Means-Ends Epistemology

Oliver Schulte
University of Alberta

October 3, 2001

Abstract. This paper describes the cornerstones of a means-ends approach to the philosophy of inductive inference. I begin with a fallibilist ideal of convergence to the truth in the long run, or in the "limit of inquiry". I determine which methods are optimal for attaining *additional* epistemic aims (notably fast and steady convergence to the truth). Means-ends vindications of (a version of) Occam's Razor and the natural generalizations in a Goodmanian Riddle of Induction illustrate the power of this approach. The paper establishes a hierarchy of means-ends notions of empirical success, and discusses a number of issues, results and applications of means-ends epistemology.

1. The Long Run In The Short Run

Inquiry begins with uncertainty. Empirical inquiry uses evidence to find the answers to the questions that prompted investigation. If these questions are of a general nature, empirical inquiry immediately faces the *problem of induction*: how to generalize beyond the evidence. David Hume observed that all such generalizations are uncertain and hence involve a risk of error, and concluded that none have a normative justification over and above our customs or inferential habits [Hume 1984]. A traditional response to Hume's skepticism seeks pure norms of inductive inference, enshrined in an "inductive logic" or a "theory of confirmation". This project motivated the work of Keynes [Keynes 1921] and Carnap [Carnap 1962], and counts modern-day Bayesians among its heirs (cf. [Earman 1992, Howson and Urbach 1989, Howson 1997]).

Hans Reichenbach proposed an original alternative approach to the problem of induction: he supplied a "pragmatic vindication" for a certain natural rule for estimating limiting frequencies of certain kinds of events, which he dubbed "the straight rule" [Reichenbach 1949], [Juhl 1994], [Salmon 1991]. The straight rule has the virtue that, if for some type of repeatable experiment the relative frequency of an outcome approaches a fixed limit, then the straight rule produces a sequence of conjectures that converges to that limit. Essentially, Reichenbach proposed a *means-ends* justification for an inference rule: his rule is guaranteed to achieve the end of converging to the right answer, if there is a right answer (limiting frequency) to be found. Reichenbach limited his pragmatic vindication to but one kind of rule for estimating probabilities.[1] His student Hilary Putnam extended the idea to a more general setting and investigated inference rules that are guaranteed to settle on the right answer in the long run, or in the "limit of inquiry". Putnam used his theory of long-run convergence to criticize Carnap's "confirmation functions" on the grounds that they were not the best means for achieving convergence to the truth [Putnam 1963]. The ideal of guaranteed convergence in the limit of inquiry is a precise formulation of a fallibilist, Peircean vision of empirical success in which inquiry may never yield certainty but nonetheless settles on the truth as more and more evidence is gathered. This idea is the philosophical core of *formal learning theory*, a mathematical framework for

[1] Reichenbach held that all inductive inference could be reduced to estimating probabilities.

studying questions of inductive inference with contributions from philosophers, logicians and computer scientists (for a survey, see [Kelly 1996]).

A long-standing criticism of Reichenbach's pragmatic vindication has been that convergence in the limit is consistent with any counterintuitive behaviour in the short run. Salmon discussed this issue in depth, and concluded that, to yield short-run constraints on inductive inferences, Reichenbach's means-ends analysis had to be augmented by pure norms of inductive rationality (for example, along Bayesian lines) [Salmon 1967, Salmon 1991].

But there is no need to abandon the spirit of Reichenbach's pragmatic, means-ends justification so quickly. Instead of appealing to "pure norms of confirmation", we may consider epistemic goals *in addition to* finding the truth. In this paper, I show that taking other epistemic aims into consideration yields strong constraints on what inference rules may conjecture in the short run.

I begin with a number of natural cognitive values: convergence to the truth, fast convergence to the truth, convergence to the truth with as few vacillations (retractions) as possible, and avoiding error. Corresponding to these objectives, I define a set of *standards of success* for inductive methods that meet them. The three most interesting of these standards derive from the aims of settling on the truth, and doing so quickly and with few vacillations. We may think of the latter two considerations as criteria of *efficient* convergence to the right answer. I illustrate the import of these efficiency criteria in two well-known inductive problems: The Occamian problem of determining whether a certain entity exists or not, and a Goodmanian "Riddle of Induction" involving a number of alternative colour predicates for emeralds. The theory of efficient long-run convergence provides a means-ends vindication of a version of Occam's Razor (namely, "do not posit the existence of entities that are observable in principle but that you haven't observed yet") as well as the natural projection rule in the Goodmanian Riddle (namely, project that "all emeralds are green" as long as only green emeralds have been observed).

The main result of this paper shows that the various standards of inductive success that stem from the cognitive goals mentioned above fall into a systematic *hierarchy of feasibility*: Some cognitive objectives are feasible whenever certain others are, but not vice versa. In this sense some epistemic objectives place more stringent requirements on inductive inquiry than others. These results establish a measure of the *inductive complexity* of a given empirical problem: An inductive problem P is easier than another problem P' if it is possible to attain a higher standard of success in P than in P'. This notion of complexity allows us to weigh on the same scale such diverse problems as the Occam problem, Goodman's Riddle of Induction, language learning and determining whether matter is finitely or infinitely divisible.

I address a number of questions that arise naturally for means-ends epistemology:

1. Consider a given standard of inductive success. What is the *general structure of inductive problems* in which that standard is feasible?

2. When a given standard of inductive success is feasible, what are the features of the inductive *methods* that attain that standard?

3. Are there *tensions and trade-offs* among various cognitive goals? If so, how great is the conflict?

4. What intuitively plausible norms for scientific reasoning have *means-ends justifications*?

5. Do the methodological recommendations from means-ends analysis depend on the *language* in which evidence and hypotheses are framed?

Let us begin with some fundamental concepts from learning theory.

2. Discovery Problems

Learning theory studies several classes of inductive problems, such as making predictions, testing hypotheses, inferring general theories, and others. This paper examines the following type of problem: Consider a collection $\mathcal{H}$ of mutually exclusive alternative hypotheses under investigation. Given some piece of evidence, which of the alternative hypotheses should the agent conjecture? Following Popper's and Kelly's usage, I refer to such problems as *discovery problems* [Popper 1968], [Kelly 1996].[2] The general definition of a discovery problem is as follows. Let a set of evidence items be given (e.g., observations of the colours of emeralds, sightings of particles, positions of planets, and so on). A **data stream** is an infinite discrete sequence of evidence items. For example, if the evidence statements are either "this emerald is green" or "this emerald is blue", then one possible data stream is the infinite sequence of observations of green emeralds. If the evidence statements are either "the Ω particle has appeared" or "the Ω particle has not appeared", a possible data stream is the infinite sequence of never observing the particle Ω. If ε is a data stream, then $\varepsilon|n$ denotes the first n observations in the data stream.

For my present purposes, I define an **empirical proposition** to be a set of data streams. The empirical content of a hypothesis is an empirical proposition, namely the set of data streams on which the hypothesis is *correct*, or true.[3] For example, the empirical content of the hypothesis "there is an Ω particle" is the set of all data streams on which an Ω particle is observed. The empirical content of the hypothesis "all emeralds are green" is just the data stream featuring only green emeralds. An empirical proposition K represents the inquirer's background knowledge about what observation sequences are possible. Now we are ready to define:

Definition 1. *A **discovery problem** is a pair $(\mathcal{H}, K)$, where K is an empirical proposition representing background knowledge, and $\mathcal{H}$ is a collection of mutually exclusive empirical hypotheses—that is, exactly one of the alternative hypotheses is correct on a given data stream.*

An **inference rule**, or **inductive method**, δ produces an empirical proposition $\delta(e)$ as its current theory in response to a finite evidence sequence e. There are of course other kinds of inductive methods, for example ones that revise "degrees of belief", as Bayesian methods do. In principle, means-ends rationality can guide agents in revising any epistemic state.[4]

Many important inductive problems from a variety of settings take the form of discovery problems. These include language learning [Osherson *et al.* 1986]; parameter estimation and "model selection" in statistics; and inferring theories in scientific disciplines

[2] [Kelly 1996] does not require the alternative hypotheses to be mutually exclusive.

[3] The operative notion of correctness may embody virtues of theories other than truth, for example empirical adequacy or problem-solving ability [Laudan 1977], [Kitcher 1993]. The results in this paper presuppose only that correctness is some relation between hypotheses and data streams.

[4] Putnam showed how means-ends analysis yields a critique of "confirmation functions" that produce "degrees of confirmation" in light of new evidence [Putnam 1963]. For learning-theoretic treatments of Bayesian updating see for example, [Earman 1992, Ch.9], [Osherson and Weinstein 1988], [Kelly *et al.* 1997], [Kelly and Schulte 1995], [Juhl 1997].

The writers who study the revision of full belief include [Gärdenfors 1988] and [Spohn 1988].

such as particle physics [Schulte 1997] and cognitive neuropsychology [Glymour 1994], [Bub 1994]. In this paper, I will consider two fairly simple but well-known and instructive inductive tasks: (1) Does a certain kind of entity exist?—I call this the *Occam problem*—and (2) a version of Goodman's celebrated *Riddle of Induction* involving various alternative colour predicates for emeralds.

