Inductive Inference from Theory-Laden Data

by

Kevin T. Kelly, Clark Glymour

February 1991

Report CMU-PHIL-18

Carnegie Mellon

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

2-25-91

Inductive Inference from Theory Laden Data

Kevin T. Kelly
Carnegie Mellon University

Clark Glymour
Carnegie Mellon University
University of Pittsburgh

1. Introduction

Consider the problem of an investigator who knows some things, who receives a stream of data, and who would like to be on the path to the truth about some matter. This is a familiar philosophical picture of the scientist's predicament; an image to be found in the work of such philosophers as Plato, Aristotle, Bacon, Mill, Peirce, Popper, and Reichenbach. In the 1960's, Hilary Putnam [17] and E. Mark Gold [6], [7] independently combined this model of the problem of inquiry with recursion theory to found a new subject, now known as formal learning theory or computational learning theory. The marriage of inductive methodology with recursion theory has proved to be very fruitful (for reviews c.f. [2], [3], and [15]). Formal learning theorists routinely separate solvable inductive problems from unsolvable problems, assess traditional methodological rules in terms of their implications for reliability, prove various inductive frameworks to be equivalent or inequivalent, and assess the effects on the intrinsic difficulty of induction of such factors as noisy data, hypothesis language syntax, and weaker and stronger notions of success . Lately, the theory has been adapted to the topic of inferring theories expressed in logical languages from data of various kinds [11], [12], [15].

Learning theoretic analysis assumes that there is a fixed language of inquiry and that there is fixed, true data to help us get to the truth in this language. Both of these assumptions are firmly rejected by many prominent philosophers of science [8] [10] [13] [19] [20] [21]. According to these philosophers, truth, syntax, and observability change during "scientific revolutions" or major breaks in scientific traditions. For example, in the move from Newtonian mechanics to the special theory of relativity, the meaning and even the syntactic valence of the relation "simultaneous with" changed. This change forced subtle changes in the meanings of many other Newtonian terms, including "mass" and "energy". So evidence phrased in these terms has a different meaning for scientists who hold different theories. Thus, philosophers often say that evidence is ineluctably "theory laden", and truth is relative to a "conceptual framework" or to a "system of beliefs".

More generally, the issue is *relativism*, the thesis that truth, meaning, and observability can shift as a function of what the process of inquiry does, or of what the inquirer believes. There are many stories in the annals of mataphysics about what truth depends upon (e.g. concepts, conventions, networks of inferential dispositions, scientific social units, history, community norms, experimental regimens, the scientist's behavior) and about how the dependence actually works.

Philosophical speculation about meaning change is not the only source of relativistic concerns. The Copenhagen interpretation of quantum mechanics teaches that the classical nature of a system (e.g. wave vs. billiard ball) is determined by the application of an observational procedure which "collapses" the wave system into a particle system, and that there is no unique truth to the matter. Relativity theory, of course, asserts that the truth about simultaneity is relative to the reference frame of the observer. Relativity is straightforwardly rampant in the social sciences. For example, the dire predictions of a panel of expert economists can precipitate "self-fulfilling prophecies" in light of the effect of these predictions on individual investors. Herds of anthropologists can alter the social relations of a small tribe by their mere presence. In these cases, the truth dependency is a straightforward causal relation between the deeds and statements of the scientist and the nature of the system under study.[1]

---

[1] From the examples, it is clear that whether or not inquiry is relativistic is itself relative to the aspirations of inquiry. For example, Einstein did not aspire merely to provide a kinematics true of some reference frame or other. He aspired, rather, to provide a full theory of the dependency itself. One way to dodge the issue of relativistic inquiry is to insist that science *always* aim for a complete theory *of the dependency*. Thus, the anthropologists would be enjoined to aim at a theory of how tribal societies respond to Western anthropologists, rather than merely at descriptions of lineage, social

Whether relativism looms due to metaphysical or semantic considerations or merely to concrete worries about uncontrollable interference between the investigator's states and those of the subject matter under study, the basic issue posed to inductive methodology remains the same. What sense does it make for science to pursue the truth when the truth feints as science lunges?

Philosophers today tend to stress short-run, rational agreement among investigator. For realists, who deny any significant relativism, convergence to the truth is a way of eventually arriving at rational agreement, since there is one truth and those who find it will agree. But relativism severs this connection, since two investigators could both arrive at their own, distinct truths without ending up in agreement. The pessimists conclude that science is an irrational process because individual theory choices cannot be agreed upon by all participants. The optimists attempt to show that there is still sufficient basis for agreement in actual case studies.

We propose a different approach. Unlike the optimists, we concede that relativism may be rampant, for all we know. And unlike both the optimists and the pessimists, we focus squarely on reliable convergence to the truth, rather than on rational agreement based on shared data in the short run. Even though truth is not unique, and even though it may depend upon us in subtle ways we do not understand *a priori*, it may still be possible for a scientist to converge to his own truth.

