The Logic of Reliable and Efficient Inquiry

Oliver Schulte

October 3, 2001

Abstract. This paper pursues a thorough-going instrumentalist, or means-ends, approach to the theory of inductive inference. I consider three epistemic aims: convergence to a correct theory, fast convergence to a correct theory and steady convergence to a correct theory (avoiding retractions). For each of these, two questions arise: (1) What is the structure of inductive problems in which these aims are feasible? (2) When feasible, what are the inference methods that attain them? Formal learning theory provides the tools for a complete set of answers to these questions. As an illustration of the results, I apply means-ends analysis to various versions of Goodman's Riddle of Induction.

1. Means-Ends Solutions for Problems of Induction

Empirical inquiry begins with questions about the world, and uses evidence to find answers. One of the major issues of epistemology is how inquiry should go about its task. This question leads immediately to the topic of *inductive inference*, how to generalize beyond the available evidence to obtain an answer to the issues under investigation. Hume's Problem of Induction and Goodman's Riddle are two classic, sharply focused illustrations of the problems associated with inductive inference. Hume asks what the foundation of beliefs is that do not follow with deductive certainty from the available evidence. Goodman's Riddle challenges us to find *epistemic* principles for choosing among various alternative generalizations. The fact that we are familiar with those predicates that occur frequently in English, for example, does not count as an answer to Goodman's challenge [Sober 1994].

A well-developed response to Hume's problem is the fallibilist proposal that, although we may never be certain of our generalizations, we can nonetheless find the right answer to the questions of inquiry in the long run, or in "the limit of inquiry". This conception of empirical success inspired the work of Peirce, James, Reichenbach, Putnam, and others. This approach to induction aims for a *means-ends* standard of inductive rationality: We ought to use those inference methods that attain the goals of inquiry. The task of methodology, then, is to determine the best means towards our epistemic aims. This paper studies three particularly interesting epistemic ends: Convergence to a correct theory, fast convergence to a correct theory, and steady convergence to a correct theory, that is, avoiding retracting one's theories as much as possible. We may think of speed and avoiding retractions as standards of *efficiency* for inductive methods that find the truth in the limit of inquiry.

Two questions arise: (1) What is the structure of inductive problems in which it is possible to attain these cognitive ends—that is, when are they *feasible*? (2) When a given cognitive aim is feasible, which methods attain it? I draw on the tools of *formal learning theory* to give a complete set of answers to these questions. Formal learning theory is a highly developed mathematical framework for carrying out the means-ends analysis of inductive problems. The results of the analysis are rewarding: the efficiency criteria that are the subject of this paper provide principled methodological recommendations for drawing general conclusions from given data. To illustrate, I analyze several

versions of Goodman's Riddle of Induction, and show that the efficient inference methods project the natural generalization ("all emeralds are green" rather than "all emeralds are grue"). This is no accident: It turns out that Goodman's Riddle has exactly the structure characteristic of efficient inductive inquiry. Other interesting inductive problems share this structure, for example the Occam-like problem of determining whether a given entity exists or not, and inferring theories of reactions among elementary particles [Schulte forthcoming, Schulte 1997]. This paper describes the structure common to all inductive problems that permit efficient inquiry, and specifies the general form of efficient inference methods. These results show that means-ends analysis does not depend on the language in which evidence and hypotheses are described.

Let us begin with some fundamental notions from learning theory.

2. Discovery Problems

Learning theory studies several broad classes of inductive problems, such as making predictions, testing hypotheses, inferring general theories, and others (for an up-to-date survey, see [Kelly 1996]). I will examine the following type of problem: Consider a collection $\mathcal{H}$ of mutually exclusive alternative hypotheses under investigation. Given a piece of evidence e, which of the alternative hypotheses in $\mathcal{H}$ should the agent conjecture? Following Popper's and Kelly's usage, I refer to such problems as *discovery problems* [Popper 1968], [Kelly 1996].[1] The general definition of a discovery problem is as follows. Let E be a set of evidence items or experimental outcomes (observations of the colours of swans, ravens, emeralds, etc., particle reactions, positions of planets, and so on). A **data stream** is an infinite discrete sequence of evidence items from E. 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 ε is a data stream, then ε_n denotes the n-th observation made along ε, and $\varepsilon|n$ denotes the first n observations in the data stream; see Figure 1. An **empirical proposition** is a set of data streams. For example, since the empirical content of the hypothesis "all emeralds are green" is just the data stream featuring only green emeralds—call it τ—I identify "all emeralds are green" with the empirical proposition $\{\tau\}$. If ε is a data stream, H an empirical hypothesis, and $\varepsilon \in H$, I say that H is **correct on**, or true of, ε.[2] 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, if a hypothesis $H \in \mathcal{H}$ is correct on a data stream ε, then no other hypothesis $H' \in \mathcal{H}$ is correct on ε.*

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.[3] Epistemologists have considered more complicated inductive methods, for example ones that revise "degrees of belief" in light of new evidence (as Bayesian methods do; see for example

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

[2] 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.

[3] This differs from Kelly's treatment of discovery problems, which requires empirical methods to produce one of the hypotheses under investigation at each stage [Kelly 1996, Ch.9].

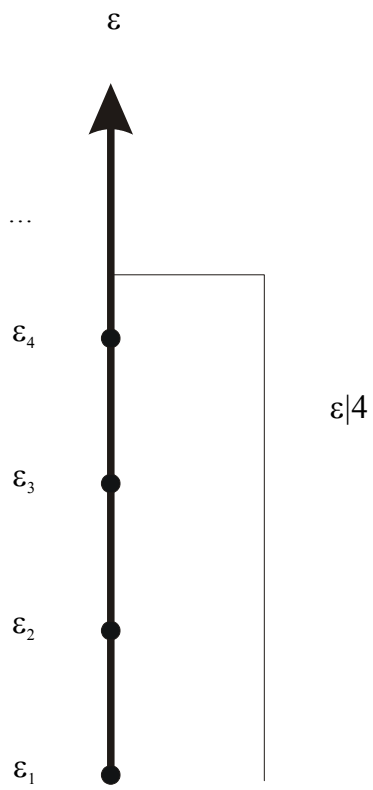


Figure 1: A Data Stream ε and Associated Concepts

[Howson and Urbach 1989]). In principle, means-ends analysis can guide agents in revising any epistemic state.[4]

Many important inductive problems from a variety of settings fit the formalism of discovery problems. To name a few, language learning [Osherson *et al.* 1986]; parameter estimation and "model selection" in statistics; and inferring theories in scientific disciplines such as particle physics [Schulte 1997] and cognitive neuropsychology [Glymour 1994], [Bub 1994]. Some examples of discovery problems will illustrate the general notions introduced in this section, as well as many of the points about inductive method that I shall be making. It turns out that Goodmanian "Riddles of Induction" serve this purpose well.

3. Riddles of Induction

3.1. Green and Grue. 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.

Goodman's 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" n, a unique colour for the emerald examined at n. Thus the empirical content of the claim "all emeralds are green" is that at each time, the emerald examined at that time is green; that is, the empirical content is the singleton $\{\varepsilon\}$, where $\varepsilon_n = \mathit{green}$ for all n. The empirical content of "all emeralds are grue" is that at times earlier than the critical time t, the emerald examined at that time is green, and that at time t and later times, the emerald examined is blue. That is, the empirical content of "all emeralds are grue" is the singleton $\{\tau\}$, where τ is the data stream such that $\tau_n = \mathit{green}$ if $n < t$ and $\tau_n = \mathit{blue}$ otherwise. Figure 2 shows these two data streams. If not all emeralds are examined, then "all emeralds are green (grue)" may be empirically adequate yet false, namely if all examined emeralds are green (grue) but the unexamined ones are not. In what follows, I shall not be concerned with this possibility.[5] In particular, for the sake of more natural expression, I will use the term "all emeralds" to implicitly mean the same as "all examined

[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 conditioning see for example, [Earman 1992, Ch.9], [Osherson and Weinstein 1988], [Kelly *et al.* 1997], [Kelly and Schulte 1995], [Juhl 1997].

[5] There are several reasons why one might want to neglect it. 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 only with the empirical adequacy of a theory, along the lines of [Van Fraassen 1980].

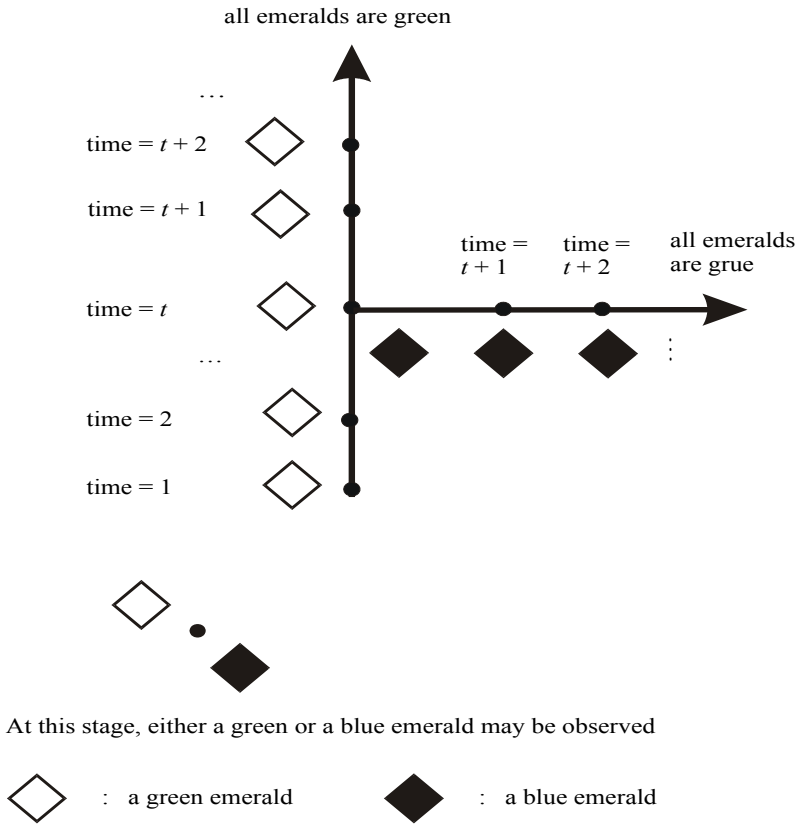


Figure 2: The empirical content of "all emeralds are green" and of "all emeralds are grue", with critical time t.

emeralds". For example, I will say that the conjecture "all emeralds are green" is correct, or true, on the data stream ε along which only green emeralds are found forever.

We obtain different versions of Goodman's Riddle of Induction by enlarging the set of alternative hypotheses with other gruesome predicates. Let me remark at the outset that my aim is not to interpret Goodman's text, and I make no claim that any of the discovery problems that I shall describe below are exactly what he had in mind. Indeed, the full strength and generality of the theory presented in this paper becomes most apparent when we apply it to a host of Riddles of Induction.