3. Two Examples of Discovery Problems

3.1. The Occam Problem. Suppose a scientist wants to investigate by empirical means whether a certain type of entity, say an Ω particle, exists. Imagine that the physicist undertakes a series of experiments (bigger accelerators, more sophisticated detection devices, pure neutrino beams). As inquiry continues indefinitely, scientists obtain increasing segments of an infinite sequence of experimental outcomes, a data stream. There are two hypotheses under investigation: (1) "the Ω particle exists" and its complement, (2) "the Ω does not exist". To keep matters simple, let's grant the physicist generous assumptions about the powers of her experimental apparatus: I assume that if an experiment detects the particle in question, then the particle exists, and that if all experiments fail to detect a particle, then the particle does not exist (in other words, the particle is *not* hidden forever). Thus the hypothesis "Ω exists" is true just in case one of the scientists' experiments registers an Ω particle.[5] A natural inference rule for this problem is to conjecture that Ω does not exist until it has been detected in an experiment. I call this rule the **Occam rule** because it follows Occam's Razor ("do not needlessly multiply entities"). From the point of view of means-ends methodology, whether or not we should follow Occam's Razor depends on the extent to which that maxim furthers epistemic aims. In Section 8, I will show that the Occam rule is indeed the optimal inference procedure for the Occam problem, with respect to certain cognitive values specified in Section 4.

3.2. The Riddle of Induction. In his "New Riddle of Induction", Nelson Goodman introduces an unusual color predicate for emeralds [Goodman 1983].

> Suppose that all emeralds examined before a certain time t are green ... Our evidence statements assert that emerald a is green, that emerald b is green, and so on ...
>
> Now let me introduce another predicate less familiar than "green". It is the predicate "grue" and it applies to all things examined before t just in case they are green but to other things just in case they are blue. Then at time t we have, for each evidence statement asserting that a given emerald is green, a parallel evidence statement asserting that emerald is grue.

The question is whether we should conjecture that all emeralds are green rather than that all emeralds are grue when we obtain a sample of green emeralds examined before time t, and if so, why. I shall treat this as a question about optimal inference in a discovery problem, in which the set of alternative hypotheses comprises the universal generalizations of the various colour predicates under consideration. To see what the empirical content of these hypotheses is, notice that they determine, for each "examination time", a unique colour for the emerald examined at that time. Thus the empirical content of the claim

[5] Instead of assuming that particles are observable, we could treat the scientist as aiming for empirically adequate theories rather than true ones, in the manner of [Van Fraassen 1980]. The hypothesis "Ω does not exist" is empirically adequate just in case scientists never observe Ω (yet false if there is a hidden Ω particle). Section 8 below has some discussion of the methodological role of unobservable particles.

"all emeralds are green" is that at each time, the emerald examined at that time is green; that is, the empirical content comprises the one data stream on which only green emeralds are observed. The empirical content of "all emeralds are grue" comprises the single data stream on which all emeralds examined before the critical time t are green, and those examined at time t and later times are blue. If not all emeralds are examined, then "all emeralds are green (grue)" may be empirically adequate yet false, namely if all emeralds that have been or will be examined are green (grue) but the unexamined ones are not. In what follows, I shall not be concerned with this possibility.[6] In particular, for the sake of more natural expression, I will use the term "all emeralds" to implicitly mean the same as "all examined emeralds"; for example I will say that the conjecture "all emeralds are green" is correct, or true, on the data stream along which only green emeralds are found.

A discovery problem is defined by the range of alternative hypotheses under consideration and given background knowledge. In the Riddle of Induction, what shall we take as the hypotheses under serious investigation, or as Goodman might say, as candidates for projection? Let's write $grue(t)$ for the grue predicate with critical time t. In this paper, I examine the **infinitely iterated Riddle of Induction**, which includes all $grue(t)$ predicates as alternative hypotheses—candidates for projection—for any natural number t.[7] The background knowledge in the infinitely iterated Riddle of Induction is that one of these alternatives is true. Figure 1 illustrates the infinitely iterated Riddle of Induction. To keep the example simple, I don't include the $bleen(t)$ predicates among the candidates for projection. It is straightforward to extend the methodological analysis to Riddles of Induction that include "bleen" predicates, but that would not yield more philosophical insight.

Now for inference rules, or as Goodman calls them, **projection rules**. The *natural projection rule* in the infinitely iterated Riddle of Induction conjectures that all emeralds are green as long as all emeralds observed so far are green. If a blue emerald is found at time t for the first time, the natural projection rule concludes that all emeralds are $grue(t)$.

From the point of view of means-ends methodology, whether an agent should follow the natural projection rule depends on whether it furthers epistemic aims. In Section 8, I will show that the natural projection rule is indeed the optimal inference procedure for the infinitely iterated Riddle of Induction, with respect to certain cognitive values—the same ones that single out the Occam rule as the best one in the Occam problem. I now turn to specifying these cognitive values.

4. Standards of Inductive Success

Epistemologists have proposed a number of desiderata for empirical inquiry, including: convergence to a correct theory, fast convergence to a correct theory, convergence with few retractions and vacillations, providing theories with interesting content, avoiding error, and defining scientific theories parsimoniously with few postulates, or "laws of nature".

[6] As in the Occam problem, we might assume that in the long run, all existing emeralds will be examined. Or we might just not care about emeralds that are forever hidden from sight; in other words, we may concern ourselves with empirical adequacy only.

[7] There are other versions of the Riddle of Induction: in the *one-shot* Riddle, we consider only two hypotheses, "all emeralds are green" and "all emeralds are grue" for some fixed critical time t; see also Section 8. In a *finitely iterated* Riddle, the alternatives include "all emeralds are green" and "all emeralds are $grue(1)$", "all emeralds are $grue(2)$", ...,"all emeralds are $grue(m)$" up to some last critical time m. Other formulations permit more than one blue-green colour change to occur (e.g., colour predicates such as "gruegr" and "bleenbl"). [Schulte forthcoming] examines all of these versions.

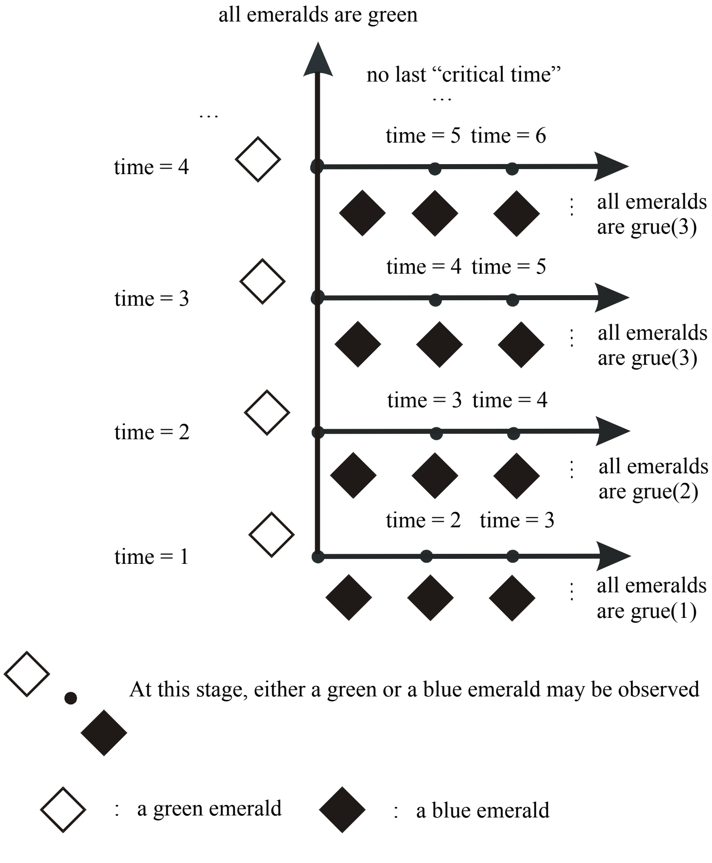

Figure 1: The Infinitely Iterated Riddle of Induction

An inductive method may perform well with respect to one or more of these aims in some circumstances but not in others. To compare the performance of methods with regard to a range of possible ways the world might be—more precisely, with regard to all the data streams consistent with background knowledge—I rely on two familiar principles from decision theory: **admissibility** and **minimax**.[8] A method is *admissible* iff it is not *dominated*. In general, an act A dominates another act A' if A necessarily yields results at least as good as those of A', and possibly better ones, where a given collection of "possible states of the world" determines the relevant sense of necessity and possibility. An act A *minimaxes* if the worst possible outcome from A is as good as the worst possible outcome from any other act. (I give precise definitions of these notions for the inductive context below.) Combining the 6 epistemic desiderata mentioned with the 2 evaluation criteria we arrive at 12 performance standards for inductive methods (for example, one of these would be "choose admissible methods for convergence to a correct theory").

I refer to these standards of empirical success as "means-ends recommendations" or as "hypothetical imperatives" for inductive inference. The three following sections define and discuss the three most interesting performance standards: admissibility with respect to convergence (reliability), admissibility with respect to convergence time (time efficiency), and minimaxing retractions (avoiding vacillations).