There are perhaps two main reasons why getting to the relative truth has not captured the imagination of relativists. First, the goal seems too hard. It is one thing, so the story goes, to find the truth in a fixed, spoon-fed framework of concepts, but it is quite another to search among different frameworks to find one that is suitable. But this observation is flawed. The assumption that truth is fixed does not make convergence to the truth easy, as the many negative results in formal learning theory already attest. And finding the truth in the system of one's choice can make the problem of finding the truth easier, for the scientist may sidestep inductive difficulties by altering auxiliary assumptions, concepts, and so forth.

---

organization and so forth. And if the philosophers are right about meaning change, then all scientists would have to append comprehensive theories of reference to their current hypotheses in physics, chemistry, and biology.

Thus we are led to the second reason for concern, namely, that getting to the relative truth may be too easy, and thus unworthy as a proposed aim for science. If truth depends upon you, then what is the point of careful inquiry? Just make your favorite theory true, and be done with it! Admittedly, if the scientist has absolute control over truth, then inductive inquiry is trivial. But *radical subjectivism*, the view that the scientist knows how to make any given theory true, is just one, trivial form of relativism. If we drop the assumption that the scientist knows how to make any theory he pleases true, the situation becomes interesting, even when he has the power to make any theory true by doing something (he knows not what). If there are constraints on his powers as well as on his knowledge, then the situation becomes still more interesting. It is not hard to imagine that in such circumstances, relativism could be rampant, and yet convergence to the relative truth might be highly non-trivial.[2] The project of this paper is to extend learning-theoretic analysis to such problems.

In this paper, we examine three precise notions of getting to the relative truth. The first requires that the scientist converge to a conceptual framework and to the correct truth value of a given hypothesis in this framework. We call this notion of success *scheme-stable truth detection*, since the scientist must eventually settle down to a particular conceptual scheme, and then must converge to the correct truth value of a given hypothesis. More leniently, we may permit the scientist to have conceptual revolutions forever, just so long as there is a time after which the truth value of the hypothesis under investigation is fixed. Then the scientist must discover this fixed truth value. We refer to this sense of success as *truth-stable truth detection*. Finally, we may be so liberal as not even to require that the scientist eventually stabilize the truth value of the hypothesis under investigation. Instead, we require only that after some time, he always gets the correct truth value of the hypothesis for the conceptual scheme he currently adopts. This notion of success we refer to as *truth detection* simpliciter.

The ontology of our setting for relativism is simple. There is some set C of things we call conceptual schemes. In applying our framework, these may actually stand for conceptual schemes, or for anything else that syntax, truth, and observability are

---

[2]The fashionable doctrines of holism and incommensurability fit nicely with the view that the scientist does not know *a priori* how his acts will affect meaning, truth and observability.

alleged to depend upon. No mathematical structure is imposed upon this set. We also assume a fixed alphabet Σ. This is innocuous, since Σ can contain every typographical symbol ever used and ever to be used. The set Σ^* of all finite strings of characters in Σ provides the raw material for evidence sentences and hypotheses. We may think of well-formedness, truth, and observability as determining subsets of Σ^*. Each such division of Σ^* according to syntax, truth, and observability is referred to as a *world of inquiry*. Since syntax, truth, and observability depend upon conceptual scheme, we represent the possible such dependencies as functions from conceptual schemes to worlds of inquiry. Such functions are referred to as *worlds-in-themselves.*

A scientist does empirical research because he is uncertain about something. In our framework, his uncertainty is represented as a set of possible worlds-in-themselves, one of which is actual. We refer to a set of possible worlds-in-themselves as a *relativistic system.* The "conceptual scheme" represents everything about the world of inquiry that depends upon the scientist and the world-in-itself represents everything about the world of inquiry over which the scientist has no control.[3]

The notion of a relativistic data stream may seem problematic. The data received depends upon the conceptual scheme of the scientist, but the conceptual scheme of the scientist depends upon the data he receives. But there is no difficulty if we stretch the circle into a spiral through time. The datum received by the scientist at stage n of inquiry is true and observable in the world-of-inquiry determined by his choice of conceptual scheme at stage n-1. Since the data received at different times may come from different worlds of inquiry, the data received at one time may be false, nonsensical, or "metaphysical" (empirically indeterminable) at another time. But the scientist can remember what sort of strings he *took to be* observable and true when his conceptual schemes were different.

---

[3]Traditionally, the conceptual scheme has been the "subjective" component of truth and the world-in-itself has been the "objective" or "mind-independent" component. But when the subjective component of truth cannot be manipulated at will (by a scientist following a method) then from a methodological perspective this subjective component raises no new questions for learning theory. Thus, a philosopher like Kant, who held that the mind's contribution to truth is invariant for our species is a naive realist so far as the logic of inquiry is concerned, since truth is fixed once for all for any given scientist.

