each $i \geq 0$, P_i entails P_{i+1} and that the P_i cover K. We must construct a single M that refutes H with certainty given K. The construction is as follows. M starts out accepting and rejects if any method M_i in the chain ever rejects. Evidently, M aspires to refute given K. So it suffices to show that M converges to the right answer given K.

Let e satisfy K. Since the presuppositions cover K, let k be least such that e satisfies P_k. Case A: $k = 0$. Then all the P_i are satisfied so for each $i \geq 1$, M_i never rejects. Thus, M always accepts, which is correct because $P_0 = H$ is satisfied. Case B: Suppose $k > 0$. Then for each $i < k$, P_i is false and P_k is true. Hence, for each $i < k$, M_i converges to the wrong answer and, hence, converges to acceptance. M_k converges to the right answer about P_{k-1}, and hence rejects. Thus, M converges correctly to rejection.

(2) $\Rightarrow$ (1): Let M refute H with certainty given K. Let $M = M_1$ and for each $i > 1$, let M_i accept no matter what. $\dashv$

Proof of Proposition 5: (3) $\Rightarrow$ (2) $\Rightarrow$ (1): Immediate, since a certain verifier is a limiting decider which is, in turn, a limiting refuter.

(1) $\Rightarrow$ (4): Let P_0 denote H. Suppose we are given an infinite sequence $((M_1, P_1), \ldots, (M_i, P_i), \ldots)$ of meta-methods, such that for each $i \geq 1$, method M_i is an aspiring limiting refuter given K, and for each $i \geq 1, M_{i+1}$ refutes P_i in the limit under P_{i+1}. Also, suppose that for each $i \geq 0$, P_i entails P_{i+1} and that the P_i cover K. We must construct a single M that refutes P_1 in the limit given K.

The constructed method M works as follows. Let $f(0), f(1), \ldots, f(k), \ldots$ be an infinitely repetitive enumeration of the natural numbers. Initialize counter $p := 0$. On the first k data points, suppose that $p := i$. Feed the first k data points to $M_{f(i)+1}$ and see if $M_{f(i)+1}$ accepts. If so, increment $p := i + 1$ and accept, and otherwise leave p set to $p := i$ and reject.

Let e satisfy K. Then by assumption, for some $i \geq 1$, P_i is satisfied by e. Let k be the least such i. Case A: Suppose that $k = 0$, so each P_i including $P_0 = H$ is satisfied. Then each M_{i+1} accepts infinitely often along e, since M_{i+1} verifies P_i in the limit when its presupposition P_{i+1} is satisfied. Hence, the counter p is incremented infinitely often, so M accepts infinitely often, as required. Case B: Suppose that $k > 0$. Then by nesting and choice of k, $H = P_0, \ldots, P_{k-1}$ are not satisfied but P_k is. Hence, M_k converges to rejection along e, say by the time n data points have been read. Since $f(0), \ldots, f(n), \ldots$ is infinitely repetitive, there is an $m \geq n$ such that $f(m) + 1 = k$. Thus, p is never incremented past m, so M converges correctly to rejection.

(4) $\Rightarrow$ (1): Suppose that M refutes H in the limit given K. We need to construct an infinite sequence $((M_1, P_1), \ldots, (M_i, P_i), \ldots)$ of meta-methods, such that for each $i \geq 1$, meta-method M_i aspires to verify P_{i-1} with certainty given K and does so under presupposition P_i. We must also show that for each $i \geq 0$, P_i entails P_{i+1} and that the P_i cover K.

The construction is as follows. For each $i \geq 0$, let R_i denote the proposition that M rejects from stage i onward. In other words,

$$R_i = \{e \in K : \forall m \geq i, M(e(0), \ldots, e(m)) \text{ rejects}\}.$$

Then define

$$\begin{aligned} P_0 &= H. \\ P_{i+1} &= H \vee R_i. \end{aligned}$$

For each finite data sequence $(x_0, \ldots, x_k)$, define

$$M_1(x_0, \ldots, x_k) = \begin{cases} \text{accept} & \text{if there is a } j \leq k \text{ such that } M(x_0, \ldots, x_j) \text{ accepts} \\ \text{reject} & \text{otherwise,} \end{cases}$$

and for each $i > 1$, define:

$$M_{i+1}(x_0, \ldots, x_k) = \begin{cases} \text{accept} & \text{if } M(x_0, \ldots, x_{i-1}) \text{ rejects or} \\ & \text{there is a } j \text{ such that} \\ & \quad i \leq j \leq k \text{ and } M(x_0, \ldots, x_j) \text{ accepts} \\ \text{reject} & \text{otherwise.} \end{cases}$$

I now verify that the construction works. It is immediate that P_i entails P_{i+1} and that each M_i aspires to verify with certainty. To see that the P_i cover K, observe that P_0 covers H and since M refutes H in the limit given K, M converges to rejection by some finite stage so $K - H$ is also covered by the P_i. For the corollary, observe that each P_i is the disjunction of H with a proposition R_i that is refutable with certainty.