5. Convergence To The Truth and Nothing But The Truth: Logical Reliability

Through the ages, skeptical arguments dating back at least to Sextus Empiricus have aimed at showing that we cannot establish generalizations from a finite sample with certainty—for the very next observation might refute our general conclusions [Sextus Empiricus 1985]. This observation is also the basis of Hume's celebrated *problem of induction* [Hume 1984]. A fallibilist response is to give up the quest for certainty and require only that science eventually settle on the right answer in the "limit of inquiry", without ever producing a signal as to what the right answer is. As William James put it, "no bell tolls" when science has found the right answer [James 1982]. This conception of empirical success runs through the work of Peirce, James, Reichenbach, Putnam and others. For discovery problems, we may render it in a precise manner as follows.

An empirical proposition P is **consistent** if it is correct on some data stream (i.e., P is not empty). An empirical proposition P **entails** another empirical proposition P' just in case P' is correct whenever P is (i.e., P is a subset of P'). If $\mathcal{H}$ is a collection of alternative hypotheses, then say that an inductive method δ **converges to the correct hypothesis on a data stream by time** n just in case forever after, the method's output is consistent and entails the hypothesis correct for that data stream. If the method converges to the correct hypothesis by some time (on a given data stream), then we say simply that it *converges to the correct hypothesis* (on that data stream).

The Occam rule has the virtue of converging to the right hypothesis on *every* data stream. If the Ω particle exists, then (by our assumptions) it will eventually be observed; at that time, the Occam rule concludes with certainty that the Ω particle exists. If the Ω particle does not exist, the Occam rule always entails this fact, right from the beginning—although never with certainty.

Similarly, in the infinitely iterated Riddle of Induction, the natural projection rule is guaranteed to converge to the right hypothesis. If a blue emerald is ever found, the natural

[8] Another prominent choice principle is to maximize (subjective) expected utility. I shall say more about expected utility at the end of Section 10.

projection rule conclusively identifies the correct generalization about emerald colours. If all emeralds are green, the natural projection rule always entails this fact, right from the beginning (without certainty). Following [Kelly 1996], I refer to inductive methods that are guaranteed to eventually entail the right answer no matter what the right answer is as (logically) *reliable*. Logical reliability is the core notion of formal learning theory [Gold 1967, Putnam 1965].

Definition 2. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. An inductive method is* ***reliable*** *for the discovery problem $(\mathcal{H}, K)$ $\iff$ the method converges to the correct hypothesis on every data stream consistent with background knowledge K.*

An unreliable method in the Occam problem would be that of a theorist who maintains his faith in the existence of the Ω particle no matter how many experiments have failed to discover it; on the data stream of infinitely many failures, this theorist converges to a wrong theory. Similarly, a "grue" aficionado would be unreliable in the infinitely iterated Riddle of Induction: if she kept conjecturing that some future emerald is blue, she would never arrive at the generalization that all emeralds are green if in fact they are.

Reliable methods succeed in finding the correct hypothesis where unreliable methods fail. Those whose aim in inquiry is to find a correct theory prefer methods that converge to the truth on a wider range of possibilities. For example, the thrust of Putnam's critique of Carnap's confirmation functions was that Carnap's confirmation functions are not the best for detecting regularities among the data, in the sense that other methods succeed in doing so over a wider range of possibilities [Putnam 1963]. (For an evaluation of Putnam's argument, see [Kelly *et al.* 1994].) This is just the decision-theoretic principle of *admissibility* applied to inductive methods. The admissibility principle yields the following criterion for comparing the performance of two inductive methods with respect to convergence time.

Definition 3. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. In the discovery problem $(\mathcal{H}, K)$, an inductive method δ* ***dominates*** *another inductive method δ'* ***with respect to convergence*** $\iff$

1. *background knowledge K entails that δ converges to the correct hypothesis whenever δ' does, and*
2. *there is some possible data stream, consistent with background knowledge K, on which δ converges to the correct hypothesis and δ' does not.*

An inductive method is ***convergence-admissible*** *for a discovery problem $\iff$ the method is not dominated in that problem with respect to convergence.*

It is clear that reliable methods are convergence-admissible because they eventually arrive at the truth on *every* data stream (consistent with given background knowledge). The converse holds as well: less than fully reliable methods are dominated with respect to convergence. Thus applying the admissibility principle to the aim of converging to the truth leads to logical reliability.

Proposition 4. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. An inductive method is convergence-admissible for the discovery problem $(\mathcal{H}, K)$ $\iff$ the method is reliable for that problem.*

The proof is in [Schulte forthcoming]. Learning theorists have studied the structure of discovery problems with reliable solutions extensively, as well as the properties of reliable methods for given discovery problems. [Kelly 1996], especially Chapter 9, and [Osherson *et al.* 1986] introduce many of the key ideas.

I now turn to other epistemic aims in addition to convergence to the truth.

6. Fast Convergence To The Truth: Data-Minimality

Other things being equal, we would like our methods to arrive at a correct theory sooner rather than later. In isolation from other epistemic concerns, minimizing convergence time is a trivial objective: An inquirer who never changes her initial conjecture converges immediately. The interesting question is which *reliable* methods converge as fast as possible. We can use the admissibility principle to evaluate the speed of a reliable method as follows.

Definition 5. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be background knowledge. In the discovery problem $(\mathcal{H}, K)$, an inductive method δ* ***dominates*** *another inductive method δ'* ***with respect to convergence time*** $\iff$

1. *background knowledge K entails that δ converges at least as soon as δ' does, and*
2. *there is some data stream, consistent with background knowledge K, on which δ converges before δ' does.*

An inductive method δ is ***data-minimal*** *for a discovery problem $(\mathcal{H}, K) \iff \delta$ is not dominated in $(\mathcal{H}, K)$ with respect to convergence time by another method δ' that is reliable for that problem.*

The term "data-minimal" expresses the idea that methods that converge as soon as possible make efficient use of the data (cf.[Gold 1967], [Kelly 1996]). Data-minimal methods are the ones that satisfy a simple, intuitive criterion. Let's say that a method *does not take its conjecture seriously* at a given stage of inquiry if background knowledge entails that the method will abandon its conjecture later no matter what evidence it receives. Then the reliable data-minimal methods are exactly those reliable methods that take their conjectures seriously at each stage of inquiry.

If a method does take its conjecture seriously at a given stage of inquiry, then there is a data stream, consistent with background knowledge and the evidence obtained so far, on which the method converges to its current hypothesis by the current stage of inquiry. In this case, I say that the method **projects** its current hypothesis. Thus the reliable data-minimal methods are exactly those reliable methods that project their conjectures at each stage of inquiry.

For example, on the data stream that never turns up the Ω particle, the Occam rule projects its conjecture "the Ω particle does not exist" at each stage of inquiry along that data stream. By contrast, consider a cautious inference procedure for the Occam problem that waits until more experiments are performed before making a conjecture about the existence of the Ω particle. While the cautious procedure is waiting, it (trivially) does not project its conjecture along any data stream, because it is not conjecturing any alternative hypothesis.

The natural projection rule in the infinitary Riddle of Induction projects "all emeralds are green" at each stage along the data stream featuring only green emeralds. By contrast, another inference rule might wait until one or more of the "critical times" have passed before generalizing. Such a cautious inference rule fails to project a universal generalization about emerald colours while it is waiting.

Here's why a data-minimal method must always take its current conjecture seriously: If a method fails to take its conjecture seriously on some finite data sequence e consistent with the given background knowledge K, then we can speed it up on some data stream ε consistent with that background knowledge and evidence e without slowing it down on other data streams, because the method is not converging on other data streams anyway. Conversely, if a reliable method δ always takes its conjectures seriously given the background knowledge K, any other method δ' that seeks to find the true hypothesis before δ on some data stream ε must disagree with the conjecture of δ at some point along ε, say at stage n. But since δ always takes its conjectures seriously, it does so after receiving evidence $\varepsilon|n$, and locks onto its conjecture $\delta(\varepsilon|n)$ on some data stream τ extending $\varepsilon|n$. Since we supposed that δ' disagrees with δ on $\varepsilon|n = \tau|n$, the theory of δ' on $\varepsilon|n = \tau|n$ must be false on the data stream τ. So the method δ' converges by time n on data stream ε, but only after stage n on data stream τ (if at all). Hence δ is faster than δ' on τ. This shows that methods that always take their conjectures seriously—that always project their conjectures—are data-minimal.

Theorem 6. *Let $\mathcal{H}$ be a collection of alternative empirical hypotheses, and let K be given background knowledge. A reliable method is data-minimal for the discovery problem $(\mathcal{H}, K) \iff$ on each finite data sequence consistent with background knowledge K, the method projects its conjecture on that data sequence given background knowledge K.*

The formal proof is in [Schulte forthcoming].