The main result of this paper is a demonstration of necessary and sufficient conditions for the existence of a method that can detect the semantic status of a given string over a given relativistic system. This theorem may be thought of as a generalization of Angluin's necessary and sufficient conditions for language acquisition from positive data [1]. The positive side of the proof, together with establishing completeness of our technique for proving relativistic problems unsolvable, involves the construction of a relativistic inductive method. The negative side of the proof may be viewed as a completeness theorem for this method, in the following sense: given a specification of how truth, syntax, and observability can possibly depend upon conceptual scheme (i.e. given a relativistic system) and given a string in Σ^*, the method will detect the semantic status of the string if and only if it is possible to do so. Similar results are given for truth-stable and scheme-stable truth detection.

Section 2 presents the concepts that make precise our three notions of convergence to the relative truth. In Section 3, we prove a locking sequence lemma for each of these notions of success. The locking sequence lemmas are useful in proving certain relativistic inductive inference problems to be unsolvable. In Section 4, we apply the locking sequence lemmas to prove a characterization theorem for each of our three notions of convergence to the relative truth. The characterization theorems provide necessary and sufficient conditions for convergence to the relative truth, stated as structural properties of the set of objects under study. The proofs of the characterization theorems involve the construction of universal inductive methods that get to the relative truth if any method can. In Section 5, we apply these characterization theorems to prove that our three notions of success are distinct in terms of the difficulties of the problems they pose to the scientist. Finally, Section 6 draws some philosophical morals from the results, and suggests some open paths for research in the area.

2. Notation and Definitions

2.1 Sequence Operations

Unless it is stated otherwise, the following operations are defined both for finite sequences and for ω-sequences.

last(τ) = the object occurring at the end of finite sequence τ.

τ_n = the item occurring in position n of sequence τ.

decr(τ) = the result of deleting the last item in finite sequence τ.

$\sigma*\tau$ is the concatenation of sequence τ and finite sequence σ.

rng(τ) = the set of all objects occurring in τ.

$\tau[n]$ = the initial segment of sequence τ of length n.

2.2 Metaphysics

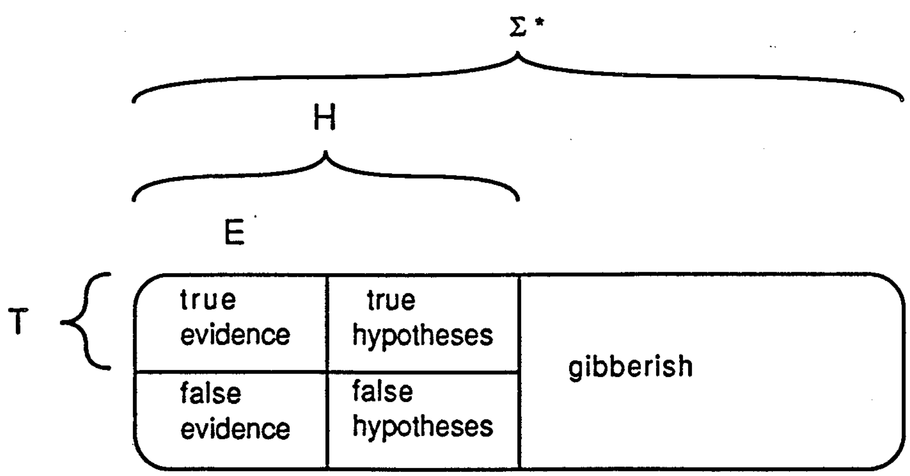
Let Σ be a countable alphabet.

Let Σ^* = the set of all finite strings on Σ (including the empty string).

A *hypothesis* is some string $s \in \Sigma^*$.

A world-of-investigation $\mathfrak{W}$ is a triple of sets <H, E, T> such that
$T, E \subseteq H \subseteq \Sigma^*$.

H represents the *well-formed* hypotheses in $\mathfrak{W}$. Note that according to our usage, a hypothesis need not be well-formed. Whether or not a given hypothesis is well-formed is one of the things the scientist must figure out for himself. E represents the "evidence language" subset of H in $\mathfrak{W}$. T represents the subset of H that is true in $\mathfrak{W}$.



A *relativistic system* is just some triple <F, C, W> where

C is an arbitrary set.
W is some set of worlds-of-investigation.

F is some set of total maps $f: C \rightarrow W$.

Let a relativistic system <F, C, W> be given.

A *conceptual scheme* is a member of C.

A *world-in-itself* is a member of F.

A world-in-itself is a specification of a particular way in which truth, observability, and syntax depend upon conceptual scheme. A relativistic system specifies the set of such dependencies that may, for all we know, be the case.