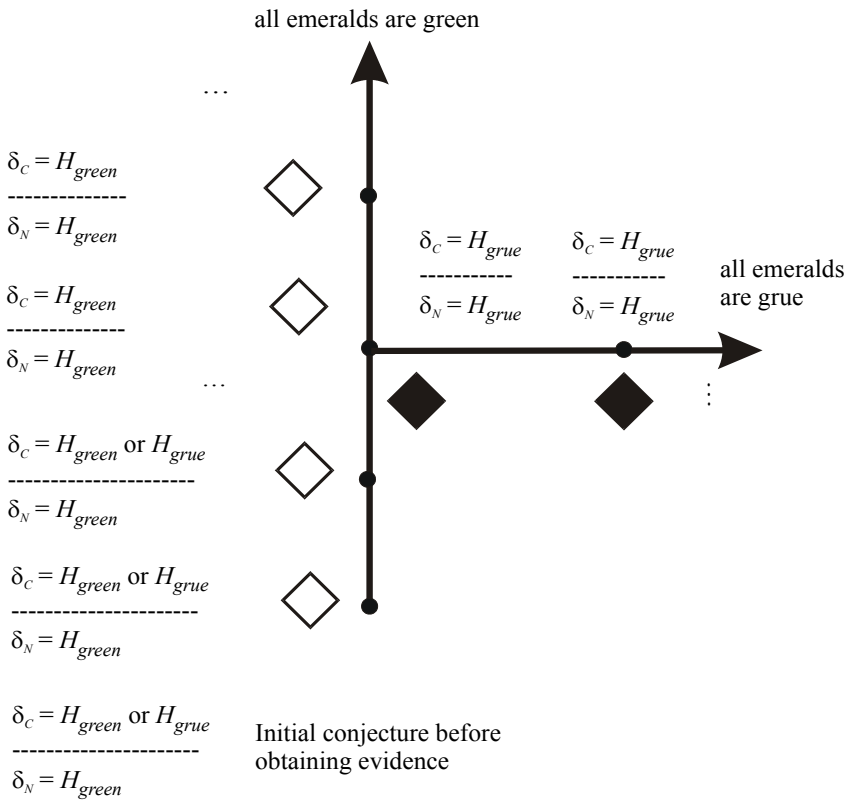
As is well-known, Goodman showed that "green" (and similarly, "blue") may be defined in terms of "grue" and "bleen". Suppose that t is the critical time, such that an emerald is grue (bleen) iff the emerald is examined before time t and found to be green (blue), or the emerald is examined at or after time t and found to be blue (green). Then an emerald is green iff it is examined before time t and found to be grue, or the emerald is examined at or after time t and found to be bleen. In the grue-bleen reference frame, we would define the empirical content of the hypothesis "all emeralds are green" to be the singleton ε such that $\varepsilon_n = \mathit{grue}$ if $n < t$, and $\varepsilon_n = \mathit{bleen}$ if $n \geq t$. As will become apparent, it does not matter to my methodological analysis whether we use the green-blue or the grue-bleen pair of predicates to define the relevant data streams. (I will return to this point in Section 8.) For ease of exposition solely, I shall continue to define "grue" predicates in terms of "green" and "blue".

3.2. The One-Shot Riddle of Induction. One reading of Goodman's Riddle is that we are taking two colour predicates under consideration: The familiar "green" and the unfamiliar "grue", where grue is defined with respect to some fixed "critical time" t. Thus we have two alternative hypotheses, "all emeralds are green" and "all emeralds are grue", or as Goodman would say, two candidates for "projection". If these are the only two serious possibilities, we may take it as background knowledge that either all emeralds are green or that all emeralds are grue; see Figure 2. I refer to this version of the Riddle of Induction as the **one-shot Riddle of Induction**. Let's consider Goodman's challenge of what to project before the critical time in the one-shot Riddle. We can sidestep the challenge by waiting until the critical time t before projecting anything. If the emerald examined at the critical time t is green, we know for certain that not all emeralds are grue. If the emerald examined at the critical time t is blue, we know for certain that not all emeralds are green.

An empirical proposition P **entails** another empirical proposition P' just in case $P \subseteq P'$. If e is a finite evidence sequence, $[e]$ denotes the empirical proposition containing all and only those data streams with e as an initial segment. Now we may define the **cautious** inference rule δ_C as follows, for any finite data sequence e (sample of examined emeralds):

1. if $[e]$ entails that H_{grue} is false, $\delta_C(e) = H_{green}$;
2. if $[e]$ entails that H_{green} is false, $\delta_C(e) = H_{grue}$;
3. else $\delta_C(e) = H_{grue} \cup H_{green}$.

Figure 3 illustrates the cautious inference rule. This inference method shows that the one-shot Riddle of Induction is a particularly easy discovery problem: An inquirer can eventually decide conclusively which of the two alternatives ("all emeralds are green" and "all emeralds are grue") is true, and we can specify in advance a deadline by which the



δ_C : the cautious projection rule

δ_N : the natural projection rule

H_{green} : "all emeralds are green"

H_{grue} : "all emeralds are grue"

Figure 3: The Cautious and the Natural Projection Rules in the One-Shot Riddle of Induction.

evidence settles the matter (namely the critical time t). This reflects the fact that the one-shot Riddle of Induction does not pose a problem of induction as classical writers such as Sextus Empiricus and Hume conceived of it. On the traditional conception, the essence of the problem is that no matter how much evidence an inquirer may have obtained, further observations may refute her generalizations.[6] The one-shot Riddle of Induction is not a problem of induction in this sense because there is a finite amount of evidence that decides which of the two alternative generalizations is correct, namely the evidence up to the critical time t.

Should an inquirer be cautious and wait until the critical time before projecting a generalization about the colour of emeralds? The cautious method δ_C leads us to the right generalization without errors along the way, and there is no risk that at the critical time, we may have to take back a conjecture that is refuted at that time. If the critical time is "tomorrow", or, say, "a week from now", this is an attractive way to proceed. But if the critical time is "twenty years from now" it may seem too long to wait for conjectures until then. One way of putting the point precisely is that the cautious method δ_C avoids errors and retractions, but is slow to settle on the right generalization about the colour of emeralds. We see this clearly if we contrast δ_C with the natural projection rule δ_N that projects "all emeralds are green" if all emeralds examined so far are green.

An empirical proposition P is **consistent** with another empirical proposition P' iff $P \cap P' \neq \emptyset$. Now we may formally define δ_N for all finite evidence sequences e (samples of green emeralds):

1. if $[e]$ is consistent with $H_{green}, \delta_N(e) = H_{green}$;
2. else $\delta_N(e) = H_{grue}$.

Figure 3 shows the natural projection rule. In what sense is the natural projection rule δ_N faster than the cautious method δ_C? If all emeralds are grue, then both δ_N and δ_C settle on the right generalization (with certainty) at the critical time t, but not before then. If all emeralds are green, again, δ_C settles on the right generalization at the critical time t, but not before then; in contrast, the natural projection rule δ_N conjectures immediately that all emeralds are green and thus settles on the truth at once (albeit without certainty). So the natural projection rule never converges after the cautious rule. And if all emeralds are green, the natural projection rule converges faster than the cautious rule (at time 1 rather than at the critical time t). In other words, the natural projection rule *dominates* the cautious one with respect to time-to-truth. I shall define this notion of dominance precisely in Section 5.

Thus in the one-shot Riddle of Induction, an inquirer has to make a choice between two conflicting values: avoiding error and retractions, on the one hand, and converging to the truth as fast as possible, on the other. In Section 7, I shall characterize the class of discovery problems in which there is a tension between these groups of desiderata, and the extent of the conflict between them in a given problem. The one-shot Riddle is an instance of a discovery problem in which minimizing convergence time conflicts with avoiding errors

[6] Sextus writes: "[The dogmatists] claim that the universal is established from the particulars by means of induction. If this is so, they will effect it by reviewing either all the particulars or only some of them. But if they review only some, their induction will be unreliable, since it is possible that some of the particulars omitted in the induction may contradict the universal. If, on the other hand, their review is to include all the particulars, theirs will be an impossible task, because particulars are infinite and indefinite. Thus it turns out, I think, that induction, viewed from both ways, rests on a shaky foundation" [Sextus Empiricus 1985, p.105].

and retractions. In fact, it is the *simplest possible* problem of this kind: There are only two possible data streams, and only two possible alternative hypotheses; any fewer data streams or hypotheses, and we have a trivial inference problem in which background knowledge entails the correct hypothesis a priori, before any evidence is obtained.

3.3. Finitely Iterated Riddles of Induction. If we are willing to investigate one grue predicate with critical time t, why not another with critical time t'? Another plausible version of the Riddle includes several gruesome predicates as candidates for projection, each defined by a specific critical time. I call such variants **iterated** Riddles of Induction. We may iterate the Riddle finitely or infinitely often; let's consider the finite case first. Suppose we iterate the Riddle m times, such that we include grue predicates with critical time $1, 2, ..., m$ as alternatives to green. Denote the grue predicate with critical time t by $grue(t)$. The hypothesis that "all examined emeralds are $grue(t)$" is correct on a data stream τ just in case $\tau_n = green$ if $n < t$ and $\tau_n = blue$ otherwise, for all $n > 0$. I denote this hypothesis by $H_{grue(t)}$. The **m-iterated Riddle of Induction** takes the alternative hypotheses to be "all emeralds are green" and "all emeralds are $grue(t)$", for any natural number $t \leq m$. If we assume as background knowledge that one of the alternative hypotheses is true, the m-iterated Riddle of Induction is the discovery problem $(\mathcal{H}^m, K^m)$, where $\mathcal{H}^m = \{H_{grue(t)} : 0 \leq t \leq m\} \cup \{H_{green}\}$, and $K^m = \bigcup \mathcal{H}^m$. Figure 4 illustrates the possible data streams and alternative hypotheses in the 3-iterated Riddle of Induction.

Like the one-shot version, no finitely iterated Riddle poses a problem of induction in the classical sense either: an inquirer might wait until the last critical time—time m—and then determine with certainty which of the possible generalizations about colour predicates is correct. We may define this cautious procedure δ_C^m for the m-iterated Riddle of Induction by $\delta_C^m(e) = K^m \cap [e]$, for all finite data sequences e. As in the one-shot Riddle, this cautious method eventually determines the correct generalizations about emerald colours, no matter what the correct generalization is, and it does so without any errors or retractions. Let's compare δ_C^m with the natural projection method δ_N^m for the m-iterated Riddle. The natural projection method δ_N^m is defined as follows, for any finite data sequence e (sample of emeralds):

1. $\delta_N^m(e) = H_{green}$ if e is consistent with H_{green};
2. otherwise $\delta_N^m(e) = K^m \cap [e]$.

The natural projection rule δ_N^m determines the correct generalization about emerald colours by time m at the latest, as the cautious procedure δ_C^m does. But the natural projection rule converges before the cautious one if all emeralds are green. Are there other projection rules that match the speed of the natural one? Indeed there are: Consider a rule δ_t^m that projects any $grue(t)$ predicate as long as the evidence is consistent with $grue(t)$, that is, as long as all examined emeralds are $grue(t)$. For the sake of definiteness, let's say that δ_t^m follows the natural projection rule δ_N^m if the evidence falsifies $H_{grue(t)}$. If all emeralds are in fact $grue(t)$—if $H_{grue}(t)$ is correct—the rule δ_t^m makes the correct conjecture from time 1 onward, whereas the natural projection rule δ_N^m conjectures that all emeralds are green until time $t - 1$, and only then changes its mind[7] to conclude that all emeralds are $grue(t)$. On the other hand, if all emeralds are green, then the natural

[7] 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.