It remains only to check that for each $i \geq 0$, M_{i+1} verifies P_i with certainty under presupposition P_{i+1}. Since we already know that each M_i aspires to verify given K, it suffices to show that for each $i \geq 0$, M_{i+1} converges to the right answer about P_i on data stream $e \in K$ just in case e satisfies P_{i+1}.

Let e satisfy K. Suppose that e satisfies $H = P_0$. Then for all $i \geq 0$, P_i is satisfied, so for each $i \geq 1$, M_i must converge correctly to acceptance. But this is indeed the case, since M accepts infinitely often along e.

Suppose that e does not satisfy H. Then since M refutes H in the limit given K, we may choose n to be least such that for all $m \geq n$, $M(e(0), \ldots, e(m))$ rejects. Then for all i such that $0 \leq i < n$, $P_i = (H \vee R_i)$ is not satisfied by e and for all $i \geq n$, P_i is satisfied by e. So methods prior to $n + 1$ mistakenly converge to acceptance, method M_{n+1} correctly converges to rejection, and methods after $n + 1$ correctly converge to acceptance. Each of these facts will now be established.

Case A: $1 \leq i < n$. Hence, $n > 0$, so by the choice of n, we have that $M(e(0), \ldots, e(n-1))$ accepts. Since $n - 1 \geq i$, M_i mistakenly converges to acceptance, as required.

Case B: $i = n$. Suppose $n = 0$. Then M_1 converges correctly to rejection, as required. Suppose $n > 0$. By the choice of n, we have that $M(e(0), \ldots, e(n-1))$ accepts and never accepts thereafter. So M_{n+1} converges correctly to rejection, as required.

Case C: $i > n$. Suppose $n = 0$. Then M never accepts. So M_1 converges correctly to rejection, as required. Suppose $n > 0$. Then $i - 1 \geq 0$ and $M(e(0), \ldots, e(i-1))$ rejects. Hence, M_i correctly converges to acceptance, as required. $\dashv$

Proof of Proposition 6: (1) $\Rightarrow$ (2): Suppose that $((M_1, P_1), \ldots, (M_i, P_i), \ldots)$ is an infinite methodological chain of aspiring limiting deciders given K such that for each $i \geq 0$, M_{i+1} decides P_i in the limit under presuppositon P_{i+1}, P_i entails P_{i+1}, $K = (P_2 \vee \ldots \vee P_n \ldots)$, and M_{i+2} converges as fast as M_{i+1} given K. We need to construct an M that decides H in the limit given K.

M assumes that all of these simulated methods have already converged. M agrees with the conjecture of M_1 if an even number of the simulated methods currently reject and reverses the conjecture of M_1 otherwise.

Let e satisfy K. Since $K = (P_2 \vee \ldots \vee P_n \ldots)$, we may choose k to be least such that e satisfies P_{k+1}. Thus, there is a stage i after which M_{k+1} is correct about P_k. Since later methods converge as fast as and are as reliable as M_{k+1}, they correctly accept from stage i onward. The finitely many methods $M_1, \ldots, M_k$ eventually all converge (possibly incorrectly), by some later stage i', since they aspire to decide in the limit given K. After the stages k and i' are passed, backward induction (cf. the proof of proposition 2) shows that M is correct.

(2) $\Rightarrow$ (3): In the other direction, let M decide H in the limit given K. Let $M_2, \ldots, M_i, \ldots$ all accept no matter what, without looking at the data. This trivial meta-methodological sequence satisfies all of the required properties. $\dashv$

Proof of Proposition 8: (1) $\Rightarrow$ (2) is immediate. (2) $\Rightarrow$ (3): Suppose we are given an infinite sequence $((M_1, P_1), \ldots, (M_i, P_i) \ldots)$ of aspiring limiting decision procedures given K, such that M_1 decides H in the limit under presupposition P_1 and for each $i > 1, M_{i+1}$ decides P_i in the limit under presupposition P_{i+1}. Also, suppose that the sequence is positively reliable and positively covers K. We must construct a single method M that decides H in the limit given K.