7. Steady Convergence To The Truth: Retraction-Minimality

Thomas Kuhn argued that one reason for sticking with a scientific paradigm in trouble is the cost of retraining and retooling the scientific community [Kuhn 1970]. The philosophical literature around "minimal change" belief revision starts with the idea that minimizing the extent of retractions is a plausible desideratum for theory change [Gärdenfors 1988]. Similarly, learning theorists have investigated methods that avoid "mind changes" [Putnam 1965, Case and Smith 1983, Sharma *et al.* 1997]. For discovery problems, this motivates a different criterion for evaluating the performance of a method on a given data stream: we want methods whose conjectures vacillate as little as possible. Consider an inductive method for a collection of alternatives. I say that a method **retracts** its conjecture on a data sequence $e_1, ..., e_n, e_{n+1}$, or **changes its mind**,[9] if

1. the method's theory on the previous data $e_1, ..., e_n$ is consistent and entails one of the alternative hypotheses,
2. and the method's theory on the current data $e_1, ..., e_n, e_{n+1}$ either entails a different hypothesis, or fails to entail any of the alternative hypotheses.

The aim of avoiding retractions by itself is trivial, because the skeptic who always conjectures exactly the evidence never retracts anything. But can we use this aim to *select among* the reliable methods the ones that avoid retractions, as we did with convergence

[9] Of course, rules do not have minds to change—only rule-followers do. But for my present purposes, there is no equally brief and vivid alternative to speaking of a rule or a method changing its mind.

time? In this Section, I consider the consequences of using the admissibility principle to do so. We may define admissibility with respect to avoiding retractions as we did with respect to convergence time in Definition 5, replacing time-to-convergence with number of mind changes.

Definition 7. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be background knowledge. In the discovery problem $(\mathcal{H}, K)$, an inductive method δ* ***dominates*** *another inductive method δ' with respect to mind changes $\iff$*

1. *background knowledge K entails that δ changes its mind no more often on a data stream than δ' does, and*
2. *there is some data stream, consistent with background knowledge K, on which δ changes its mind less often than δ' does.*

A method δ is ***retraction-minimal*** *for a discovery problem $(\mathcal{H}, K) \iff \delta$ is not dominated in $(\mathcal{H}, K)$ with respect to mind changes by another method δ' that is reliable for that problem.*

What does retraction-minimality amount to? Avoiding mind changes pulls an inquirer away from generalizations that may have to be taken back later, and towards the "skeptic's" caution about going beyond the available evidence. But long-run reliability forces a skeptic to take a chance eventually and conjecture a hypothesis that goes beyond the evidence (if the discovery problem is a genuine problem of induction, in which the evidence does not eventually entail the correct hypothesis conclusively). But because long-run reliability is consistent with any behavior in the short run, the skeptic may delay this moment and the risk of having to retract her theories, for as long as she pleases. For any amount of time that the skeptic may choose to delay taking an inductive risk, she could have delayed more and been just as reliable in the long run. It follows that a reliable method is retraction-minimal just in case it never changes its mind; that is, just in case it never goes beyond the available evidence. Hence if there is a reliable retraction-minimal method, it must be possible to "wait and see" until the evidence settles which theory is true. The next proposition formulates this phenomenon precisely.

Proposition 8. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. Then there is a reliable retraction-minimal method for the discovery problem $(\mathcal{H}, K) \iff$ the background knowledge K entails that eventually the evidence will conclusively establish one of the alternative hypotheses as the correct one.*

The proof is in Section 11.

Proposition 8 points to an interesting methodological phenomenon: In discovery problems that are genuinely inductive, myopically avoiding retractions at each stage of inquiry leads to inferences that are not reliable in the long run. An example will help to illustrate this.

In their famous Turing address, Newell and Simon formulated the *physical symbol system hypothesis*, according to which there is some computer program whose intelligence matches that of humans [Newell and Simon 1976].[10] If cognitive scientists experience a

[10] The other part of the physical symbol system hypothesis is the converse: That any intelligent system must have the capacities of a computer.

series of failures in building an intelligent system, they must eventually abandon the physical symbol system hypothesis, or else risk the possibility that they go on for eternity searching for the path to machine intelligence when there is none (cf. [Kelly 1996], [Glymour and Kelly 1990]). But just *when* should they become pessimistic about the prospects of cognitive science? Suppose that the researchers try programs $m_1, m_2, ..., m_n$, in vain, and now consider whether to give up the physical symbol system hypothesis, or else to try one more program m_{n+1} before they conjecture that machine intelligence is impossible. As far as long-run reliability is concerned, it makes no difference whether they give up the belief in machine intelligence after seeing the failures on the first n machines, or after trying another one. But with respect to retractions, maintaining faith in cognitive science until the system m_{n+1} has been tried dominates giving up the belief in cognitive science beforehand. For if m_{n+1} is successful, the researchers need not have retracted their belief in the physical symbol system hypothesis.

But if m_{n+1} fails too, AI researchers may reason in the same way again: trying one more system m_{n+2} before recanting their faith in artificial intelligence might save them a retraction, without risking any additional ones. If the researchers continue to avoid changing their belief in cognitive science in this way, they will never recognize that machine intelligence is impossible if it actually is.

The cognitive scientists' dilemma is an instance of the general problem of when exactly an inquirer or a group of inquirers should abandon their current paradigm. The scientists must eventually jump ship if they want to avoid following the wrong paradigm forever. But as Thomas Kuhn observed, there is typically no particular point at which the revolution must occur [Kuhn 1957]. Indeed, short-run considerations such as avoiding the embarrassment of dismissing their prior work—that is, avoiding retractions—pull scientists in the direction of conservatism. [Kelly *et al.* 1997, Sec. 4] discusses the problem of deciding among scientific paradigms from a reliabilist perspective.

We saw that applying admissibility to the aim of avoiding retractions yields a standard of performance that is too high for the interesting inductive problems, in which we cannot simply rely on the evidence and background knowledge to eliminate all but the true alternative. Learning theorists have examined another decision criterion by which we may evaluate the performance of a method with respect to retractions: the classic minimax criterion. Unlike retraction-minimality, minimaxing retractions is feasible even when there is a problem of induction. Indeed, this criterion turns out to be a very fruitful principle for deriving constraints on the short-run inferences of reliable methods.

8. Steady Convergence To The Truth: Minimaxing Retractions

The minimax principle directs an agent to consider the *worst-case* results of her options and to choose the act whose worst-case outcome is the best. So to minimax retractions with respect to given background assumptions, we consider the maximum number of times that a method might change its mind if the background assumptions are correct. Suppose background knowledge entails that an inductive method δ changes its mind no more than n times on any data stream, whereas background knowledge does not rule out that another method δ' may change its mind more than n times on some data stream. Then the principle of minimaxing retractions directs us to prefer the method δ to the method δ'. The principle of minimaxing retractions by itself is trivial, because the skeptic who always conjectures exactly the evidence never retracts anything. But using the minimax criterion to *select among* the reliable methods the ones that minimax retractions yields interesting results, as we shall see shortly. The following definition makes precise how we may use the minimax criterion in this way.

Definition 9. *Suppose that δ is a reliable discovery method for alternative hypotheses $\mathcal{H}$ given background knowledge K. Then δ* **minimaxes retractions** $\iff$ *there is no other reliable method δ' for the discovery problem $(\mathcal{H}, K)$ such that the maximum number of times that δ might change its mind, given background knowledge K, is greater than the same maximum for δ'.*

If there is no bound on the number of times that a reliable method may require a change of mind to arrive at the truth (for examples of such discovery problems see [Kelly 1996, Ch.4], or the Hypergrue problem described in [Schulte forthcoming]), there is no maximum number of mind changes for reliable methods, and the minimax criterion has no interesting consequences.[11] But if we know that a reliable method can succeed in identifying the correct hypothesis without ever using more than n mind changes, the principle selects the methods with the best such bound on vacillations. I say that a method identifies a true hypothesis from a collection of alternatives $\mathcal{H}$ given background knowledge K **with at most n retractions** if background knowledge entails that (1) the method will identify the correct alternative hypothesis, and (2) that the method changes its mind at most n times on any given data stream. The goal of minimaxing retractions leads us to seek methods that succeed with as few mind changes as possible; learning theorists refer to this paradigm as discovery *with bounded mind changes* [Kelly 1996, Ch.9], [Sharma *et al.* 1997].

To get a feel for what minimaxing mind changes requires, consider a theorist whose initial conjecture in the Occam problem is that the Ω particle does exist. Reliability demands that she eventually change her mind if the Ω particle fails to appear. But after the point at which she finally concludes that the particle will not be found, a more powerful experimental design could turn it up, forcing her to change her mind for the *second* time. By contrast, the Occam rule never retracts its conjecture if the Ω particle does not exist, and changes its mind exactly once if it does appear, to conclude (with certainty) that it exists. Hence the Occam rule changes its mind at most once, whereas the theorist beginning with faith in the existence of the Ω particle may have to change her mind twice in the worst case; see Figure 2. So reliability and the principle of minimaxing retractions select the Occam rule over the latter inference procedure in this particular problem. If we combine this with data-minimality, minimizing convergence time requires a theorist to immediately make a conjecture about whether the Ω particle exists or not, and reliability together with minimaxing retractions dictate that this conjecture must be "the Ω particle does not exist". Hence we have the following result.

Proposition 10. *The only inductive method that is reliable, data-minimal and minimaxes retractions in the Occam problem is the Occam rule.*

Thus reliability and efficiency considerations underwrite a variant of Occam's rule, namely: Do not posit the existence of entities that are observable in principle, but that have not yet been observed. Note that this means-ends recommendation agrees with Occam's Razor only about entities that are observable in principle: The very same epistemic aims may *require* a particle theorist to *posit* the existence of hidden particles! Briefly, it

[11] [Sharma *et al.* 1997] presents some interesting generalizations that apply even in this case. Briefly, the idea is that methods annouce mind-change bounds together with a conjecture, and may dynamically revise their current mind-change bound as they receive further evidence.