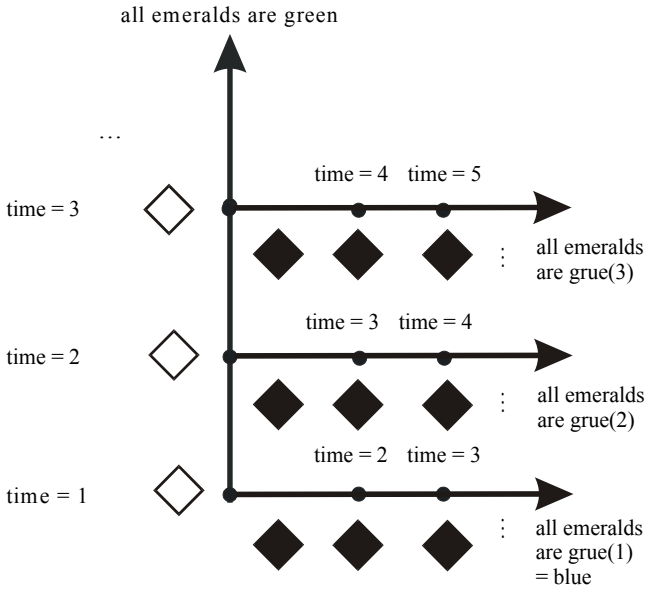


Figure 4: The 3-Iterated Riddle of Induction

projection rule is faster than any of the $grue(t)$ rules. Thus the natural projection rule δ_N^m and the δ_t^m rules do not dominate each other, as far as time-to-truth is concerned. Yet the natural projection rule does better than all but one of the δ_t^m rules at *avoiding retractions*—in the sense that the natural projection rule may change its mind at most once, whereas the unnatural projection rules might have to change their mind twice. The one exception is the rule δ_m^m rule that starts projecting $H_{grue(m)}$, that is, the colour predicate with the latest critical time; this rule also changes its mind at most once. (For a specific illustration of these observations, the reader may wish to trace the conjectures of δ_N^3, the natural projection rule in the three-iterated Riddle of Induction—diagrammed in Figure 4—and its alternatives δ_2^3 and δ_3^3.) Section 6 investigates avoiding retractions as a performance criterion in detail.

The upshot is that, in the finitely m-iterated Riddle of Induction, efficiency criteria such as minimizing convergence time and avoiding retractions select those rules that project either "all emeralds are green" or "all emeralds are $grue(m)$" so long as these hypotheses are consistent with the evidence, but rule out all other generalizations under consideration.

3.4. The Infinitely Iterated Riddle of Induction. There is no reason why we should consider gruesome predicates only up to a certain last critical time m. After all, when that critical time m comes, an inquirer may well consider the possibility that all emeralds are $grue(m+1)$. Accordingly, the **infinitely iterated Riddle of Induction** includes all $grue(t)$ predicates as candidates for projection, for any natural number t. Formally, the infinitely iterated Riddle of Induction is the discovery problem $(\mathcal{H}^\omega, K^\omega)$ where $\mathcal{H}^\omega = \{H_{grue(t)} : t \in \omega\} \cup \{H_{green}\}$, and $K^\omega = \bigcup \mathcal{H}^\omega$. Figure 5 illustrates the infinitely iterated Riddle of Induction.

A fundamental difference between the infinitely iterated Riddle and the finitely iterated versions is that the infinitely iterated Riddle poses a problem of induction in the classical sense: No matter how many green emeralds have been observed, the next one might be blue. Another way of saying that the infinitely iterated Riddle is a classical problem of induction is that the cautious projection rule is no longer guaranteed to eventually settle on the correct generalization about emerald colours: if all emeralds are green the cautious rule will never make the inductive leap to this generalization. On the other hand, for each $H_{grue(t)}$ hypothesis, it is still the case that the evidence will eventually decide—namely by time t—whether $H_{grue(t)}$ is correct or not. This is an important logical asymmetry between "all emeralds are green" and "all emeralds are $grue(t)$". This asymmetry has strong methodological consequences: We shall see in Section 6 that the only efficient projection rule for the infinitely iterated Riddle, in the sense of minimizing convergence time and avoiding retractions, is the natural one: conjecture that all emeralds are green as long as all emeralds examined so far have been found to be green.

4. Convergence to the Truth and Nothing But the Truth

For different versions of the Riddle of Induction, the preceding sections compared various inductive methods with respect to whether they reliably settled on a correct hypothesis, whether they did so quickly, and whether they did so with as few vacillations—retractions or mind changes—as possible. The remainder of this paper is a systematic study of these criteria as they apply to discovery problems in general. I begin with reliable convergence to a correct hypothesis.

Through the ages, skeptical arguments dating back to Sextus Empiricus have aimed at

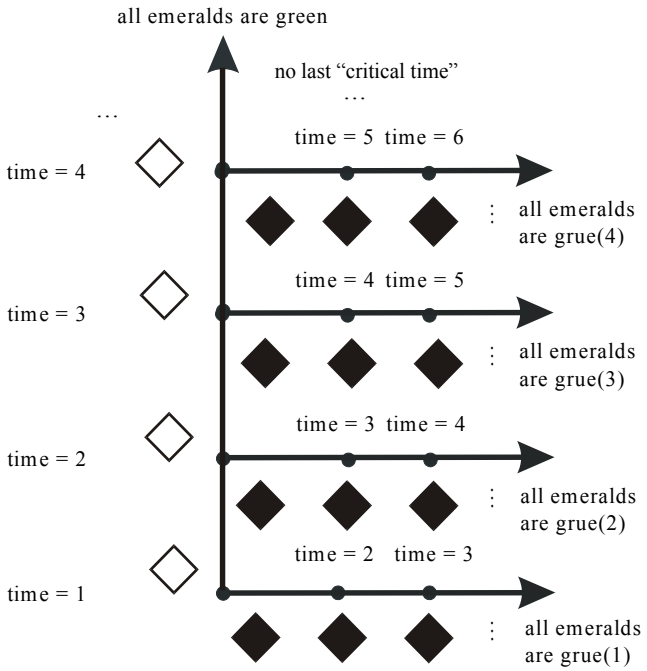


Figure 5: The Infinitely Iterated Riddle of Induction

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]. Hume formulated this observation as his celebrated *problem of induction* [Hume 1984]. One 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 that it has done so. 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.

Definition 2. *Let $\mathcal{H}$ be a collection of alternative hypotheses, ε a data stream, and let H be the hypothesis from $\mathcal{H}$ that is correct on ε.*

1. *An inductive method δ* ***converges to the correct hypothesis on*** *ε* ***by time*** *n $\iff$ for all later times $n' > n$, $\delta(\varepsilon|n')$, δ's theory on the data observed on ε up to time n' is consistent (i.e., $\delta(\varepsilon|n') \neq \emptyset$) and entails H.*

2. *An inductive method δ* ***converges to the correct hypothesis on*** *$\varepsilon \iff$ there is a time n by which δ converges to H on ε.*

As I noted, in the one-shot and finitely iterated versions of the Riddle, both the natural and the cautious projection rule will eventually converge to the correct generalizations, no matter what the correct generalization is. In the infinitely iterated version of the Riddle, we have the same guarantee for the natural projection rule, but not for the cautious one: If all emeralds are green, the cautious rule never makes the required inductive leap to generalizing beyond the evidence. On the other hand, an inquirer may be cautious for as long as she pleases even in the infinitely iterated Riddle, as long as she *eventually* projects that all emeralds are green, when it is in fact the case that all emeralds are green. Long-run convergence to the truth permits an agent to be skeptical for as long as she pleases, but not forever.

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 concept of formal learning theory.

Definition 3. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. An inductive method δ is* ***reliable*** *for the discovery problem $(\mathcal{H}, K)$ $\iff$ δ converges to the correct hypothesis on every data stream ε consistent with background knowledge K.*

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 [Putnam 1963] was that Carnap's confirmation functions are not the best for detecting regularities among the data, because there are other methods that succeed in doing so over a wider range of possibilities. (For an evaluation of Putnam's argument, see [Kelly *et al.* 1994].) This is just the fundamental decision-theoretic principle of *admissibility* applied to inductive methods. In general, an act A is admissible if it is not dominated. An act B *dominates* an act A if B yields outcomes that are necessarily at least as good as those that A produces, and possibly better, where a given collection of "possible states of the world" determines the relevant

sense of (epistemic) necessity and possibility. In our methodological setting, background knowledge specifies the relevant possibilities (data streams). The admissibility principle yields the following criterion for comparing the performance of two inductive methods with respect to convergence over a range of possible data streams.

Definition 4. *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*** $\Longleftrightarrow$

1. *on every data stream consistent with K, δ converges to the correct hypothesis on ε if δ' does, and*

2. *on some data stream ε consistent with K, δ converges to the correct hypothesis on ε and δ' does not.*

An inductive method δ is ***convergence-admissible*** *for a discovery problem $(\mathcal{H}, K)$ $\Longleftrightarrow$ δ 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 5. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. An inductive method δ is convergence-admissible for the discovery problem $(\mathcal{H}, K)$ $\Longleftrightarrow$ δ is reliable for that problem.*

Learning theorists have studied extensively the structure of discovery problems with reliable solutions, as well as the properties of reliable methods for given discovery problems (see for example, [Kelly 1996], especially Chapter 9, and [Osherson *et al.* 1986]). I continue with my exploration of epistemic aims in addition to reliable convergence to the truth.

5. Fast Convergence to the Truth

Time is a resource of inquiry. An inquirer who wants a correct theory as soon as possible prefers his methods to stabilize to a true belief sooner rather than later. Let us call the time that a method δ requires to settle on a hypothesis from a collection of alternatives $\mathcal{H}$, on a given data stream ε, the **modulus** of δ on ε; I denote the modulus by $mod(\delta, \varepsilon)$. Formally, $mod(\delta, \varepsilon)$ = the first time n by which δ converges to the correct hypothesis on ε. If a method δ fails to converge to a true hypothesis on a data stream ε, then I take its modulus on ε to be infinite, so $mod(\delta, \varepsilon) = \omega$. 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 6. *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*** $\Longleftrightarrow$

1. *on every data stream* ε *consistent with* $K, mod(\delta, \varepsilon) \leq mod(\delta', \varepsilon)$, *and*
2. *for some data stream* ε *consistent with* $K, mod(\delta, \varepsilon) < mod(\delta', \varepsilon)$.

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* $(\mathcal{H}, K)$.

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, Osherson *et al.* 1986]). Data-minimal methods are exactly the ones that satisfy a simple, intuitive criterion. Let's say that a method δ *projects* its current conjecture H at a given stage of inquiry if δ converges to H along *some* data stream consistent with background knowledge and the evidence obtained by that stage. For example, in the three versions of Goodman's Riddle, the natural projection rule projects its current generalization at each stage of inquiry. The cautious rule, on the other hand, does not project its current hypothesis before the critical time, because it is not making a conjecture as to which of the alternative hypotheses is correct. This leads to the following definitions.