The constructed method M works as follows. Let $f(0), f(1), \ldots, f(k), \ldots$ be an infinitely repetitive enumeration of the natural numbers. Initialize counter $p := 0$. After seeing the first k data points, suppose that $p = i$. Feed the first k data points to $M_{f(i)+1}$ and see if $M_{f(i)+1}$ rejects. If so, increment $p := i + 1$, and otherwise leave p set to $p = i$. Then return the current output of M_1 if an even number of methods among $M_2, \ldots, M_{f(p)+1}$ reject and return the result of reversing the current output of M_1 othewise.

Let e satisfy K. Then by the positive covering condition, for some $i \geq 0$, P_{i+1} is satisfied by e. Thus, M_{i+1} converges to acceptance, say at stage j. Then since f is infinitely repetitive, there is some stage $j' \geq j$ after which p is no longer incremented. Let m be the terminal value of p. Thus, M_{m+1} has converged to acceptance by stage j'. By positive reliability, e satisfies P_{m+1}. Since all of the methods are aspiring limiting decision procedures, there is some possibly later stage j'' by which all methods prior to $M_{f(p)+1}$ have converged. Thereafter, by the backward induction argument of proposition 2, M conjectures correctly about H.

(3) $\Rightarrow$ (1): In the other direction, suppose we are given an arbitrary method M that decides H in the limit given K. We must construct an infinite, positively reliable sequence $((M_1, P_1), \ldots, (M_i, P_i), \ldots)$ of refuting meta-methods that positively covers K.

Without loss of generality, we can assume that M starts out accepting prior to seeing any data (given a limiting decider M', the result M of forcing M' to accept prior to seeing any data is still a limiting decider). Let O be the proposition that M retracts an odd number of times prior to convergence and let E be the proposition that M retracts an even number of times. Let R_i be the proposition that M retracts at most i times. Let $H = P_0$. Now, for each $i \geq 1$, define

$$P_i = \begin{cases} (R_{i-1} \vee O) & \text{if } i \text{ is odd} \\ (R_{i-1} \vee E) & \text{if } i \text{ is even.} \end{cases}$$

and define, for each finite data sequence ϵ,

$$M_i(\epsilon) = \begin{cases} \text{reject} & \text{if } M(\epsilon) \text{ uses at least } i \text{ retractions} \\ \text{accept} & \text{otherwise.} \end{cases}$$

By construction, each M_i is an aspiring refuter given K.

Next, we must establish the positive covering condition. Since M decides H in the limit given K, M uses only finitely many retractions on each data stream e satisfing K. Suppose M uses j retractions on e, so e satisfies R_j and hence e satisfies P_{j+1}. Thus, the P_i cover K.

To establish the positive reliability condition, suppose for arbitrary $i \geq 1$ that M_i never rejects along e. We must show that e satisfies P_{i-1}

Case A: suppose $i = 1$. Then M never retracts along e. Since M starts out accepting, M always accepts. So since M decides $H = P_0$ in the limit, e satisfies P_0.

Case B: suppose $i > 1$. Then M retracts at most $i - 1$ times so e satisfies R_{i-1}. If i is even, then R_{i-1} entails $(R_{i-2} \vee O)$, which is just P_{i-1}. Similarly, if i is odd, then R_{i-1} entails $(R_{i-2} \vee E)$, which is just P_{i-1}. So in either case, e satisfies P_{i-1}, as required.

It remains only to check that for each $i \geq 1$, M_i refutes P_{i-1} with certainty under presupposition P_i.

When $i = 1$, we must show that M_1 refutes $P_0 = H$ under presupposition $P_1 = (R_0 \vee O)$. Let $e \in K$.

Case A: suppose e satisfies $P_1 = (R_0 \vee O)$.

Case A.1: suppose e satisfies R_0. Then M never retracts and hence converges to acceptance. Since M is correct, $H = P_0$ is satisfied by e. Since M never retracts, M_1 converges correctly to acceptance, as required.

Case A.2: suppose e satisfies O. So M retracts an odd number of times starting with acceptance, and hence M converges to rejection. Since M converges to the right answer, e does not satisfy $H = P_0$. But M_1 converges correctly to rejection after the first retraction is observed, as required.

Case B: suppose e does not satisfy $P_1 = (R_0 \vee O)$. Then M uses some even number of retractions greater than zero, starting with acceptance. Thus, M accepts H in the limit, and since M converges to the right answer, e satisfies $H = P_0$. But since M retracts at least once, M_1 converges incorrectly to rejection, as required.

Now consider the case in which $i > 1$.