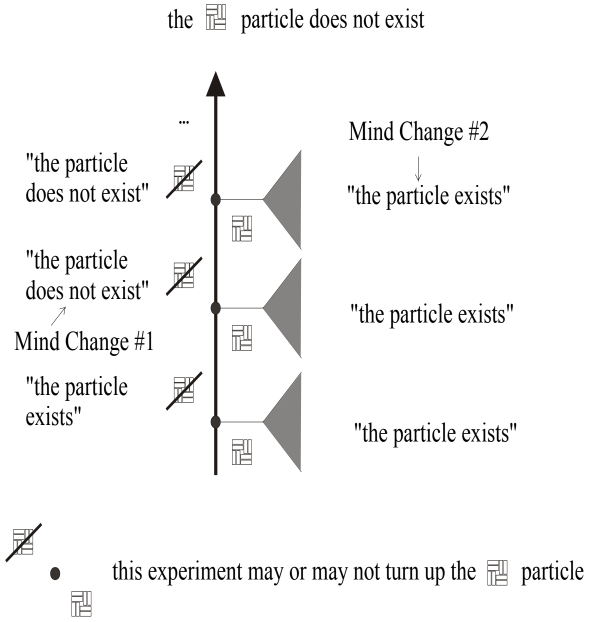
the particle does not exist
...
Mind Change #2
"the particle does not exist"
"the particle exists"
"the particle does not exist"
"the particle exists"
Mind Change #1
"the particle exists"
"the particle exists"
this experiment may or may not turn up the particle

Figure 2: In the Occam Problem, violating Occam's Razor may lead to two retractions.

can be shown that reliability and minimaxing retractions mandate a theorist to choose a "closest fit" to the available evidence of observed particle reactions; that is, these principles select the theory that conjectures that the reactions seen so far are the only possible ones. If we follow current practice in particle physics and seek conservation principles (selection rules or quantum properties) that describe which particle reactions are possible, we find that the logic of conservation principles entails that sometimes the theory that provides the closest fit to the available evidence must introduce hidden particles into its ontology. [Schulte 1997] develops a model of theory discovery in particle physics and shows in details why this is so; there is no room to pursue this application of reliabilist methodology here.

The same means-ends vindication as for Occam's Rule holds good for the natural projection rule in Goodman's Riddle of Induction. First, data-minimality requires projecting one of the alternative colour predicates at each stage of inquiry. Second, consider an unnatural projection rule that conjectures that "all emeralds are $grue(n)$" when a sample of k green emeralds has been obtained, for some $k < n$. If green emeralds continue to be found past the critical time n, this projection rule must eventually change its mind to project "all emeralds are green"—otherwise it would converge to a mistaken generalization about the colour of emeralds (namely "all emeralds are $grue(n)$") and would be unreliable. But after the point at which the unnatural rule projects that all emeralds are green, a blue emerald may be found, forcing the rule to change its mind for the *second* time.

By contrast, the natural projection rule changes its mind at most once: not at all if all emeralds are green, and once to conclude (with certainty) that all emeralds are $grue(k)$ if the k-th emerald is found to be blue. Since the unnatural projection rule retracts its conjecture twice in the worst case, the principle of minimaxing retractions rejects it in favour of the natural projection rule;[12] see Figure 3. The next proposition asserts that the *only* reliable, data-minimal projection rule that minimaxes retractions is the natural one. The argument for this is essentially the same as for Proposition 10; the formal proof is in [Schulte forthcoming].

Proposition 11 [with Kevin Kelly]. *In the Riddle of Induction, the only reliable and data-minimal projection rule that minimaxes retractions is the natural one.*

The similarity of Figures 2 and 3 suggests that the Occam problem and the infinitely repeated Riddle of Induction share a common structure, and that it is this structure which makes the natural inferences optimal by our means-ends criteria. The structure in question is the *topology of the possible data streams* in relation to the alternative hypotheses under investigation. [Schulte forthcoming] gives an explicit characterization of this structure for arbitrary discovery problems. More precisely, [Schulte forthcoming] specifies necessary and sufficient conditions on the topology of the data streams that in a given discovery problem must hold if it is possible to reliably identify the correct hypothesis with a bounded number of retractions in that problem. This characterization theorem shows that two facts are crucial in singling out the natural projection rule as the optimal one in the infinitely repeated Riddle of Induction. First, if all emeralds are *green*, no finite sample of emeralds will be observed that entails this fact. Second, if all emeralds are $grue(t)$, for some critical time $t > 0$, then by contrast we will observe a finite sequence of emeralds that falsifies all alternatives to "all emeralds are $grue(t)$" under consideration—namely $t - 1$

[12] Kevin Kelly was the first to observe this fact.

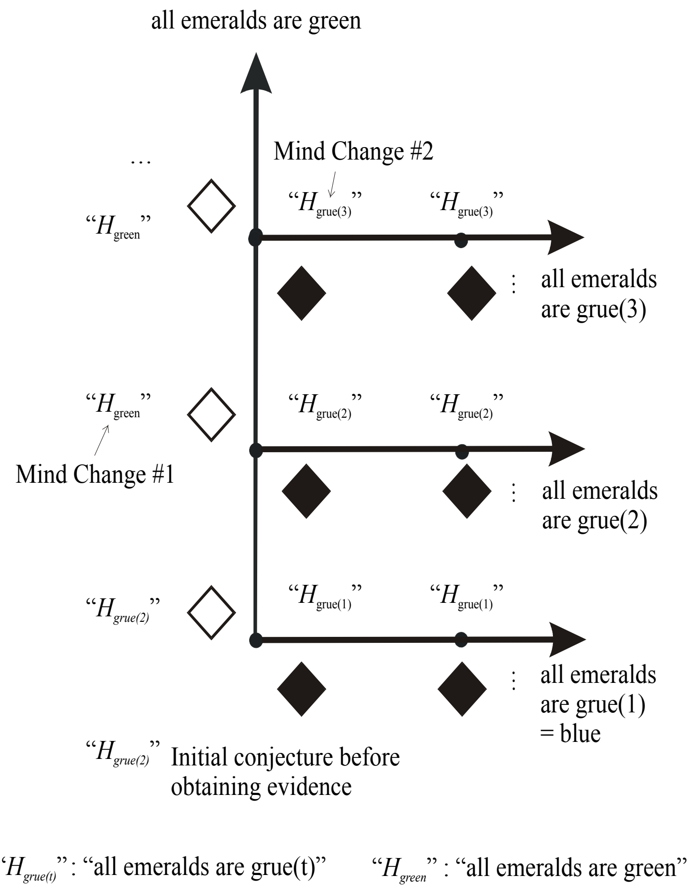

Figure 3: An unnatural projection rule may have to retract its conjecture twice—in this case, if all emeralds are $grue(3)$.

green emeralds followed by a blue one. Similarly, the characterization shows that the two crucial facts in the Occam problem are, first, that no finite number of experiments entails that the Ω particle will never be found, whereas second, if the Ω particle is discovered, this conclusively establishes its existence.

We saw that in the Occam problem and in the infinitely iterated Riddle of Induction, the results of our means-ends analysis depend only on *logical* relations of verification or falsification between evidence and hypothesis. The characterization theorem cited establishes that this is the case in any discovery problem. Since acceptable translations from one language into another preserve logical entailment, it follows that the methodological recommendations that means-ends analysis issues do not depend on the language in which we choose to express evidence and hypotheses. In particular, they do not change if we describe samples of emeralds and universal generalizations about their colour in a $grue - bleen$ vocabulary rather than the familiar $green - blue$ one. The result is that the optimal projection rule projects the translation of "all emeralds are green" (that is, the optimal rule projects "all emeralds are $grue(t)$ until time $t - 1$ and $bleen(t)$ thereafter" as long as that generalization is consistent with the evidence).

There is an important feature of our means-ends criteria that the two examples from this paper do not illustrate: In general, the goal of avoiding retractions—minimaxing retractions—conflicts with the goal of minimizing time-to-truth. The one-shot Riddle of Induction illustrates this tension (in fact, it is one of the simplest possible discovery problems featuring this conflict); see Figure 4.

In the one-shot Riddle, it is easy to identify the correct universal generalization with no mind changes simply by waiting until the "critical time" t. On the other hand, no data-minimal method can do so: By Theorem 6, a data-minimal method must project a universal generalization before time t. But whichever generalization a method elects to project, it might be falsified at the critical time t, forcing the method to retract its conjecture. In general, data-minimality requires an inquirer to make "bold" quick generalizations, whereas avoiding retractions leads to more "cautious" inferences.