Definition 7. *Let* δ *be a discovery method for a discovery problem* $(\mathcal{H}, K)$ *and let* e *be a finite data sequence,* $H \in \mathcal{H}$ *one of the alternative hypotheses.*

1. δ ***projects*** H ***at*** e ***along data stream*** $\varepsilon \iff$ *for all evidence sequences* e' *such that* $e \subseteq e' \subset \varepsilon$*:*
   (a) $\delta(e')$ *entails* H, *and*
   (b) $\delta(e')$ *is consistent.*
2. δ *projects* H *at* e ***given background knowledge*** $K \iff$ *there is a data stream* ε *consistent with* K *such that* δ *projects* H *at* e *along* ε.
3. δ *projects* ***its current hypothesis*** *at* e *given* $K \iff$ *there is some (unique) hypothesis* $H \in \mathcal{H}$ *such that* δ *projects* H *at* e *given* K.

If a method fails to project its current hypothesis, then the method is certain to eventually abandon the hypothesis no matter what future evidence it receives. So I use the phrase "method δ takes its current conjecture seriously" as an alternative to "method δ projects its current hypothesis".

The next theorem says that data-minimal methods are exactly those that always take their conjectures seriously.

Theorem 8. *Let* $\mathcal{H}$ *be a collection of alternative empirical hypotheses, and let* K *be given background knowledge. A reliable method* δ *is data-minimal for the discovery problem* $(\mathcal{H}, K) \iff$ *for each finite data sequence* e *consistent with* K, δ *projects its current hypothesis at* e *given* K.

It follows from Theorem 8 that in all three versions of the Riddle of Induction, the natural projection rule is data-minimal, whereas the cautious projection rule is not. Note that data-minimal reliable methods always produce consistent theories because a reliable method does not converge to an inconsistent (and hence false) theory.

6. Steady Convergence to the Truth

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 literature around "minimal change" belief revision shows 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, Sharma *et al.* 1997, Case and Smith 1983]. For discovery methods, 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.

Let δ be a discovery method for a collection of alternatives $\mathcal{H}$. I say that δ **retracts** its conjecture on a data stream ε at time $n+1$, or **changes its mind** at $\varepsilon|n+1$, if $\delta(\varepsilon|n)$ is consistent and entails a hypothesis H from $\mathcal{H}$, but $\delta(\varepsilon|n+1)$ either is inconsistent or does not entail H. I denote the number of times that a method δ changes its mind on a data stream ε by $MC(\delta, \varepsilon)$; formally, $MC(\delta, \varepsilon) = |\{n : \delta \text{ changes its mind at } \varepsilon|n+1\}|$. If δ does not stabilize to a hypothesis on a data stream ε, then $MC(\delta, \varepsilon)$ is infinite.

As with convergence to a true theory and convergence time, we can define a standard of performance for inductive methods by applying the admissibility principle to the aim of avoiding retractions; I refer to this performance standard as **mind-change–minimality**. It turns out that mind-change–minimality imposes stringent demands on inductive methods: A reliable mind-change–minimal method exists for a given discovery problem just in case the evidence is guaranteed to eventually entail the correct hypothesis [Schulte forthcoming, Prop.8]. In other words, mind-change–minimality is unattainable when there is a genuine problem of induction in the traditional sense. When there is no genuine problem of induction, as in the finitely iterated Riddles, the mind-change–minimal methods are the skeptical "wait-and-see" methods that wait until the evidence settles the question at hand.

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. Minimaxing retractions is possible even when there is a problem of induction. Indeed, this criterion turns out to be a very fruitful principle for deriving plausible constraints on the short-run inferences of reliable methods.

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 knowledge K, we consider the maximum number of times that a method might change its mind assuming that K is true, which is given by $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\}$.[8] If $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\} < \max\{MC(\delta', \varepsilon) : \varepsilon \in K\}$, 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 to 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* $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\} > \max\{MC(\delta', \varepsilon) : \varepsilon \in K\}$.

[8] If $\{MC(\delta, \varepsilon) : \varepsilon \in K\}$ has no maximum, let $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\} = \omega$.

If there is no bound on the number of times that a reliable method may have to change its mind to arrive at the truth, $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\}$ is infinite for all reliable methods δ, and the minimax criterion has no interesting consequences (for examples of such discovery problems see [Kelly 1996, Ch.4], or the Hypergrue problem described at the end of Section 6).[9] But if we can guarantee that a reliable method δ can succeed in identifying the correct hypothesis without ever using more than n mind changes, the principle selects the method with the best such bound on vacillations. I say that a method δ identifies a true hypothesis from a collection of alternatives $\mathcal{H}$ given background knowledge K **with at most n mind changes** if δ is a reliable method for $\mathcal{H}$ given K, and $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\} \leq n$. 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].

To get a feel for what minimaxing mind changes is like, let us consider the three versions of Goodman's Riddle of Induction. In the one-shot version, we can simply wait until the critical time when the evidence entails which of the two possible colour predicates is correct. This method succeeds with 0 retractions and hence minimaxes mind changes. Note that making a conjecture before the critical time—as data-minimality requires by Theorem 8—does not minimax retractions: if a reliable method conjectures at time $t' < t$ that all emeralds are green or that all emeralds are $grue(t)$, the subsequent emerald colours may be such that the method has to change its mind again. For example, if a method δ projects that all emeralds are green after the first emerald is found to be green, as the natural projection rule does, the emerald examined at the critical time t may be blue, forcing δ to change its mind. Similarly, in the finitely iterated version of the Riddle, minimaxing retractions requires a method to be "skeptical" and to not go beyond the evidence in its conjectures.

The infinitely iterated Riddle of Induction is different, because in this problem it is not possible to reliably converge to the correct hypothesis with 0 retractions: After some run of green emeralds, a reliable method δ must project that all emeralds are green. Otherwise δ fails to conjecture that all emeralds are green when in fact they are, and hence is unreliable. So let k be the first time such that after k green emeralds have been examined, δ projects that all examined emeralds are green. Then if a future emerald examined after time k turns out to be blue, δ has to (eventually) retract its hypothesis that all examined emeralds are green, or else again δ is unreliable. Since this argument applies to any reliable method δ, in the infinitely iterated Riddle it is impossible to reliably find the correct colour generalization with no retractions. The reason is that in that version of the Riddle, background knowledge and the available evidence never conclusively establish that all emeralds are green, and hence reliable methods must take an "inductive leap" and eventually go beyond the evidence if the data continue to be consistent with that hypothesis. This is so whenever there is a genuine problem of induction: If the evidence, together with background knowledge, never entails that H is true, then after some finite amount of evidence, a reliable method must make an inductive leap to conjecture H, and after that point the evidence may be such that H is false. Conversely, if there is no genuine problem of induction in a given discovery problem, a method can reliably identify the correct hypothesis by waiting until the evidence settles the matter. Hence it is possible

[9] However, [Jain and Sharma 1997] define ordinal-valued mind-change bounds for learning methods (see also [Sharma *et al.* 1997], [Freivalds and Smith 1993]). Even when there is no *finite* worst-case bound on the number mind changes that a reliable learner may have to undergo in a given discovery problem, there may be an *ordinal* bound.

to reliably solve a discovery problem with 0 retractions just in case the problem is not genuinely inductive. This leads to the following characterization of discovery problems that require no mind changes.

Proposition 10. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. Then there is a reliable discovery method for the discovery problem $(\mathcal{H}, K)$ that never retracts its conjectures $\Longleftrightarrow$ for each data stream ε consistent with background knowledge K, there is a time t such that $K \cap [\varepsilon|t]$ entails the hypothesis $H \in \mathcal{H}$ that is correct on ε.*

Thus *all* reliable methods require at least one mind change to solve the infinitely iterated Riddle. The natural projection rule requires *at most* one mind change: If all examined emeralds are in fact green, the rule converges to the correct belief with 0 retractions. And if all examined emeralds are $grue(t)$ for some critical time t, the natural projection rule changes its mind once at time t and never thereafter. Since there is a reliable discovery method for the infinitely iterated Riddle that requires at most one mind change, minimaxing retractions rules out all methods that might use more than one. Which reliable projection rules change their mind at most once? Strikingly, none of the unnatural ones that project a grue predicate after examining a sample of green predicates.[10] For consider a reliable method δ that examines k green emeralds and then projects that all examined emeralds are $grue(t)$, for some later time $t > k$. Now suppose that we continue to find green emeralds beyond time t. Then the evidence falsifies δ's earlier conjecture, and eventually δ retracts "all examined emeralds are $grue(t)$" since δ is reliable—one retraction. Indeed, if all examined emeralds continue to be green, δ must eventually project that all emeralds are green, say at time l. But then at some later time $l' > l$, a blue emerald may be found, again falsifying δ's conjecture, and forcing δ to change its mind for a second time.

Since this argument applies to any projection rule that projects a $grue(t)$ predicate after finding nothing but green emeralds, reliability and minimaxing retractions allow only those rules that, for a finite "waiting time" don't go beyond the evidence while finding green emeralds, and then eventually project "all emeralds are green". Now by Theorem 8, data-minimality does not permit inductive methods to "wait for more evidence" before conjecturing one of the alternative hypotheses. And as we have seen, minimaxing retractions requires them to project "all emeralds are green" when all emeralds observed so far are green. Data-minimal methods must also immediately conclude that all emeralds are $grue(t)$ when the first blue emerald appears at time t. Thus reliability together with data-minimality and minimaxing retractions single out the natural projection rule as the *only* optimal one. The next proposition records this observation.

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

Next, let's consider a finitely iterated Riddle, say with the latest critical time at stage m. As we saw in Section 3.3, it is possible to find the correct generalization in this version by simply waiting until stage m. This cautious procedure δ_C^m requires 0 retractions, and hence minimaxes retractions, but it is not data-minimal because it does not project any hypothesis until stage m. Suppose we apply the criterion of minimaxing retractions to

[10] Kevin Kelly was the first to notice this fact.

select among the *data-minimal* reliable projection rules the ones with the best bound on mind changes—which methods are optimal in that sense? By Theorem 8, a data-minimal rule must immediately project a generalization (for example "all emeralds are $grue(m)$"). But that generalization may turn out to be false (for example if the m-th examined emerald is green). Then the data-minimal rule has to change its mind. Thus all reliable data-minimal projection rules in the m-iterated Riddle must change their mind at least once (for $m > 1$). Which ones don't change their mind more than once? There are exactly two: the natural projection rule, and the δ^m_m rule that projects "all emeralds are $grue_m$" so long as that conjecture is consistent with the evidence. Contrast this with a data-minimal rule δ^m_{m-1} that begins by projecting "all emeralds are $grue(m-1)$". Then if the first $m-1$ emeralds are green, the conjecture "all emeralds are $grue(m-1)$" is falsified, and δ^m_{m-1} has to change its mind immediately (since δ^m_{m-1} is data-minimal). But no matter whether δ^m_{m-1} then projects that all emeralds are green or that all emeralds are $grue(m)$, its conjecture may be falsified again, and δ^m_{m-1} must change its mind for a second time. This shows that among the data-minimal projection rules in the m-iterated Riddle of Induction, only the natural projection rule and δ^m_m minimax retractions.