For example, let Σ = the standard keyboard symbols, let C = {a, b}, let $W = \{\mathfrak{W}_1, \mathfrak{W}_2\}$ where $\mathfrak{W}_1 = \langle H_1, E_1, T_1 \rangle$, and $\mathfrak{W}2 = \langle H_2, E_2, T_2 \rangle$, and where H_1 = the well-formed first-order sentences on a non-logical vocabulary consisting of unary predicate P, H_2 = the well-formed first-order sentences on a non-logical vocabulary consisting of binary predicate Q, E_1 = the atomic sentences of H_1, E_2 = the literal sentences of H_2, T_1 = the truths in H_1 according to some chosen relational structure for H_1, and T_2 = the truths in H_2 according to some chosen relational structure for H_2, except that the truth conditions are all reversed so that negations count as assertions and assertions as negations. Let $f_1 = \{\langle a, \mathfrak{W}_1 \rangle, \langle b, \mathfrak{W}_1 \rangle\}$ and let $f2 = \{\langle a, \mathfrak{W}_2 \rangle, \langle b, \mathfrak{W}_1 \rangle\}$, and let $F = \{f_1, f_2\}$. In f_1, truth, syntax and observability do not depend upon conceptual scheme, since f1 is a constant function. In f_2, vocabulary, observability, and the meaning of negation all depend in a radical way upon conceptual scheme. The triple <F, C, W> is a relativistic system.

In subsequent discussions, F, C, and W are all to be understood as constituents of some fixed, arbitrary, relativistic system.

The *truth values* are {T, F, U}.

Define $tv: \Sigma^* \times W \rightarrow \{T, F, U\}$ as follows, where $\mathfrak{W} = \langle H, E, T \rangle$:

$tv(s, \mathfrak{W}) = T$ if $s \in T$
$tv(s, \mathfrak{W}) = F$ if $s \in H - T$.
$tv(s, \mathfrak{W}) = U$ if $s \notin H$.

Intuitively, **T** means "true", **F** means "false" and **U** means "no truth value".

2.3 Data Presentations

a *data presentation* is an ω-sequence of strings in Σ^*.

SEQ = the set of all finite segments of data presentations.

2.4 Truth Detectors

A *truth detector* is a function δ: $\Sigma^* \times SEQ \to C \times \{T, F, U\}$.

A truth detector is given a hypothesis to investigate and is given some finite sequence of data. On the basis of this data, the truth detector is required to produce a choice of conceptual scheme and a guess as to the truth value of the given hypothesis. Think of the truth detector as investigating a single hypothesis as the data sequences get ever larger through the course of inquiry. Since we will not examine questions of computability in this paper, we may assume without loss of generality that truth detectors are total functions.

Suppose $\delta(s, \sigma) = \langle c, b \rangle$, where s is a hypothesis, $\sigma \in SEQ$, $c \in C$ and $b \in \{T, F, U\}$. Then

$\delta(s, \sigma)_1 = c$

$\delta(s, \sigma)_2 = b$.

2.5 Data Presentations for Worlds and Detectors

Let t be a data presentation. Then define:

$t[\delta, c, s]_n = t_k$ where k is the nth position in t such that $\delta(s, t[k-1])_1 = c$.

$t[\delta, c, s]$ = the sequence $\langle t[\delta, c, s]_1, t[\delta, c, s]_2, \ldots, t[\delta, c, s]_n, \ldots \rangle$

$t[\delta, c, s]$ may be thought of as the result of deleting each position in t such that δ does not choose conceptual scheme c at the previous position in t. Hence, $t[\delta, c, s]$ is the total data presented in t for conceptual scheme c when the truth detector choosing conceptual schemes is δ.

$ev(\mathfrak{W}) = T \cap E$, where $\mathfrak{W} = \langle H, E, T \rangle$.

$ev(\mathfrak{W})$ should be thought of as the total evidence true of world-of-investigation $\mathfrak{W}$.