One way to describe the conflict between minimizing convergence time and avoiding retractions is to ask how many mind changes a *data-minimal* reliable method might require, and contrast this with the number of mind changes required by a method that does not minimize convergence time. [Schulte forthcoming] characterizes the structure of discovery problems that permit a data-minimal method to reliably find the correct hypothesis with a bounded number of mind changes. Together with the analogous results for methods that are not data-minimal, the two characterizations provide a precise measure of the extent to which the aim of avoiding retractions conflicts with minimizing time-to-truth in a given discovery problem. When this tension obtains, an inquirer must strike a subjective balance, as in any case of conflicting aims. But when there is a data-minimal method that minimaxes retractions, an inquirer can epistemically "have it all". In such cases, the methods that avoid retractions and minimize convergence time seem to have special intuitive appeal—witness the natural projection rule in the infinitely iterated Riddle of Induction and the Occam Rule in the Occam problem.

Data-minimal methods that minimax retractions are not only intuitively appealing, they also avoid a standard objection to the minimax principle: The minimax principle tells us to take the bird in the hand rather than the two in the bush, as the saying goes—but we may well be willing to take a chance on winning big. With regard to retractions, the ideal case is to converge to the truth with no retractions. But in the *best* case, data-minimal methods converge to the truth with no more retractions at *any* stage of inquiry, because

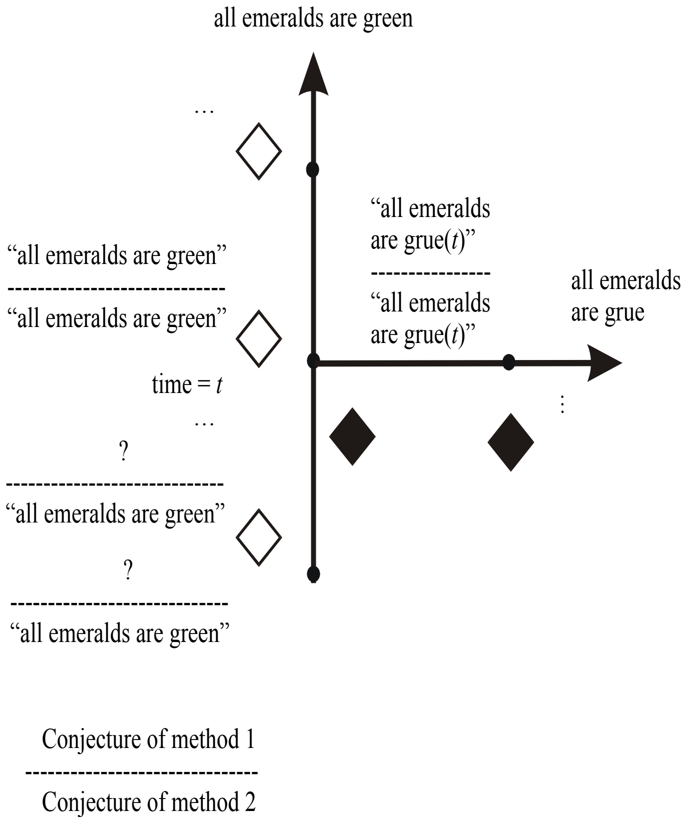

Figure 4: Conflicting Aims in the One-Shot Riddle of Induction: Method 1 minimizes convergence time (is data-minimal) and may have to retract its conjecture once. Method 2 does not minimize convergence time but never retracts its conjecture.

they always project their current conjecture. In terms of the old saying, data-minimal methods that minimax retractions take the bird in the hand—and then go for the two in the bush.

9. The Hierarchy of Hypothetical Imperatives for Inductive Inference

As we have seen, the standards of efficiency defined in Section 4 can differ greatly in the strength of the demands that they impose in inductive inquiry. Data-minimality imposes only the fairly mild constraint that a method must take its conjectures seriously at each stage of inquiry (Section 6). Retraction-minimality by contrast does not allow an inquirer to generalize beyond the data, a demand which is impossible to meet if inquiry is to find informative truths. Between these two extremes lies the criterion of minimaxing retractions, which does allow for genuine generalizations, but selects some over others, as we saw in the Occam problem and the Riddle of Induction. Is this an accident of the examples we have chosen, or are there general principles that relate the strength of one set of epistemic aims to another? In this section I will show that indeed, the subject of means-ends epistemology has a tidy structure: The various performance standards for inductive methods fall into an exact hierarchy of *feasibility*.

I say that a performance standard is **feasible** in a discovery problem posed by a collection of alternative hypotheses and given background knowledge if there is a reliable method for solving the problem that attains the standard in question. For example, minimaxing convergence time requires a reliable method to deliver the correct hypothesis by an a priori fixed deadline. So if minimaxing convergence time is feasible, there must be a deadline by which the evidence conclusively verifies one of the alternative hypotheses. I record this observation in the following proposition; the proof is in Section 11.

Proposition 12. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. Then there is a reliable method for the discovery problem $(\mathcal{H}, K)$ that minimaxes convergence time $\iff$ background knowledge K entails that there is a time n by which the evidence will conclusively establish one of the alternative hypotheses as the correct one.*

It is not hard to show that whenever this is possible, the other performance standards are feasible. In this sense minimaxing convergence time is the most stringent demand on inductive methods. It turns out that many performance standards are equally stringent, in the sense that one is feasible just in case the other is. Moreover, inductive methods can attain any one of the performance standards from this class just in case background assumptions guarantee that eventually the evidence will entail the correct hypothesis—in other words, just in case there is no problem of induction.

I say that an inductive method makes an *error* on a data stream ε at stage n iff on receiving the evidence from data stream ε up to stage n, the method produces a conjecture that is false on ε. Just as we can count the steps to convergence and the number of retractions, we can count the number of errors that a method commits on a given data stream. Then we may define the concepts of minimaxing error and admissibility with respect to error in the manner of Definitions 5 and 9. (The explicit definition is in Section 11.) This adds two more members to the class of performance standards that are feasible in exactly the same stringent conditions, according to our next theorem.

Theorem 13. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. The following conditions are equivalent:*

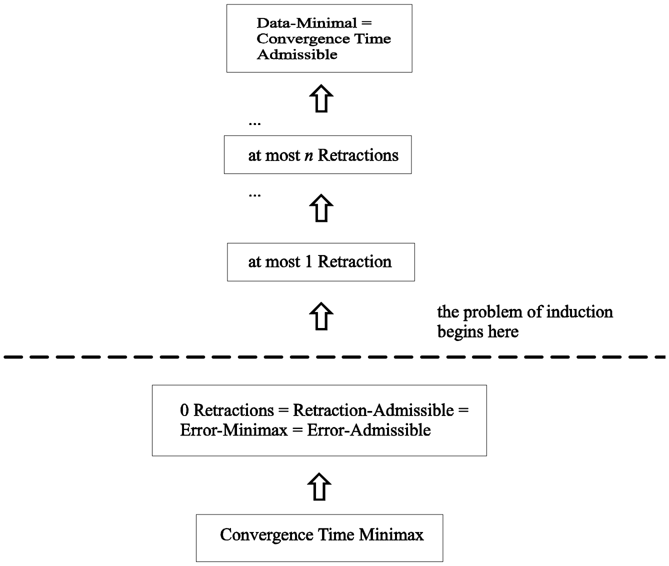

Figure 5: The Hierarchy of Hypothetical Imperatives for Inductive Inference

1. *The background knowledge K entails that eventually the evidence will conclusively establish one of the alternative hypotheses as the correct one.*

2. *There is a reliable discovery method for the discovery problem $(\mathcal{H}, K)$ that succeeds with no retractions.*

3. *There is a retraction–admissible reliable discovery method for the discovery problem $(\mathcal{H}, K)$.*

4. *There is an error-admissible reliable discovery method for the discovery problem $(\mathcal{H}, K)$.*

5. *There is a reliable discovery method δ for the discovery problem $(\mathcal{H}, K)$ that succeeds with a bounded number of errors.*

We can order problems that permit reliable methods to minimax retractions by whether they are solvable with at most 0 mind changes, at most 1 mind change,..., at most n mind changes etc., which yields an infinite subhierarchy of feasibility. Finally, whenever there is reliable method for a discovery problem, there is a data-minimal one (see [Schulte forthcoming]); thus data-minimality is the least demanding of our performance standards. Figure 5 displays the full feasibility hierarchy of standards of inductive success.

The hierarchy provides a *scale of inductive complexity* by which we can measure the difficulty of a discovery problem: A problem P is easier than a problem P' if a more stringent standard of inductive success is feasible in P than in P'. Thus the Occam problem and the Riddle of Induction are equally difficult in the sense that it is possible to reliably identify the correct hypothesis with at most one mind change in handling both of these problems. Problems such as determining whether human cognition is computable or whether matter is only finitely divisible go beyond the hierarchy from Figure 5: in the natural models of these problems, it is impossible to reliably find the right answer with a bounded number of mind changes (see [Kelly 1996, Kelly *et al.* 1997]). A problem of intermediate complexity would be to investigate a hypothesis of the form "there are exactly n different types of x" (n different elementary particles, n different forms of intelligent life, etc.), which requires no more and no less than two mind changes at worst. Thus we see again that means-ends analysis uncovers common structure among inductive problems that superficially appear quite different, revealing them to be equally complex.

Finally, the hierarchy explains why minimaxing retractions is a powerful principle for deriving short-run constraints on inductive method: unlike many of the other more demanding standards, the principle applies even when there is a genuine problem of induction. But unlike data-minimality, minimaxing retractions is stringent enough to yield strong constraints on how inductive inference ought to proceed.