Proposition 12. *In a finitely iterated Riddle of Induction, let m be the last "critical time". Let δ^m_m be the projection rule that projects "all emeralds are $grue(m)$" until the evidence falsifies this conjecture and let δ^m_N be the natural projection rule. The two rules δ^m_m and δ^m_N are the only reliable and data-minimal rules that succeed with at most one mind change.*

To sum up: In the infinitely iterated Riddle, it is possible to have a data-minimal projection rule that minimaxes retractions. The only such rule is the natural one. In the finitely, m-iterated Riddle, minimizing time-to-truth conflicts with avoiding retractions. If we put minimizing time-to-truth first, and then use the criterion of minimaxing retractions to select among reliable data-minimal projection rules, we have a choice between the natural projection rule and the δ^m_m rule ("all emeralds are $grue(m)$").

These examples show that reliability and efficiency criteria yield interesting, principled and plausible recommendations for what inductive methods should conjecture in the short run. The same is true if we apply the means-ends analysis to more puzzling Riddles of Induction. For example, consider a Riddle that allows for n colour changes from blue to green and back. That is, a data stream is possible in this Riddle just in case a blue emerald follows a green one, or vice versa, no more than n times.[11] The infinitely iterated Riddle of Induction has $n = 1$. The alternative hypotheses are the universal generalizations whose empirical content is exactly one data stream. In the n-colour change Riddle, it is possible to reliably identify the correct universal generalization about emerald colours (that is, the actual data stream) with at most n retractions. And the only data-minimal reliable method that accomplishes this and minimaxes retractions projects that all emeralds are green as long as all emeralds examined so far are green. On the other hand, there is another Riddle in which it is impossible to succeed with any bounded number of retractions, so the directive to minimax retractions does not apply; I refer to this Riddle as the "Hypergrue Problem". (The problem and its name are due to Kevin Kelly.) In the Hypergrue Problem, the number of possible colour changes increases with time, such that if the first colour change occurs at stage t, then t more colour changes are possible afterwards—but no more than t.

[11] Defining the alternative possibilities in terms of blue-green colour changes is not "privileging green" over grue, just convenient.

To analyze these examples and other discovery problems, we do well to equip ourselves with some fundamental mathematical tools for investigating data-minimality and the mind-change complexity of inductive problems.

7. The Topology of Fast and Steady Reliable Inquiry

Proposition 10 characterizes the discovery problems in which reliable methods need no mind changes. This section generalizes the proposition to any bound n on the number of times that a reliable method might change its mind. What is the *structure* of those discovery problems that don't require more than n retractions? I answer this question in terms of the *topology* of the alternative hypotheses[12], determined by the given background knowledge. The finite versions of the Goodmanian Riddle show that data-minimality—minimizing convergence time—may require extra mind changes. In other problems, such as the infinitely iterated Riddle of Induction, there is no such conflict. I characterize the number of mind changes that a reliable *data-minimal* method might have to undergo in a given discovery problem. Together, the two characterizations measure the extent to which data-minimality conflicts with avoiding retractions in a given discovery problem.

7.1. Necessary and Sufficient Conditions for Discovery With Bounded Mind Changes. A reliable discovery method δ identifies a correct hypothesis from a collection of alternatives $\mathcal{H}$ with at most n mind changes given background knowledge K if δ does not change its mind more than n times on any data stream ε consistent with K. That is, δ succeeds with at most n mind changes if δ is a reliable discovery method for $\mathcal{H}$ given K, and $\max\{MC(\delta, \varepsilon) : \varepsilon \in K\} \leq n$. The next theorem characterizes what background knowledge K must be like if there is a discovery method δ that reliably identifies a correct hypothesis from a collection of alternatives $\mathcal{H}$ and never changes its mind more than n times on any data stream ε consistent with K. I define the characteristic condition inductively, starting with discovery without any mind changes. Consider an initial conjecture H. Suppose that H is not certain (i.e., the background knowledge K does not entail H). Then any reliable discovery method starting with H has to change its mind if H is false. If after this mind change, still n more mind changes are required, a total of $n + 1$ mind changes may result. So if a reliable discovery method δ whose initial conjecture is H never requires more than n mind changes, there must be some point at which δ can change its mind and incur no more than $n - 1$ mind changes whenever H is false. The structures that meet this requirement look like "feathers" ([Kelly 1996, Ch.4]). I write $F_n(K, H)$ for "K is an n-feather for H". The intended interpretation of $F_n(K, H)$ is "every reliable discovery method starting with H requires at least $n + 1$ mind changes given K".

Definition 13. *Let $\mathcal{H}$ be a collection of empirical hypotheses, and let K be background knowledge. Let $\mathcal{H}(\varepsilon)$ stand for the (unique) member of $\mathcal{H}$ correct on $\varepsilon \in K$. Write $F_n(K, H)$ for "K is an n-feather for H", and define this notion as follows.*

1. $F_0(K, H) \iff \exists \varepsilon \in K - H$.
2. $F_{n+1}(K, H) \iff \exists \varepsilon \in K - H. \forall k$: $F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$.

To illustrate this definition, the background knowledge K in the infinitely iterated Riddle of Induction is a 0-feather for H_{green} ("all emeralds are green")—that is, $F_0(K, H_{green})$ holds—because H_{green} is not true on every data stream consistent with K (see Figure 5).

[12] [Schulte and Juhl 1997] and [Kelly 1996, Ch.4] define the relevant topology on the space of data streams.

But K is not a 1-feather for H_{green}, that is, $\neg F_1(K, H_{green})$ holds: for every data stream ε in K on which H_{green} is false—on which a blue emerald is observed at, say time k—there is an initial segment $\varepsilon|k$, such that $K \cap [\varepsilon|k]$ is not a 0-feather for $H_{grue(k)}$ ("all emeralds are $grue(k)$"), that is, $H_{grue(k)}$ is entailed by $K \cap [\varepsilon|k]$. By contrast, K is a 1-feather for any $H_{grue(t)}$, that is, $F_1(K, H_{grue(t)})$ holds: A given hypothesis $H_{grue(t)}$ is false on the sequence τ of all green emeralds. And no initial segment $\tau|k$ entails $H_{green} = \mathcal{H}(\tau)$. So $K \cap \tau|k$ is a 0-feather for H_{green}.

Figure 6 displays the general structure of 0 and 1-feathers, and Figure 7 illustrates 2-feathers and 3-feathers.

The next lemma shows that feather structures characterize how many mind changes are required by a discovery method that starts with a certain initial conjecture.

Lemma 14. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let background knowledge K be given. Then there is a reliable discovery method δ for the discovery problem $(\mathcal{H}, K)$ such that*

1. *δ succeeds with at most n mind changes, and*
2. *$\delta(\emptyset)$ is consistent and entails H*

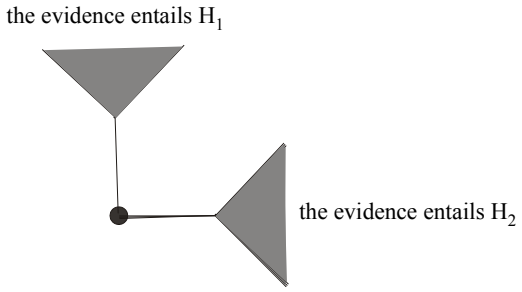
$\iff$ (K, H) is not an n-feather (that is, $\neg F_n(K, H)$).

The last complication is that a reliable method may delay conjecturing any of the alternative hypotheses. In fact, minimaxing retractions can require arbitrarily long delays. In the one-shot Riddle of Induction, for example, a method has to wait until the critical time t before projecting either of the two alternatives. Thus the full characterization of discovery with bounded mind changes is this: A reliable method must use at least $n+1$ mind changes, if and only if, there is one data stream ε consistent with background knowledge K such that every initial segment $\varepsilon|k$ is an n-feather for every hypothesis H under consideration.

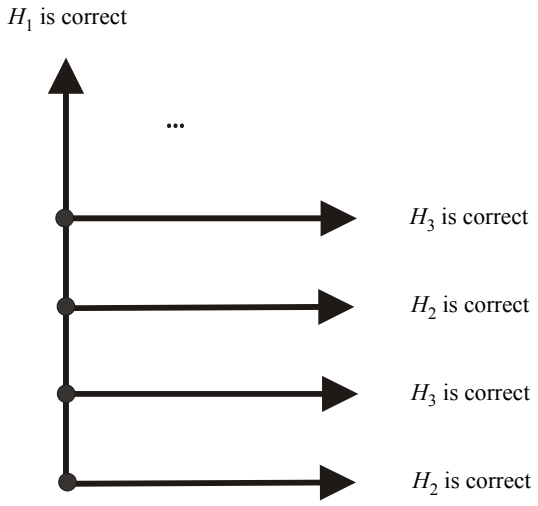
Theorem 15. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let background knowledge K be given. Then there is a reliable discovery method δ for $\mathcal{H}$ given K that succeeds with at most n mind changes $\iff$ for every data stream ε consistent with K, there is a time k such that $(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$ is not an n-feather (that is, $\neg F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$.*

From Theorem 15, we can derive a universal method δ_{MC} for reliably identifying a correct hypothesis from $\mathcal{H}$ given K when (K, H) is not an n-feather. Say that the dimension of (K, H) is n if (K, H) is an n-feather but not an $n+1$ feather. The universal method δ_{MC} begins by conjecturing nothing but the evidence until $(K \cap [e], H)$ is of dimension n, for some $H \in \mathcal{H}$; then δ_{MC} conjectures H. Let a finite data sequence $e * x$ be given (i.e., e followed by one datum x); if there is an H' such that $(K \cap [e * x], H')$ is of lower dimension than $(K \cap [e * x], H)$, then $\delta_{MC}(e * x) = H'$; otherwise $\delta_{MC}(e * x) = \delta_{MC}(e)$.

Theorem 15 settles the status of the "Hypergrue" problem posed at the end of Section 6: it is possible to reliably find the correct generalization about colour predicates, but not with a bounded number of mind changes. For the positive part of the claim, consider the natural projection method that conjectures that all emeralds are green until a blue-green colour change occurs. If all emeralds are green, this method converges to the correct generalization. Otherwise, let t be the first time at which a blue emerald is observed. By the definition of the Hypergrue problem, at most t more colour changes are now possible.