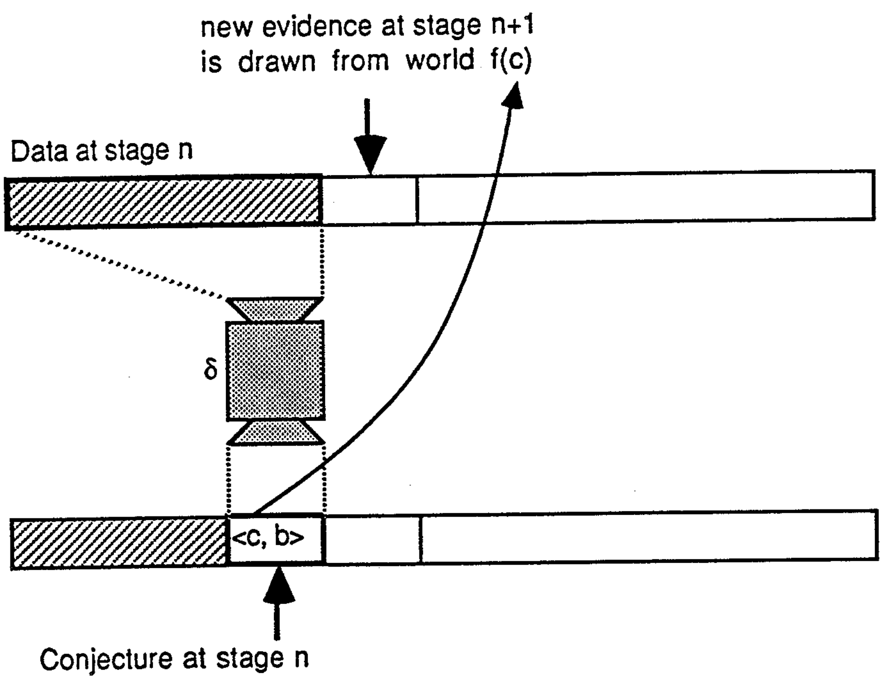
Let δ be a truth detector, s be a hypothesis, f be a world-in-itself, and c be a conceptual scheme.

$\sigma \in$ SEQ is *sound* for δ, f, s $\Leftrightarrow \forall c$, $rng(\sigma[\delta, c, s]) \subseteq ev(f(c))$.

data presentation t is *sound for* δ, f, s $\Leftrightarrow \forall c \in C$ $rng(t[\delta, c, s]) \subseteq ev(f(c))$.

Soundness requires that the data read at stage n+1 be true with respect to the conceptual scheme produced by δ at stage n. This relationship is clarified in the following figure.

Relativistic Data Presentations



We require not only that data presentations be sound, but also that they be complete. Completeness demands that all the data in world-of-inquiry f(c) be presented to δ only when δ selects c infinitely often. Otherwise, an infinite data set for f(c) could not possibly be presented to δ.

data presentation t is *complete for* δ, f, s $\Leftrightarrow$
$\forall c \in C$, $t[\delta, c]$ is infinite $\Rightarrow rng(t[\delta, c, s]) = ev(f(c))$.

data presentation t is *for* δ, f, s $\Leftrightarrow$
t is complete for δ, f, s and t is sound for δ, f, s.

$PRES(f, \delta, s) = \{t: t \text{ is for } \delta, f, s\}$.

2.6 Reliable Detection

Let $\langle F, C, W\rangle$ be a relativistic system.

Now we define various notions of convergence to a truth value on a data presentation.

> δ *converges* to c, b on t, s $\Leftrightarrow$
> $\exists n\ \forall m > n\ \delta(s, t[m])_1 = c$ and $\delta(s, t[m])_2 = b$
>
> δ *converges* to b on t, s $\Leftrightarrow$
> $\exists n\ \forall m > n\ \delta(s, t[m])_2 = b$.
>
> δ *detects* s on t $\Leftrightarrow$
> $\exists n\ \forall m > n\ \delta(s, t[m])_2 = tv(s, f(\delta(s, t[m])_1))$.

Let $\langle F, C, W\rangle$ be a relativistic system. Let $f \in F$, and let s be a hypothesis. Let $t \in PRES(f, \delta, s)$. Let δ be a truth detector.

> δ *scheme-stably* detects s on t $\Leftrightarrow$
> δ detects s on t and
> $\exists c, b$ such that δ converges to c, b on t, s.
>
> δ *truth-stably detects* s on t $\Leftrightarrow$
> δ detects s on t and
> $\exists b$ such that δ converges to b on t, s.
>
> δ [scheme-stably, truth-stably] *detects* s in f $\Leftrightarrow$
> $\forall t \in PRES(\delta, f, s)$, δ [scheme-stably, truth-stably] detects s on t.
>
> δ [scheme-stably, truth-stably] *detects* s over F $\Leftrightarrow$
> $\forall f \in F$, δ [scheme-stably, truth-stably] detects s in f.

Detection requires that after some time, δ always produces a truth value that is correct for the conceptual scheme produced at the same moment as this truth value. Truth-stable detection requires, in addition, that after some time the truth value of s is stabilized by δ's choices of conceptual scheme. Scheme-stable detection requires, in addition, that δ stabilizes to a unique conceptual scheme.

3. Locking Sequence Lemmas

The basic notion of locking sequences is familiar to learning theorists working with non-relativistic systems [15], [16]. In the present section, we generalize the notion to the relativistic case.

Loosely speaking, a data sequence σ *locks* a scientist onto a world-in-itself f if the scientist produces only conjectures correct for f upon seeing σ and he continues to produce conjectures correct for f until he sees data unsound for f. That is, a locking sequence is data that the scientist finds absolutely compelling for f until further data proves that he is wrong.