The hierarchy of hypothetical imperatives does not cover the principle of maximizing expected utility. A full discussion of the differences between an analysis based on expected utility and the approach of this paper would lead into deep and complex issues in both decision theory and epistemology. I will leave this for another occasion, and just give some indication of the main differences between the two approaches.

One difference is that my treatment distinguishes different cognitive aims, rather than subsuming them under a single utility function. This allows us to study explicitly the objective relationships and trade-offs between different cognitive values.

Another reason, perhaps more fundamental, why expected utility does not appear in the feasibility hierarchy is that the hierarchy measures the *structural* (topological) complexity of a discovery problem, establishing a scale of ever more complex problems.[13] Probability measures, on the other hand, tend to collapse these differences in structural complexity.[14]

Lastly, decision theorists typically take probabilities to be *subjective* "degrees of belief". Thus whether a method maximizes the subjective expectation of a cognitive value will depend on a subjective ranking of possibilities by personal probability. In contrast, the project of means-ends methodology is to establish *objective* comparisons between the performance of inductive methods, with the design of optimal methods tied to the *intrinsic structure* of a given inductive problem. As we have seen, examining the admissibility and worst-case performance of inductive methods achieves this goal (see also [Kelly 1996], [Schulte forthcoming, Sec.7]).

[13] The characterization theorems in [Kelly 1996] and [Schulte forthcoming] provide explicit characterizations of this structural complexity. It is worth noting that the structural complexity of inductive problems is of a kind familiar from analysis and recursion theory.

[14] [Kelly 1996, Ch.13] shows exactly why this is so for the question of whether there is a reliable method for a given discovery problem: It turns out that for any discovery problem and any *countably additive* probability measure, there is an inductive method that has a probability of 1 in that measure of converging to the correct hypothesis. For the auxiliary cognitive values considered in this paper, there remains an interesting open question: Let a discovery problem and a probability measure be given. (Countably additive measures are an important special case.) Is it the case that if there is a reliable method that maximizes the expectation of one cognitive value considered in this paper, then there is a reliable method that maximizes the expectation of the others? For example, if a reliable method minimizes the expected number of mind changes, is there also a reliable method that minimizes the expected convergence time? If the answer is yes, we would have another example of how probabilities level differences in the feasibility of performance standards and the corresponding structural complexity of inductive problems.

10. Conclusion: A Map of Means-Ends Methodology

I will end my survey of means-ends epistemology with an outline of the subject, to put the results that I have covered in place, and to draw the reader's attention to some issues and applications that I have not had the space to explore in detail.

Means-ends analysis begins with the notion of an *inductive problem*. This paper considered discovery problems, which are defined by the inquirer's background assumptions about what the (epistemically) possible observation sequences are, and a set of mutually exclusive alternative hypotheses under investigation. The next step is to specify a *standard of success* for inductive methods, a combination of an epistemic value and an evaluation criterion. We then ask: When is a given standard of success *feasible*? The answer to this question is a set of results that characterize what the structure of an inductive problem must be like in order for a given performance standard to be feasible. These results establish a tidy *hierarchy of feasibility* that ranks inductive problems by the demands they place on empirical inquiry—a scale of inductive complexity. Problems of the same complexity share the same structure; more precisely, they share the same topology of possible data sequences and alternative hypotheses. We find these structural similarities across problems that on the surface appear quite different: For example, Goodman's Riddle of Induction, the Occamian question whether a given entity exists or not, and the task of determining whether a given reaction among elementary particles is possible or not, have the same topology of data streams.

The next question is what *methods* attain a given performance standard *when* the standard is feasible. The answer is another set of characterization theorems that describe exactly what the successful methods are like. These results allow us to reinterpret traditional proposals, such as axioms for belief revision and Popper's conjectures-and-refutations scheme, as means-ends recommendations: They turn out to be the means of choice for certain epistemic ends [Schulte 1997]. This clarifies the normative status of such methodological principles: They are good advice for agents with certain epistemic values, but agents with other aims need not, and sometimes *should* not, follow them.

When we apply the characterizations of successful methods to derive what the optimal inferences for a particular *problem* are, we find new means-ends justifications for interesting and plausible methodological recommendations. These applications include the natural projection rule in Goodman's Riddle of Induction, a version of Occam's Razor, and a instrumentalist interpretation of the role of conservation laws and hidden particles in particle physics [Schulte 1997].

This paper has examined some of the most natural and interesting epistemic aims. We can equally well ask what means would be optimal for achieving other cognitive objectives. [Schulte 1997] carries out this investigation for content, avoiding error and producing succinct (finitely axiomatizable) theories. In principle, all we need to provide a means-ends analysis for further theoretical goals is a sufficiently clear description of the goals in question.

Acknowledgments

Kevin Kelly, John King-Farlow, Bernard Linsky and Mohan Matthen provided much helpful discussion and comments on drafts of this paper. Kevin Kelly made extensive suggestions for improving the exposition.

11. Proofs

Let $[e]$ be the empirical proposition that the evidence e has been observed; that is, $[e]$ is the set of all data streams extending e.

Proposition 10. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. Then there is a reliable retraction-minimal method for the discovery problem $(\mathcal{H}, K) \iff$ on every data stream ε, there is a time n such that $[\varepsilon|n]$ and K entail the correct hypothesis H from $\mathcal{H}$.*

Proof. ($\Leftarrow$) Given the right-hand side, the "skeptical" method δ whose theory is always exactly the evidence conjoined with background knowledge K reliably identifies a true hypothesis from $\mathcal{H}$ and never retracts its beliefs.

($\Rightarrow$) Let δ be a reliable retraction-minimal discovery method for $\mathcal{H}$. Because beginning with the trivial conjecture $\delta(\emptyset) = K$ does not increase the number of mind changes of a method on any data stream, we may without loss of generality assume that $\delta(\emptyset) = K$. Now suppose that on some data stream ε consistent with K, δ changes its mind at least once. Then there is a first time $m_0 > 0$ at which δ makes a non-vacuous conjecture, that is, $\delta(\varepsilon|m_0)$ is consistent and entails some hypothesis H from $\mathcal{H}$, and a first time $m_1 > m_0$ at which δ makes a mind change, that is, $\delta(\varepsilon|m_1)$ does not entail H or $\delta(\varepsilon|m_1) = \emptyset$. Now the following method δ' dominates δ with respect to mind changes given K: If $\varepsilon|m_0 \subseteq e \subset \varepsilon|m_1$, let $\delta'(e) = K$. Otherwise $\delta'(e) = \delta(e)$. Then δ' conjectures K along ε until $\varepsilon|m_1$, so δ' changes its mind along ε exactly one less time than δ. And clearly δ' never uses more mind changes than δ does. So δ is dominated with respect to mind changes given K. This shows that if there is a reliable retraction-minimal method for $\mathcal{H}$ given K, then there is a reliable method δ that never changes its mind along any data stream in K. Such a method δ never entails more than the evidence. For suppose that $[e] \cap K$ does not entail $\delta(e)$. Then there is a data stream ε consistent with e and K on which $\delta(e)$ is false. So to succeed on ε, δ must change its mind at least once (after e), contrary to supposition. Since δ never entails more than the evidence and is reliable, eventually the evidence and background knowledge must conclusively establish which alternative in $\mathcal{H}$ is the true one, no matter what it is. ■

Proposition 14. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. Then there is a reliable method for the discovery problem $(\mathcal{H}, K)$ that minimaxes convergence time $\iff$ there is a time n such that on every data stream ε, $[\varepsilon|n]$ and K entail the true hypothesis H in $\mathcal{H}$.*

Proof. ($\Leftarrow$) Suppose that there is a deadline n by which background knowledge and the data entail which hypothesis is correct. Without loss of generality, assume that n is the earliest such time. Then a reliable method δ can simply conjecture the evidence until time n . In the worst case (and in the best case), δ converges to the correct hypothesis by time n. To show that δ minimaxes convergence time, I establish that no other method δ' can achieve a better guarantee on the time that it requires to find the truth. This is immediate if $n = 0$. Otherwise, let ε be any data stream such that $K, \varepsilon|n-1$ do not entail which hypothesis from $\mathcal{H}$ is correct. Since we chose n to be the earliest time by which background knowledge and the evidence always entail the correct hypothesis, there is such a data stream. If $\delta'(\varepsilon|n-1)$ is inconsistent or does not entail a hypothesis $H \in \mathcal{H}$, then clearly δ' does not converge on ε before time n. Otherwise suppose that $\delta'(\varepsilon|n-1)$ entails some hypothesis H. Then by the choice of ε and n, there is another hypothesis H' in $\mathcal{H}$

such that H' is true on some extension τ of $\varepsilon|n-1$, consistent with K, and $\delta'(\varepsilon|n-1)$ does not entail H'. Hence δ' converges to the correct hypothesis on τ only after time n. Therefore every method requires, on some data stream consistent with background knowledge K, at least n pieces of evidence before settling on a correct hypothesis, and our "wait-and-see" method minimaxes convergence time.