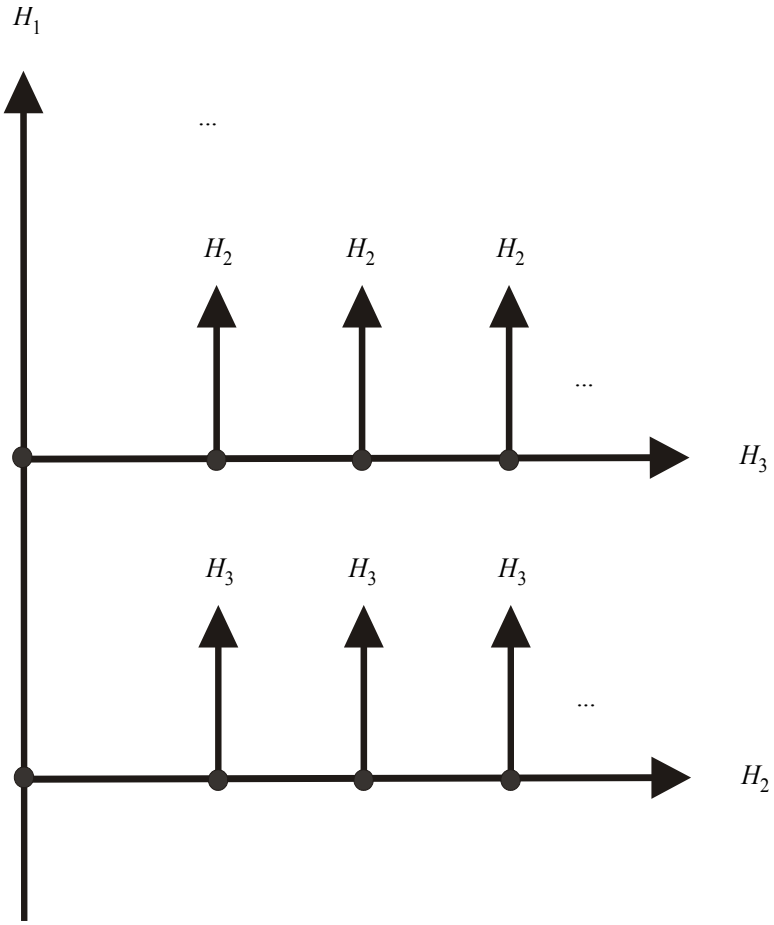


(a) A 0-feather, but not a 1-feather for H_1 and for H_2.



(b) A 0-feather, but not a 1-feather for H_1.
A 1-feather, but not a 2-feather for H_2 and H_3.

Figure 6: "Feather" Structures characterize Discovery with Bounded Mind Changes. The figure illustrates 0-feathers and 1-feathers.



A 1-feather, but not a 2-feather for H_1.
A 2-feather, but not a 3-feather for H_2 and H_3.

Figure 7: 2-feathers and 3-feathers

This means that the evidence together with the applicable background knowledge K^{hyper} does not constitute a $t+1$ feather, and hence the universal method δ_{MC} now settles on the correct generalization about emerald colours with no more than t mind changes. For the impossibility claim, let an inductive method δ be given and suppose that δ changes its mind at most n times on any data stream consistent with the background knowledge K^{hyper} specified in the Hypergrue problem. Let e be the evidence sequence consisting of $n+1$ green emeralds. Then $(K^{hyper} \cap [e], \{\varepsilon\})$ is an n-feather for every universal generalization $\{\varepsilon\}$ consistent with $K^{hyper} \cap [e]$, and so no reliable method finds the correct generalization about emerald colours with at most n mind changes given $K^{hyper} \cap [e]$. In particular, δ is not reliable for the Hypergrue problem given $K^{hyper} \cap [e]$, and hence not given K^{hyper}.[13]

7.2. Necessary and Sufficient Conditions for Data-Minimal Discovery With Bounded Mind Changes. Sometimes it is possible to minimax retractions with a reliable and data-minimal method. In such cases inductive inquiry can epistemically "have it all", and the inferences of reliable and efficient methods have special intuitive appeal. Other problems such as the finitely iterated versions of the Riddle pose a hard choice between avoiding retractions and time-to-truth. In what problems does this tension arise, and how serious is it?

The next theorem determines the exact extent to which data-minimal methods may have to undergo extra mind changes to solve an inductive problem, compared to slower methods whose convergence time is not optimal. I begin with a variant of Definition 13. If a reliable data-minimal discovery method can succeed with n mind changes, then whenever the previous conjecture H of a data-minimal method is refuted, the method must be able to immediately change its mind to a conjecture H' after which no more than $n-1$ mind changes are required. Another way of putting the matter is that the universal method δ_{MC} for discovery with bounded mind changes is not data-minimal unless for some hypothesis H', $(K \cap [e * x], H')$ is of lower dimension than $(K \cap [e * x], H)$ whenever $e * x$ falsifies the previous conjecture H of δ_{MC}. Clause 2b of the next definition reflects this observation. The intended interpretation of DM-$F_n(K, H)$—read "background knowledge K is a data-minimal–n-feather for hypothesis H"—is "a reliable data-minimal method whose initial conjecture is H requires at least $n+1$ mind changes".

Definition 16. *Let $\mathcal{H}$ be a collection of empirical hypotheses, and let K be background knowledge.*

1. *DM-$F_0(K, H) \iff \exists \varepsilon \in K - H$*
2. *DM-$F_{n+1}(K, H) \iff$ if H is consistent with K, then $\exists \varepsilon \in K - H$ such that*
   (a) $\forall k$:*DM-$F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$*, or
   (b) $\exists k : K \cap [\varepsilon|k]$ *entails that H is false, and $\forall H'$ in $\mathcal{H}$ consistent with $K \cap [\varepsilon|k]$:* *DM-$F_n(K \cap [\varepsilon|k], H')$.*

To illustrate this definition, consider the twice iterated Riddle of Induction with background knowledge K^2. This background knowledge is a DM-0-feather for all universal

[13] [Jain and Sharma 1997] and [Sharma *et al.* 1997] present an interesting generalization of mind change complexity that allows us to classify the difficulty of Hypergrue in a more informative way. By their definition of ordinal-valued mind change bounds (originally due to [Freivalds and Smith 1993]), the Hypergrue problem can be solved with $\omega \cdot 2$ mind changes.

generalizations—that is, DM-$F_0(K, \{\varepsilon\})$ holds for any $\varepsilon \in K^2$—because K^2 does not entail any universal generalization. But (K^2, H_{green}) and $(K^2, H_{grue(2)})$ are *not* DM-1-feathers—that is, $\neg$DM-$F_1(K, H_{green})$ and $\neg$DM-$F_1(K, H_{grue(2)})$ hold—because, first, along every data stream ε on which, for example, H_{green} is false, eventually the evidence entails the correct hypothesis. And second, there is no evidence that falsifies H_{green} without entailing an alternative hypothesis. But $(K^2, H_{grue(1)})$ *is* a DM-1-feather: For the evidence sequence e featuring one green emerald falsifies $H_{grue(1)}$, but does not entail an alternative hypothesis; formally, for all $\{\varepsilon\}$ consistent with $K^2 \cap [e]$, we have that DM-$F_0(K^2 \cap [e], \{\varepsilon\})$ holds. Accordingly, a data-minimal projection rule whose first conjecture is "all emeralds are $grue(2)$" might have to change its mind twice.

Since data-minimal methods must immediately project one of the alternative hypotheses, a data-minimal method cannot wait for evidence before making a conjecture; otherwise the characterization is analogous to Theorem 15.

Theorem 17. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let background knowledge K be given. Then there is a reliable method δ for the discovery problem $(\mathcal{H}, K)$ such that*

1. *$\delta(\emptyset)$ is consistent and entails H, and*
2. *δ requires at most n mind changes, and*
3. *δ is data-minimal*

$\iff$ (K, H) is not a data-minimal n-feather (that is, $\neg$DM-$F_n(K, H)$).

Let's return to the version of the Riddle of Induction from the end of Section 6 that allows at most n blue-green colour changes to occur. In this Riddle, any reliable data-minimal projection rule δ that minimaxes retractions conjectures that all emeralds are green as long as all emeralds examined so far are green. Here's why: By Theorem 8, any data-minimal reliable projection rule δ must immediately entail some conjecture H. Suppose that $\delta(\emptyset)$ is inconsistent with H_{green}. Then along the data stream ε of all green emeralds, δ must eventually change its mind to H_{green}; let the first such time be k. Then $K \cap [\varepsilon|k]$ is still an $n-1$-feather (and thus a data-minimal $n-1$-feather), and so δ may be forced to change its mind n more times, for a total of $n+1$ mind changes. Thus any data-minimal reliable projection rule projects "all emeralds are green" as long as all emeralds observed so far are green.

8. Conclusion

Means-ends epistemology takes the answer to the question "why ought we to draw this inference rather than that one?" to be of the form "because this inference method is the best for the aims of inquiry". This paper studied three natural and interesting epistemic objectives: reliable convergence to a correct theory, fast convergence to a correct theory, and steady convergence to a correct theory (avoiding retractions). We may consider the latter two criteria as standards of efficiency for reliable empirical methods.

I investigated the structure of those inductive problems in which these goals are feasible, and gave necessary and sufficient conditions that characterize this structure. I specified the principles that guide inductive methods designed to attain these epistemic goals where they are feasible. As an illustration, I applied these results to various versions of Goodman's Riddle of Induction. In the infinitary version of Goodman's Riddle, there

is only one efficient inference rule for this problem; it turns out to be the natural rule (project that "all emeralds are green" as long as all observed emeralds are green). The characterization results established in this paper show that whether an inductive problem has the requisite structure for reliable efficient inquiry depends only on logical relations of entailment between possible evidence items and the hypotheses under investigation. Since acceptable translations from one language into another preserve logical entailment, it does not matter to the methodology of reliable efficient inquiry what language we use to describe evidence and hypotheses. In particular, the means-ends solution to Goodman's Riddle does not depend on whether we choose the "blue-green" or the "grue-bleen" language for describing the problem.

This may surprise the reader—did Goodman not show that the blue-green and grue-bleen pairs of predicates are interchangeable, in the sense that one pair can be defined in terms of the other? The answer is that my analysis turns on logical (and topological) asymmetries between the *universal generalizations* of these predicates. These asymmetries depend on a pragmatic factor, namely which hypotheses we are willing to entertain as candidates for projection (more generally, they depend on what hypotheses are consistent with the inquirer's background assumptions). When we allow grue predicates for all critical times as candidates for projection, "all emeralds are green" is different from "all emeralds are grue (with a given critical time t)", because the classical problem of induction arises for the former but not for the latter. That is to say, no matter how many green emeralds have been found, our background knowledge allows that the next one may be blue, and thus that "all emeralds are green" may be false. In contrast, if all emeralds up to and including the critical time t are grue, the only candidate for projection is "all emeralds are grue", because in that case the evidence falsifies "all emeralds are green" as well as all "grue" predicates with critical times other than t.

The characterization theorems for reliable efficient inquiry invite us to apply means-ends analysis to other inductive problems. For example, it turns out that the Occam-like problem of determining whether a given entity exists shares with Goodman's Riddle the structure characteristic of reliable efficient inquiry, and so does the problem of inferring theories of reactions among elementary particles [Schulte forthcoming, Schulte 1997]. In these cases too standards of efficiency yield interesting methodological recommendations: A version of Occam's Razor in the Occam problem, and a form of "choose the closest fit to the data" in the particle problem. Thus one important strength of means-ends analysis is that it reveals epistemologically significant, common structure among superficially quite different empirical problems. The characterization theorems show that this structure does not depend on the language in which evidence and hypotheses are described.

The goal of this paper was to contribute an explicit, general characterization of the structure and logic of reliable efficient inquiry to the foundations of means-ends epistemology, and to illustrate some of the rewarding applications of this approach to questions about inductive inference.

Acknowledgments

I am indebted to Kevin Kelly, Clark Glymour and Cory Juhl for helpful discussion and comments. Bernard Linsky and an anonymous referee made valuable suggestions for the organization of this paper.

9. Proofs

Proposition 5. *Let H be a collection of alternative hypotheses, and let K be given background knowledge. An inductive method δ is convergence-admissible for the discovery problem ($\mathcal{H}$, K) $\Longleftrightarrow$ δ is reliable for that problem.*

Proof. ($\Longleftarrow$) Immediate.