A locking sequence lemma tells us that whenever a scientist succeeds in a world-in-itself, there is a finite data sequence that locks the scientist onto this world. For each of our three senses of getting to the relative truth we isolate a respective notion of locking sequence and we prove a corresponding locking sequence lemma. The locking sequence lemmas are useful in proving that no reliable scientist exists for a given relativistic system. They will also serve as lemmas in the proofs of our respective characterizations of truth detectability, truth-stable detecability, and scheme-stable detectability.

3.1 Locking Sequence Lemma for Detection Simpliciter.

$\sigma \in$ SEQ is *locking* for δ, f, s $\Leftrightarrow$

σ is sound for δ, f, s and

$tv(s, f(\delta(s, \sigma)_1)) = \delta(s, \sigma)_2$ and

$\forall \tau \in$ SEQ if

(a) $\sigma \subseteq \tau$ and

(b) τ is sound for δ, f, s

then $\delta(s, \tau)_2 = tv(s, f(\delta(s, \tau)_1))$.

Lemma 3.1: If δ detects s in f then

$\forall \gamma$ sound for f, δ,

$\exists \sigma \in$ SEQ such that

$\gamma \subseteq \sigma$ and

σ is locking for δ, f, s .

Proof: see Appendix. ■

3.2 Locking Sequence Lemma for Truth-Stable Detection

$\sigma \in$ SEQ is *truth-locking* for δ, f, s, b $\Leftrightarrow$
σ is sound for δ, f, s and
$\delta(s, \sigma) = \langle c, b \rangle$ and
$tv(s, f(c)) = b$ and
$\forall \tau \in$ SEQ if
(a) $\sigma \subseteq \tau$ and
(b) τ is sound for δ, f, s
then $\delta(s, \tau)_2 = b = tv(s, f(\delta(s, \tau)_1))$.

Lemma 3.2: If δ truth-stably detects s in f, then
$\forall \gamma$ sound for f, δ,
$\exists \sigma \in$ SEQ, $\exists b \in \{T, F, U\}$ such that
$\gamma \subseteq \sigma$ and
σ is truth-locking for δ, f, s, b.

Proof: Similar to the proof of Lemma 3.2 except that during each fooling stage 2n, we search for a τ such that either $\delta(s, \tau)_2 \neq \delta(s, t\{2n-1\})_2$ or $\delta(s, \tau)_2 \neq tv(s, f(\delta(s, \tau)_1))$. ■

3.3 Locking Sequence Lemma for Scheme-Stable Detection

$\sigma \in$ SEQ is *scheme-locking* for δ, f, s, c $\Leftrightarrow$
σ is sound for δ, f, s and
$\delta(s, \sigma) = \langle c, tv(s, f(c)) \rangle$ and
$\forall \tau \in$ SEQ if
(a) $\sigma \subseteq \tau$ and
(b) τ is sound for δ, f, s
then $\delta(s, \tau)_2 = \delta(s, \sigma)_2$ and $\delta(s, \tau)_1 = c$.

Lemma 3.3: If δ scheme-stably detects s in f then
$\forall \gamma$ sound for f, δ,
$\exists \sigma \in$ SEQ, $\exists c \in C$ such that
$\gamma \subseteq \sigma$ and
σ is scheme-locking for δ, f, s .

Proof: Similar to the proof of Lemma 3.2, except that during each fooling stage 2n, we search for a τ such that $\delta(s, \tau)_2 \neq \delta(s, t\{2n-1\})_2$ or $\delta(s, \tau)_1 \neq \delta(s, t\{2n-1\})_1$. ■

4. Characterization Theorems

For each string and relativistic system, either the string is reliably detectable over the system or not. A characterization theorem provides necessary and sufficient conditions for detectability entirely in terms of the structure of relativistic systems. In this section, we present characterization theorems for truth detectability, for truth-stable detectability, and for scheme-stable detectability. The characterizations do not hold for arbitrary relativistic systems. The result for scheme-stable detectability is valid only for systems in which C and F are countable. The characterizations for truth detectability and truth-stable detectability are valid only when C is finite and F is countable. The issues that arise when we relax the finitude of C are discussed in Section 6.

There is a useful, alternative perspective on the characterization theorems. Think of an *adaptive detector* as a map $\phi(F, s, \sigma)$, such that the result of fixing argument F is a truth detector. An adaptive detector may be thought of as using "background knowledge" F to try to detect s. For each adaptive detector, there is some range of pairs <F, s> such that $\phi(F, \_, \_)$ detects s over F. We may then speak of an adaptive detector as being *complete* if for each pair <F, s>, it detects s over F if some truth detector can. From this point of view, each of our characterization theorems involves the construction of a complete adaptive detector. Accordingly, the proofs of the characterization theorems may be viewed as completeness theorems for the adaptive detection systems so constructed.

4.1 Truth Detectability Characterized

4.1.1 Definitions

The following sequence of definitions is necessary in order to state the characterization theorem.