Case I: suppose i is odd. Then $P_i = (R_{i-1} \vee O)$.

Case I.A: suppose e satisfies $P_i = (R_{i-1} \vee O)$.

Case I.A.1: suppose e satisfies $P_{i-1} = (R_{i-2} \vee E)$. If e satisfies O, then e satisfies R_{i-2}, so M_i correctly converges to acceptance, as required. If e satisfies E, then e satisfies R_{i-1}, so M_i converges correctly to acceptance, as required.

Case I.A.2: suppose e does not satisfy $P_{i-1} = (R_{i-2} \vee E)$. So e satisfies O but not R_{i-2}. Since i is odd, e does not satisfy R_{i-1} either, else M retracts exactly $i - 1$ times, which is an even number of times. Hence, M retracts at least i times, so M_i converges correctly to rejection, as required.

Case I.B: suppose e does not satisfy $P_i = (R_{i-1} \vee O)$. Then e satisfies E and hence satisfies $P_{i-1} = (R_{i-1} \vee E)$. Also, e does not satisfy R_{i-1}, so M retracts at least i times, and hence M_i converges incorrectly to rejection, as required.

Under case II, in which i is even, the same argument works, if one switches O with E everywhere. $\dashv$

17 Bibliography

Case, J. and C. Smith (1983) "Comparison of Identification Criteria for Machine Inductive Inference", *Theoretical Computer Science* 25: 193-220.

Arthur Donovan, Larry Laudan, and Rachel Laudan (1992) *Scrutinizing Science*, Baltimore: Johns Hopkins University Press.

Dreyfus, Hubert (1979) *What Computers Can't Do*, New York: Harper and Rowe.

Feyerabend, Paul (1975) *Against Method*: London: Verso.

Gold, E. M. (1965) "Language Identification in the Limit", *Information and Control* 10: 447-474.

Hàjek, P. (1978) "Experimental Logics and Π^0_3 Theories", *Journal of Symbolic Logic* 42: 515-522.

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

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

Kelly, K. and C. Glymour (1992) *Inductive Inference from Theory-Laden Data*, *Journal of Philosohical Logic* 21: 391-444.

Kelly, K. and O. Schulte (1995a) "Church's Thesis and Hume's Problem." *Proceedings of the XIIth Joint International Congress for Logic, Methodology and the Philosophy of Science.* Florence 1995.

Kelly, K. and O. Schulte (1995b) "The Computable Testability of Theories Making Uncomputable Predictions", *Erkenntnis* 43:29-66.

Kelly, K., O. Schulte, and C. Juhl (1997) "Learning Theory and the Philosophy of Science", *Philosophy of Science* 64: pp. 245-267.

Kuhn, Thomas (1962) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

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

Laudan, Larry (1984) *Science and Values*, Berkeley: University of California Press.

Laudan, Larry (1996) *Beyond Positivism and Relativism*, Boulder CO: Westview Press.

Levi, I. (1983) *The Enterprise of Knowledge*, Cambridge: M.I.T. Press.

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

Mayo, D. (1996) *Error and the Growth of Experimental Knowledge*, Chicago: University of Chicago Press.

Miller, D. (1979) "Popper's Qualitative Theory of Verisimilitude. *British Journal for the Philosophy of Science* 24:166-177.

Mormann, Thomas (1988) "Are All False Theories Equally False? A Remark on David Miller's Problem and Geometric Conventionalism." *British Journal for the Philosophy of Science* 39:505-519.

Nola, R. (1999) "On the Possibility of a Scientific Theory of Scientific Method", *Science and Education*, In press.

Osherson, D., M. Stob, and S. Weinstein (1986) *Systems that Learn.* Cambridge: MIT Press.

Karl R. Popper (1968), *The Logic of Scientific Discovery*, New York: Harper.

Putnam, H. (1963) "'Degree of Confirmation' and Inductive Logic", in *The Philosophy of Rudolph Carnap*, A. Schilpp, ed. LaSalle, IL: Open Court.

Putnam, H. (1965) "Trial and Error Predicates and a Solution to a Problem of Mostowski", *Journal of Symbolic Logic* 20: 49-57.

Schulte, O. (1998) *Hard Choices in Scientific Inquiry*, Doctoral Thesis, Department of Philosophy, Carnegie Mellon University.

Schulte, Oliver (1999a) "Means-Ends Epistemology". *The British Journal for the Philosophy of Science.* In Press.

Schulte, Oliver (1999b) "The Logic of Reliable and Efficient Inquiry". *The Journal of Philosophical Logic.* In Press.