($\Longrightarrow$) I show the contrapositive. Suppose that δ is not reliable for $(\mathcal{H}, K)$. Then there is a data stream ε consistent with K such that δ fails to converge to the truth on ε. Let $H \in \mathcal{H}$ be the hypothesis correct on ε, and define δ' like this: Conjecture H while the data are consistent with ε. If the observations deviate from ε, δ' follows δ. Then δ' converges to the correct hypothesis H on ε, and converges to the same hypothesis as δ on all other data streams. Hence δ' dominates δ with respect to convergence, and δ is not convergence-admissible. ■

Theorem 8. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. A reliable method δ is data-minimal for the discovery problem ($\mathcal{H}$, K) $\Longleftrightarrow$ for each finite data sequence e consistent with K, δ projects its current hypothesis at e given K.*

Proof. ($\Longrightarrow$) I show the contrapositive. Suppose that there is some finite evidence sequence e (consistent with K) such that δ does not projects its conjecture at e given K. Let e_1 be a shortest data sequence that extends e such that δ does project an hypothesis H from $\mathcal{H}$ that is entailed by $\delta(e_1)$ along some data stream $\varepsilon \in K$. Since δ does not project $\delta(e)$, e_1 must properly extend e; hence we may take $e_1 = e_0 * x$, where x is the last datum that appears in e_1. Now define δ' by $\delta'(e_0) = \delta(e_1)$, and $\delta'(e') = \delta(e')$ at all data sequences e' different from e_0. I show that δ' weakly dominates δ. By construction, δ' projects the hypothesis H along ε at e_0. Thus $mod(\delta', \varepsilon) \leq lh(e_0)$. By contrast, the choice of e_0 implies that δ does not stabilize to H along ε at e_0, so $lh(e_0) < mod(\delta, \varepsilon)$. Hence $mod(\delta', \varepsilon) < mod(\delta, \varepsilon)$; that is, δ' converges on ε faster than δ does. Furthermore, on no data stream consistent with background knowledge K does δ' converges after δ. For the only place at which δ and δ' differ is e_0, and by assumption δ is not converging at e_0 on any data stream ε consistent with K , since e_0 is shorter than e_1 and so δ doesn't take its conjecture at e_0 seriously given K. This establishes that δ' weakly dominates δ.

($\Longleftarrow$) Suppose that δ always takes its conjectures seriously given K. Consider some other reliable method δ' that converges faster than δ on some data stream $\varepsilon \in K$ (i.e., $mod(\delta', \varepsilon) < mod(\delta, \varepsilon)$). Let H be the hypothesis correct on ε, and let k be the first time after δ' converges on ε (that is, $k \geq mod(\delta', \varepsilon)$) such that $\delta(\varepsilon|k)$ entails H' which is inconsistent with H. Now by hypothesis, δ projects its conjecture H' along some data stream $\tau \in K$ at $\varepsilon|k$; that is, $k \geq mod(\delta, \tau)$. Since δ' projects H along ε at k, δ' does not entail H' at $\varepsilon|k = \tau|k$. Thus $mod(\delta', \tau) > k \geq mod(\delta, \tau)$. So δ' does not dominate δ in convergence time. Since any method δ' that dominates δ in convergence time must be faster than δ on some data stream $\varepsilon \in K$, this argument shows that δ is data-minimal. ■

Proposition 10. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let K be given background knowledge. Then there is a reliable discovery method for the discovery problem ($\mathcal{H}$, K) that never retracts its conjectures $\Longleftrightarrow$ for each data stream ε consistent with background knowledge K, there is a time t such that $K \cap [\varepsilon|t]$ entails the hypothesis $H \in \mathcal{H}$ that is correct on ε.*

Proof. ($\Longrightarrow$) I show the contrapositive. Suppose that for some data stream $\varepsilon \in K$, $K \cap [\varepsilon|t]$ does not entail the correct hypothesis H at any time t. Let δ be a reliable method for $(\mathcal{H}, K)$; at some stage n, δ stabilizes to H along ε, such that $\delta(\varepsilon|n)$ entails H. By assumption, $K \cap [\varepsilon|n]$ does not entail H, and thus there is a data stream $\tau \in K$ extending $\varepsilon|n$ on which H is false. Since δ is reliable, δ changes its mind on τ after stage n from H to another hypothesis. Thus δ retracts its hypothesis at least once, and so does any other reliable method for $(\mathcal{H}, K)$.

($\Longleftarrow$) If the right-hand side holds, the cautious method δ_C that waits until the evidence is reliable and never retracts its conjectures. ■

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

Proof. The natural projection rule δ_N conjectures that all emeralds are green until it encounters a blue one; suppose that the k-th emerald is blue. Then δ_N concludes that all emeralds are $grue(k)$. If all emeralds are green, then δ_N converges to the right generalization immediately. Otherwise, δ_N changes its mind, for the first time, after the first blue emerald turns up. Hence—assuming that all emeralds are green or $grue(k)$ for some k—δ_N finds the correct generalization with at most one mind change. Finally, δ_N always takes its conjectures seriously, so by Theorem 8, δ_N is data-minimal.

Now consider any projection rule δ that is reliable, data-minimal and minimaxes mind changes; I show that $\delta = \delta_N$. Since δ is reliable, δ must eventually infer that all emeralds are green if only green emeralds are observed. Let m be the minimal number of green emeralds from which δ generalizes that all emeralds are green. I argue that $m = 1$, that is, δ must immediately infer that all emeralds are green when one green emerald is observed. For suppose otherwise ($m > 1$). Since δ is data-minimal, δ must project some hypothesis other than "all emeralds are green" before the m-th green emerald is observed. That is, δ changes its mind when the m-th emerald appears. But after δ has inferred that all emeralds are green from the sample of m green emeralds, a blue emerald may be found, say at time k, which establishes that all emeralds are $grue(k)$. Since δ is reliable and data-minimal, δ must then change its mind for the *second* time to conclude that all emeralds are $grue(k)$. So if $m > 1$, then δ does not minimax retractions; thus δ infers that all emeralds are green after seeing the first green emerald. Since δ is data-minimal, δ projects this hypothesis as long as all observed emeralds are green. And again by data-minimality, δ concludes immediately that all emeralds are $grue(k)$ if the k-th emerald is blue. Thus $\delta = \delta_N$; that is, the only reliable and data-minimal projection rule that minimaxes retractions in the Riddle of Induction is the natural projection rule. ■

Proposition 12. *In a finitely iterated Riddle of Induction, let m be the last "critical time". Let δ_m^m be the projection rule that projects "all emeralds are $grue(m)$" until the evidence falsifies this conjecture and let δ_N^m be the natural projection rule. The two rules δ_m^m and δ_N^m are the only reliable and data-minimal rules that succeed with at most one mind change.*

Proof. It is clear that both δ_m^m and δ_N^m are reliable, data-minimal and use at most one mind change. Consider some other data-minimal rule δ that is reliable for the m-iterated Riddle of Induction. Since δ is reliable, its initial conjecture on no evidence must be "all emeralds are $grue(m')$" for some $m' < m$. The evidence sequence e comprising m'

green emeralds is then consistent with background knowledge and falsifies δ's conjecture. Since δ is data-minimal, δ must change its mind at e immediately. Now there are at least two alternative hypotheses consistent with background knowledge and e, namely $H_{grue(m)}$ and H_{green}. So no matter what δ conjectures at e, the conjecture may turn out to be false, in which case δ has to change its mind for a second time. Since this argument applies to any reliable and data-minimal projection rule in the m-iterated Riddle of Induction, the only reliable and data-minimal rules that succeed with at most one mind change are δ^m_m and δ^m_N. ■

Lemma 14. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let background knowledge K be given. Then there is a reliable discovery method δ for the discovery problem $(\mathcal{H}, K)$ such that*

1. *δ succeeds with at most n mind changes, and*
2. *$\delta(\emptyset)$ is consistent and entails H*

$\Longleftrightarrow (K, H)$ is not an n-feather (that is, $\neg F_n(K, H)$).

Proof. The proof is by induction on n.

Base Case, $n = 0$.

($\Leftarrow$) Suppose that (K, H) is not a 0-feather. Then H is a priori certain, that is, K entails H. So the method δ that always conjectures H reliably identifies the truth from $\mathcal{H}$ with 0 retractions.

($\Rightarrow$) Suppose that (K, H) is a 0-feather. Then there exists a data stream $\varepsilon \in K - H$. Let δ be any reliable method that starts with H (i.e., $\delta(\emptyset)$ entails H). Since δ is reliable, δ changes its mind on ε at least once. Hence every reliable method that starts with H may change its mind at least once.

Inductive Step: Assume the hypothesis for n and consider $n + 1$.

($\Leftarrow$) Suppose that (K, H) is not an n+1-feather. Then for every data stream $\varepsilon \in K - H$, there is a time k such that $\neg F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$. By inductive hypothesis, for each such point $\varepsilon|k$, we may choose a method $\delta_{\varepsilon|k}$ and a hypothesis $H_{\varepsilon|k}$ such that $\delta(\emptyset) = H_{\varepsilon|k}$ and δ succeeds with at most n mind changes given $K \cap [\varepsilon|k]$.

Now define a discovery method δ that reliably identifies a correct hypothesis from $\mathcal{H}$ given K with no more than $n + 1$ mind changes, starting with H:

1. $\delta(\emptyset) = K \cap H$;
2. If there is a time k such that
   (a) $0 < k \leq lh(e)$, and
   (b) $(K \cap [\varepsilon|k])$ is not an $n + 1$-feather for some H' in $\mathcal{H}$,

   then let k be the least such time and conjecture $\delta_{\varepsilon|k}(e)$.
3. Otherwise, conjecture $K \cap [e] \cap H$.

To see that δ succeeds with at most $n + 1$ mind changes, consider any data stream $\varepsilon \in K$.

Case 1: Clause 3 always obtains along ε. Then δ converges to H with 0 retractions. Since (K, H) is not an $n + 1$-feather, we have that $\varepsilon \in H$, and δ is correct.

Case 2: Clause 2 obtains at some point k along ε. Assume that k is the first such point. Then on ε, time k is the earliest at which δ might change its mind. After time k, δ follows $\delta_{\varepsilon|k}$ and hence succeeds with at most n mind changes. Hence overall, δ changes its mind at most $n+1$ times along ε. Since this is true for any data stream ε consistent with background knowledge K, δ requires at most $n+1$ mind changes.

($\Rightarrow$) Suppose that (K, H) is an $n+1$-feather. Then there is a data stream $\varepsilon \in K - H$ such that for all times k, $(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$ is an n-feather (i.e., $F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$ holds). Let δ be any reliable discovery method that starts with H (i.e., $\delta(\emptyset) \models H$). Then some time along ε, δ changes its mind to $\mathcal{H}(\varepsilon)$; let k be the first such time. By inductive hypothesis, any method δ' that begins with $\mathcal{H}(\varepsilon)$ requires at least $n+1$ mind changes on some data stream $\tau \in K \cap [\varepsilon|k]$. In particular, the following method δ' does.