A *clue* is a finite subset of $C \times \Sigma^*$.

clue D is *sound for* f $\Leftrightarrow \forall <c, e> \in D, e \in ev(f(c))$.

clue D *involves* $c \in C \Leftrightarrow \exists e \in \Sigma^*$ such that $<c, e> \in D$.

cs(D) = $\{c \in C$: D involves c$\}$.

clue$_\delta(\sigma) = \{<c, e>: \exists n\ 1 \leq n \leq \text{length}(\sigma)$ and $\sigma_n = e$ and $c = \delta(s, \sigma[n-1])_1\}$.

D *is contained in* σ mod δ, s $\Leftrightarrow D \subseteq \text{clue}_\delta(\sigma)$.

Let <F, C, W> be a relativistic system and let $F' \subseteq F$.

$az_s(F)$ =
the *agreement zone* of F' mod s =
{c: ∀f, g ∈ F' tv(s, f(c)) = tv(s, g(c))}.

$az_s(F)$ = the agreement zone of F mod s.

a *path* is an element of (F X Clue)*.

Let p = <<f_1, D_1>, ..., <f_n, D_n>> be a path.

clues(<<f_1, D_1>, ..., <f_n, D_n>>) = {D_1, ..., D_n}.

worlds(<<f_1, D_1>, ..., <f_n, D_n>>) = {f_1, ..., f_n}.

lastclue(<<f_1, D_1>, ..., <f_n, D_n>>) = D_n.

lastworld(<<f_1, D_1>, ..., <f_n, D_n>>) = f_n.

When p is empty, lastclue(p) and lastworld(p) are undefined.

f *may extend* p mod F, s ⇔
(1) clues(p) is sound for f and
(2) if p is non-empty then
∀c ∈ az_s(worlds(p)), ev(f(c)) ⊆ ev(lastworld(p)(c)) and
(3) if p is non-empty then
az_s(worlds(p)) - az_s(worlds(p), f) ≠ ∅ and
(4) az_s(worlds(p), f) ≠ ∅.

A *tree* is a subset of (F X Clue)*.

Let 𝕿 be a tree. We now define the notion of *revolution tree*. Revolution trees are so named because they will instruct our truth detector when to have conceptual revolutions (i.e. when to change his conceptual scheme).

𝕿 is a *revolution tree* mod F, s ⇔
∀p*<f, D> ∈ 𝕿,
(1) f may extend p mod F, s and
(2) cs(D) ⊆ az_s(worlds(p)) and
(3) D is sound for f.

𝕿 is *complete* mod F, s ⇔
(1) <> ∈ 𝕿 and
(2) ∀p ∈ 𝕿 ∀f ∈ F, if f may extend p mod F, s then ∃D p*<f, D> ∈ 𝕿.

𝕿 is *safe* mod F, s ⇔
∀ non-empty p ∈ 𝕿 ∀f ∈ F
if
(1) clues(p) is sound for f and

(2) $az_s(\text{worlds}(p), f) = \varnothing$
then $\exists c \in az_s$ such that
$(\text{worlds}(p))\ \ ev(f(c))\ -\ ev(\text{lastworld}(p)(c)) \neq \varnothing$.

4.1.2 Example: A Complete, Safe, Revolution Tree

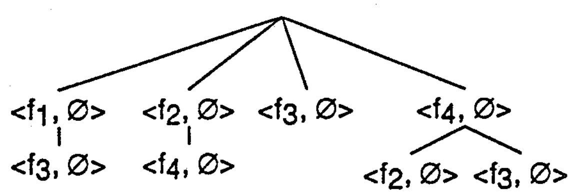
We now provide an example of a complete, safe revolution tree for a finite relativistic system. It should be understood, however, that the concept is in no way limited in application to finite systems. Let relativistic system <F, C, W> be presented by the following matrix, in which each row represents a world-in-itself in F and each column represents a conceptual scheme.

	1	2	3	4
f1	F / $N-\{0\}$	T / $N-\{0\}$	T / $N-\{0\}$	T / $N-\{0\}$
f2	T / N	F / $N-\{1\}$	T / N	F / $N-\{1\}$
f3	T / $N-\{0\}$	F / $N-\{0,1\}$	T / $N-\{0\}$	F / $N-\{0,1\}$
f4	T / N	T / $N-\{1\}$	F / N	T / $N-\{1\}$