($\Rightarrow$) I show the contrapositive. Suppose that for every bound n , there is a data stream ε such that $K, \varepsilon|n$ do not entail which hypothesis in $\mathcal{H}$ is correct. By the same argument as for the converse implication, this means that for every reliable method δ, and every bound n, there is some data stream ε consistent with K on which δ converges to the correct hypothesis only after time n. Hence there is no global bound on the convergence time of any reliable method δ. ■

Definition 16. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be background knowledge. In the discovery problem $(\mathcal{H}, K)$, an inductive method δ* ***dominates*** *another inductive method δ'* ***with respect to error*** $\iff$

1. *on every data stream ε consistent with K, δ makes no more errors than δ', and*
2. *for some data stream ε consistent with K, δ makes fewer errors than δ' does.*

An inductive method δ is ***error-minimal*** *for a discovery problem $(\mathcal{H}, K) \iff \delta$ is not dominated in $(\mathcal{H}, K)$ with respect to error by another method δ' that is reliable for $(\mathcal{H}, K)$.*

For given background knowledge K, let $MaxErr(\delta, K) = \max\{$the number of errors that δ commits on $\varepsilon : \varepsilon \in K\}$ be the maximum number of errors that δ makes on any data stream consistent with the given background knowledge. I say that δ **succeeds with a bounded number of errors** given K iff there is a finite number n such that $MaxErr(\delta, K) < n$. As in the case of discovery with bounded mind changes, if a method δ minimaxes errors, its succeeds with a bounded number of errors (namely the lowest possible such bound).

Theorem 15. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. The following conditions are equivalent:*

1. *The background knowledge K entails that eventually the evidence will conclusively establish one of the alternative hypotheses as the correct one.*
2. *There is a reliable discovery method for the discovery problem $(\mathcal{H}, K)$ that succeeds with no retractions.*
3. *There is a retraction–admissible reliable discovery method for the discovery problem $(\mathcal{H}, K)$.*
4. *There is an error-admissible reliable discovery method for the discovery problem $(\mathcal{H}, K)$.*
5. *There is a reliable discovery method δ for the discovery problem $(\mathcal{H}, K)$ that succeeds with a bounded number of errors.*

Proof. The equivalence of claims 1 and 2 follows from the arguments in the proof of Proposition 8. By the same Proposition, claim 3 is equivalent to claim 1. I prove the equivalence of claims 4 and 5 with claim 1.

$(1 \Rightarrow 4, 5)$ If background knowledge K guarantees that eventually the evidence will conclusively entail the correct hypothesis, the method δ that conjectures nothing but the evidence reliably identifies the correct hypothesis from H and never makes an error. Hence δ both is error-admissible and minimaxes error.

$(1 \Leftarrow 4)$ Let δ be an error-admissible method. Suppose that δ makes an error on some data stream ε consistent with K, say at time k. Then the following method δ' dominates δ in error: δ' agrees with δ everywhere but at $\varepsilon|k$, and $\delta'(\varepsilon|k) = K \cap \varepsilon|k$. Then δ' makes strictly fewer errors than δ on ε, and never makes more. So the only error-admissible methods are those that never make an error on any data stream consistent with K. Hence no error-admissible method makes a conjecture that goes beyond the evidence. As we saw in the proof of Proposition 8, such a method is reliable only if background knowledge K guarantees that the evidence eventually entails the true hypothesis.

$(1 \Leftarrow 5)$ I show the contrapositive. Suppose that background knowledge K does not guarantee that eventually the evidence will conclusively establish one of the alternative hypotheses as the correct one. Then there is some hypothesis H in $\mathcal{H}$ true on some data stream ε consistent with H and K such that H is never entailed along ε given K. Now consider some possible bound n on the number of errors that a reliable discovery method δ may make on any data stream in K. Since δ is reliable, δ eventually converges to H on ε, say at time k. Thus δ conjectures H on $\varepsilon|k, \varepsilon|k+1, ..., \varepsilon|k+n$. Since H is never entailed along ε with respect to background knowledge K, there is data stream τ extending $\varepsilon|k+n$, consistent with K, on which H is false. Thus on τ, δ makes at least $n+1$ errors. Since this argument applies to any bound n, there is no bound on the number of errors that δ might make on a data stream. Thus there is no reliable discovery method that minimaxes errors unless background knowledge K guarantees that the evidence conjoined with K will eventually entail one of the alternative hypotheses. ■

References

[Bub 1994] Bub, J. (1994). "Testing Models of Cognition Through the Analysis of Brain-Damaged Performance," *British Journal for the Philosophy of Science.* 45:837–855.

[Carnap 1962] Carnap, R. (1962). "The Aim of Inductive Logic", in *Logic, Methodology and Philosophy of Science*, ed. E. Nagel, P. Suppes and A. Tarski. Stanford: Stanford University Press.

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

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

[Gärdenfors 1988] Gärdenfors, P. (1988). *Knowledge In Flux: modeling the dynamics of epistemic states.* Cambridge: MIT Press.

[Glymour 1994] Glymour, C. (1994). "On the Methods of Cognitive Neuropsychology," *British Journal for the Philosophy of Science.* 45:815–835.

[Glymour and Kelly 1990] Glymour, C. and Kelly, K. (1990). "Why You'll Never Know if Roger Penrose is a Computer," *Behavioral and Brain Sciences.*Vol. 13.

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

[Goodman 1983] Goodman, N. (1983). *Fact, Fiction and Forecast.* Cambridge, MA: Harvard University Press.

[Howson 1997] Howson, C. (1997). "A Logic of Induction," *Philosophy of Science.* 64:268–290.

[Howson and Urbach 1989] Howson, C. and Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach.* La Salle, Ill: Open Court.

[Hume 1984] Hume, D. (1984). *An Inquiry Concerning Human Understanding,* ed. C.Hendell. New York: Collier.

[James 1982] James, W. (1982). "The Will To Believe," in *Pragmatism.* ed. H.S. Thayer. Indianapolis: Hackett.

[Juhl 1994] Juhl, C. (1994). "The Speed-Optimality of Reichenbach's Straight Rule of Induction," *British Journal for the Philosophy of Science* 45:857–863.

[Juhl 1997] Juhl, C. (1997). "Objectively Reliable Subjective Probabilities," *Synthese* 109:293-309.

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

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

[Kelly *et al.* 1994] Kelly, K., Juhl, C. and Glymour, C. (1994). "Reliability, Realism, and Relativism", in *Reading Putnam*, ed. P. Clark. London: Blackwell.

[Kelly *et al.* 1997] Kelly, K., Schulte, O. and Juhl, C. (1997). "Learning Theory and the Philosophy of Science," *Philosophy of Science.* 64: 245–267.

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

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

[Kuhn 1957] Kuhn, T. (1957). *The Copernican Revolution.* Cambridge, MA: Harvard University Press.

[Kuhn 1970] Kuhn, T. (1970). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

[Laudan 1977] Laudan, L. (1977). *Progress and Its Problems.* Berkeley: University of California Press.

[Newell and Simon 1976] Newell, A. and Simon, H. (1976). "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM* 19: 113–126.

[Osherson *et al.* 1986] Osherson, D., Stob, M. and Weinstein, S. (1986). *Systems That Learn.* Cambridge, Mass: MIT Press.

[Osherson and Weinstein 1988] Osherson, D. and Weinstein, S. (1988). "Mechanical Learners Pay a Price for Bayesianism," *Journal of Symbolic Logic* 53: 1245–1252.

[Popper 1968] Popper, K. (1968). *The Logic Of Scientific Discovery.* New York: Harper.

[Putnam 1963] Putnam, H. (1963). "'Degree of Confirmation' and Inductive Logic," in *The Philosophy of Rudolf Carnap,* ed. A. Schilpp. La Salle, Ill: Open Court.

[Putnam 1965] Putnam, H. (1965). "Trial and Error Predicates and a Solution to a Problem of Mostowski," *Journal of Symbolic Logic* 30: 49–57.

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

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

[Salmon 1991] Salmon, W. (1991). "Hans Reichenbach's Vindication of Induction," *Erkenntnis.* 35:99–122.

[Schulte and Juhl 1997] Schulte, O. and Juhl, C. (1997). "Topology as Epistemology". *The Monist* 79:141–148.

[Schulte 1997] Schulte, O. (1997). "Hard Choices in Scientific Inquiry". Doctoral Dissertation, Department of Philosophy, Carnegie Mellon University.

[Schulte forthcoming] Schulte, O. (forthcoming). "The Logic of Reliable and Efficient Inquiry," *The Journal of Philosophical Logic.*

[Sextus Empiricus 1985] Sextus Empiricus (1985). *Selections from the Major Writings on Skepticism, Man and God,* ed. P. Hallie, trans. S. Etheridge. Indianapolis: Hackett.

[Sharma *et al.* 1997] Sharma, A., Stephan F. and Ventsov, Y. (1997). "Generalized Notions of Mind Change Complexity", *Proceedings of the Conference of Learning Theory (COLT) 1997.*

[Spohn 1988] Spohn, W. (1988). "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States," *Causation in Decision, Belief Change and Statistics II*, ed. Skyrms, B. and Harper, W. Dordrecht: Kluwer.

[Van Fraassen 1980] Van Fraassen, B. (1980). *The Scientific Image.* Oxford: Clarendon Press.