1. $\delta'(\emptyset) = \mathcal{H}(\varepsilon)$;
2. if $e \subseteq \varepsilon|k$, $\delta'(e) = \mathcal{H}(\varepsilon)$;
3. if $\varepsilon|k \subseteq e$, $\delta'(e) = \delta(e)$.

By construction, on $K \cap [\varepsilon|k]$, δ' changes its mind only after $\varepsilon|k$, and hence changes its mind on $K \cap [\varepsilon|k]$ exactly when δ does. Hence δ changes its mind at least $n+1$ times on some data stream $\tau \in K \cap [\varepsilon|k]$. Since δ also changes its mind before $\tau|k = \varepsilon|k$, δ requires at least $n+2$ mind changes. Hence any reliable method starting with H may change its mind at least $n+2$ times when (K, H) is an $n+1$-feather, which completes the inductive step. ■

Theorem 15. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let background knowledge K be given. Then there is a reliable discovery method δ for the discovery problem $(\mathcal{H}, K)$ that succeeds with at most n mind changes $\iff$ for every data stream ε consistent with K, there is a time k such that $(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$ is not an n-feather (that is, $\neg F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$).*

Proof. ($\Leftarrow$) Suppose that for every data stream ε consistent with K, there is a time k such that $(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$ is not an n-feather (i.e., $\neg F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$. By Lemma 14, for each such point $\varepsilon|k$, we may choose a method $\delta_{\varepsilon|k}$ and a hypothesis $H_{\varepsilon|k}$ such that $\delta(\emptyset) = H_{\varepsilon|k}$ and δ succeeds with at most n mind changes given $K \cap [\varepsilon|k]$. Now define a discovery method δ that reliably identifies a correct hypothesis from $\mathcal{H}$ given K with no more than n mind changes:

1. If there is a time k such that
   (a) $0 < k \leq lh(e)$, and
   (b) $(K \cap [\varepsilon|k], H')$ is not an n-feather for some H' in $\mathcal{H}$,

   then let k be the least such time and conjecture $\delta_{\varepsilon|k}(e)$.
2. Otherwise, conjecture $K \cap [e]$.

For any data stream $\varepsilon \in K$, there eventually comes a first time k when $(K \cap [\varepsilon|k], H')$ is not an n-feather for some H' in $\mathcal{H}$. After time k, δ follows $\delta_{\varepsilon|k}$ and hence succeeds with at most n mind changes. Before time k, δ conjectures only the evidence and hence does not change its mind.

($\Longrightarrow$) Conversely, suppose that there is a data stream $\varepsilon \in K$ such that for all times k $(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$ is an n-feather (i.e., $F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$. Let δ be any reliable discovery method. Then there is a first time k along ε at which δ conjectures $\mathcal{H}(\varepsilon)$. By the same argument as in Lemma 14, δ requires at least $n+1$ mind changes on some data stream $\tau \in K \cap [\varepsilon|k]$. ■

Theorem 17. *Let $\mathcal{H}$ be a collection of alternative hypotheses, and let background knowledge K be given. Then there is a reliable method δ for the discovery problem $(\mathcal{H}, K)$ such that*

1. *$\delta(\emptyset)$ is consistent and entails H, and*
2. *δ requires at most n mind changes, and*
3. *δ is data-minimal*

$\Longleftrightarrow$ (K, H) is not a data-minimal n-feather (that is, $\neg$DM-$F_n(K, H)$.)

Proof. The proof is by induction on n. The base case follows as in the proof of Theorem 15.

Inductive Step: Assume the hypothesis for n and consider $n+1$.

($\Rightarrow$) Suppose that (K, H) is a data-minimal $n+1$-feather, that is, DM-$F_{n+1}(K, H)$ holds. Let δ be any reliable data-minimal discovery method for $(\mathcal{H}, K)$ that starts with H. It follows from Theorem 8 that H is consistent with K. So by Definition 16 of DM-$F_n(K, H)$, we have that there is an ε such that

1. $\forall k$: DM-$F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$, *or*
2. $\exists k$: $K \cap [\varepsilon|k]$ entails that H is false, and $\forall H'$ in $\mathcal{H}$ consistent with $K \cap [\varepsilon|k]$: DM-$F_n(K \cap [\varepsilon|k], H')$.

Case 1: $\forall k$: DM-$F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$. The argument proceeds as in the proof of Theorem 15: A reliable data-minimal method δ must eventually change its mind, say at $\varepsilon|k$, to $\mathcal{H}(\varepsilon)$. But then by the assumption of this case, DM-$F_n(K \cap [\varepsilon|k], \mathcal{H}(\varepsilon))$, so δ requires at least $n+1$ more mind changes by inductive hypothesis.

Case 2: $\exists k : K \cap [\varepsilon|k]$ entails that H is false, and $\forall H'$ in $\mathcal{H}$ consistent with $K \cap [\varepsilon|k]$: DM-$F_n(K \cap [\varepsilon|k], H')$. Let k be the first time that witnesses the condition of this case. Since $\varepsilon|k$ falsifies H given K, and δ is data-minimal, it follows from Theorem 8 that δ changes its mind at $\varepsilon|k$, to a hypothesis H' that is consistent with $K \cap [\varepsilon|k]$. But then we have that DM-$F_n(K \cap [\varepsilon|k], H')$, so by inductive hypothesis, δ requires at least $n+1$ more mind changes. Hence in either case, δ requires at least $n+2$ mind changes.

($\Leftarrow$) Suppose that (K, H) is not a data-minimal $n+1$-feather, that is, $\neg$DM-$F_n(K, H)$. At each point $\varepsilon|k$ for which there is some H' in $\mathcal{H}$ such that $(K \cap [\varepsilon|k], H')$ is not a data-minimal n-feather (i.e., $\neg$DM-$F_n(K \cap [\varepsilon|k], H')$), apply the inductive hypothesis to $(K \cap [\varepsilon|k], H')$ and choose a method $\delta'_{\varepsilon|k}$ and a hypothesis $H_{\varepsilon|k}$ with the properties that

1. $\delta'_{\varepsilon|k}(\emptyset) = H_{\varepsilon|k}$;
2. $\delta'_{\varepsilon|k}$ identifies a correct hypothesis from $\mathcal{H}$ given $K \cap [\varepsilon|k]$ with at most n mind changes;
3. $\delta'_{\varepsilon|k}$ is data-minimal given $K \cap [\varepsilon|k]$.

If H is consistent with $K \cap [\varepsilon|k]$, I modify $\delta'_{\varepsilon|k}$ as follows: Choose $\tau_{\varepsilon|k} \in K \cap [\varepsilon|k] \cap H$. Set $\delta_{\varepsilon|k}(e) = K \cap [e] \cap H$ if $e \subset \tau$, and $\delta_{\varepsilon|k}(e) = \delta'_{\varepsilon|k}$ otherwise. Note that $\delta_{\varepsilon|k}$ is data-minimal and reliable given $K \cap [\varepsilon|k]$ since $\delta'_{\varepsilon|k}$ is.

Since (K, H) is not a data-minimal $n+1$-feather, H is consistent with K. Choose a data stream $\tau \in K$ that makes H true. Now define a data-minimal discovery method δ that reliably identifies a correct hypothesis from $\mathcal{H}$ given K with no more than $n+1$ mind changes:

1. If $e \subset \tau, \delta(e) = K \cap [e] \cap H$;
2. Else if there is a time k such that
   (a) $0 < k \leq lh(e)$ and
   (b) $(K \cap [e|k], H')$ is not an n-feather for some H' in $\mathcal{H}$

   then let k be the least such time and conjecture $\delta_{\varepsilon|k}(e)$.
3. else conjecture H.

By definition δ starts with H, i.e. $\delta(\emptyset) = K \cap [e] \cap H$. I show that δ and identifies the correct hypothesis from $\mathcal{H}$ using no more than $n+1$ mind-changes. Let $\varepsilon \in \mathcal{K}$.

Case 1: Clause 1 always obtains along ε. Then $\varepsilon = \tau$, and so δ stabilizes to the correct hypothesis H along ε (immediately).

Case 2: Clause 1 fails at some point m along ε . I consider two further cases.

Case 2a: Clause 2 is satisfied at some k along ε; let k be the first such time. Two more subcases.

Case 2a1: $k \geq m$. Then δ conjectures H until $\varepsilon|k$ (by Clause 3) and then follows $\delta_{\varepsilon|k}$, which identifies the correct hypothesis from $\mathcal{H}$ given $K \cap [\varepsilon|k]$ along ε. If $\varepsilon = \tau_{\varepsilon|k}$, then δ again does not change its mind along ε at all. Otherwise $\delta_{\varepsilon|k}$ might change its mind at some time $k' \geq k$ from H to follow $\delta'_{\varepsilon|m}$, and thereafter requires at most n mind-changes . Hence δ identifies the correct hypothesis along ε using at most $n+1$ mind-changes.

Case 2a2: $k < m$. Then $\delta_{\varepsilon|k}$ projects H along τ until $\tau|m = \varepsilon|m$. Thereafter $\delta_{\varepsilon|k}$ follows $\delta'_{\varepsilon|k}$.

Case 2b: Clause 2 always fails along ε. Then by the definition of a DM-$n+1$-feather, H is true on ε. By construction δ stabilizes to H along ε (immediately).

Thus in all cases, δ converges to the truth. To see that δ is data-minimal, note that by inductive hypothesis δ projects its conjecture at any evidence sequence e on which Clause 2 obtains. So the only case to consider is when evidence e deviates from τ but Clause 2 does not obtain anywhere along e. This implies that

1. $\delta(e)$ entails H, and
2. $K \cap [e]$ is consistent with H.

The first observation holds by Clause 3 of the definition of δ. The second follows from the second clause of the definition of a DM-$n+1$-feather (Definition 16) and the fact that $K \cap [e]$ is not a DM-n-feather for some $H' \in \mathcal{H}$ (because otherwise Clause 2 obtains, contrary to supposition). If Clause 2 never obtains along some data stream $\varepsilon \in K \cap [e]$, then δ projects H at e by Clause 3 of the definition of δ. Otherwise Clause 2 obtains eventually on all data streams extending e. In particular, Clause 2 must obtain (for the

first time) on some data sequence $e' \supseteq e$ such that $K \cap [e']$ is consistent with H. But then $\delta_{e'}$ and hence δ projects H at e'. Since δ maintains H between e and e' (by Clause 3), δ projects H at e. So δ always takes its conjectures seriously, and thus Theorem 8 implies that δ is data-minimal. ■

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.

[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.

[Freivalds and Smith 1993] Freivalds, R. and Smith C. (1993). "On the Role of Procrastination in Machine Learning," *Information and Computation* 107:237–271.

[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.

[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 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.

[Jain and Sharma 1997] Jain, S. and Sharma, A. (1997). "Elementary Formal Systems, Intrinsic Complexity and Procrastination", *Information and Computation* 132:65–84.

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

[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.

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

[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.

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

[Schulte forthcoming] Schulte, O. (forthcoming). "Means-Ends Epistemology". *The British Journal for the Philosophy of Science.*

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

[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).*

[Sober 1994] Sober, E. (1994). "No Model, No Inference: A Bayesian Primer on the Grue Problem," in *Grue!*, ed. Stalker, D. Open Court: Chicago, Ill.

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