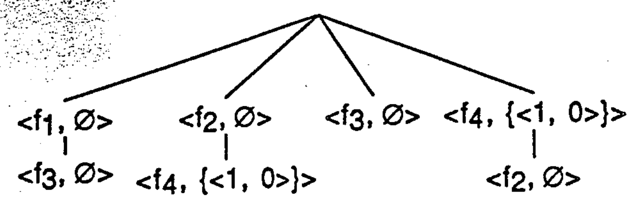
There are four worlds-in-themselves in F, called f_1, f_2, f_3, and f_4. Each world-in-itself is defined over four conceptual schemes, 1, 2, 3, and 4. Cell i,k corresponds to world-of-inquiry $f_k(i)$. Let s be the string whose truth value is to be detected over F. We assume that in system F, the string s is never observable. The truth value in the upper left corner of cell $f_k(i)$ corresponds to the truth value assigned to s in $f_k(i)$. We put $\Sigma^* - \{s\}$ into 1-1 correspondence with the natural numbers. The set of natural numbers in the lower right of cell $f_k(i)$ is the set of all code numbers of strings in $ev(f_k(i))$ (i.e. the well-formed, true, observable strings in world of inquiry $f_k(i)$). This will be our standard representation of relativistic systems in the balance of the paper.

There is a complete, safe revolution tree for F, s. To find it, we apply the following procedure. We start constructing $\mathfrak{T}$ by putting the empty path <> into $\mathfrak{T}$. Thereafter, for each path p in $\mathfrak{T}$, we extend the path by each $f \in F$ that may extend p mod F, s. No path in $\mathfrak{T}$ can have a length greater than 4 by the definition of "may extend" and the fact that $|C| = 4$, so our process of additions terminates at least by then. Next, beginning at the root, we examine each path p in $\mathfrak{T}$ in order to verify safety, under the assumption that each clue in the path is empty. If safety is violated, we add data to some clue along the path until safety is no longer violated by this path. Then we eliminate all paths in $\mathfrak{T}$ that extend p and that involve worlds-in-themselves for which some clue in p is not sound.

Let us proceed: $\mathfrak{T}[0] = \{<>\}$. By the definition of "may extend", each $f \in F$ may extend $<>$ mod F, s. So, if we omit initial segments of paths to avoid clutter, we have $\mathfrak{T}[1] = \{<f_1>, <f_2>, <f_3>, <f_4>\}$. Now it gets a bit more challenging to figure out who may extend whom. Recall that the agreement zone of a singleton path is the set C of all conceptual schemes. By exhaustively considering all pairs of distinct worlds (twelve in all) to see which may extend which, we conclude that $\mathfrak{T}[2] = \{<f_1, f_3>, <f_2, f_4>, <f_3>, <f_4, f_2>, <f_4, f_3>\}$. At the next level, we must pay attention to the agreement zone of each path in $\mathfrak{T}[2]$. Again, checking to see who may extend whom, we have $\mathfrak{T}[3] = \mathfrak{T}[2]$. Since we have arrived at a fixed point in our construction, $\mathfrak{T} = \mathfrak{T}[2]$. Now we assume that each world in $\mathfrak{T}$ is paired with the empty clue.



Now we must add clues to make $\mathfrak{T}$ safe. For each path p in $\mathfrak{T}$, we must check whether there is an $f \in F$ such that $az_s(\text{worlds}(p), f) = \emptyset$ and such that $\forall c \in az_s(\text{worlds}(p))$ $ev(f(c)) \subseteq ev(\text{lastworld}(p))$. No pair of worlds in F has an empty agreement zone, so it suffices if we examine paths of length 2 for possible violations of safety. Checking each path of length 2 in order against every other world in F, we find that f_1 violates safety on paths $<f_2, f_4>$,$<f_4, f_2>$, and $<f_4, f_3>$, and that there are no other violations of safety in $\mathfrak{T}$. Seeing a 0 in the data under conceptual scheme 1 can eliminate f_1 from consideration. So if we add the pair $<1, 0>$ to the clue for f_4 in path $<f_2, f_4>$, then safety is no longer violated by this path. Next consider the path $<f_4, f_3>$. We can't see a 0 under scheme 1 in f_3. Hence, we go up to f_4, where we can see a 0 under 1 and we add $<1, 0>$ to the clue for f_4. This forces us to cut away path $<f_4, f_3>$, and at the same time eliminates the violation of safety in path $\langle f_4, f_2 \rangle$. The resulting complete, safe revolution tree $\mathfrak{T}$ is



That having this tree is tantamount to having a method that detects s over $\langle F, C, W \rangle$ will be proved in the next section.

4.1.3 Characterization Theorem for Relativistic Truth Detectability

Theorem 4.1: If C is finite and F is countable then

s is detectable over F $\Leftrightarrow$
there is a complete, safe, revolution tree mod F, s.

Proof: $\Rightarrow$ Suppose that C is finite, F is countable and that some δ detects s over F. Using δ, we will construct a complete, safe, revolution tree for F, s.

Stage 0: $\mathfrak{T}[0] = \{\langle\rangle\}$.

Stage n+1: $\forall p \in \mathfrak{T}[n]$ of length n, $\forall f$ that may extend p mod F, s, define lock(p, f) inductively as follows: lock($\langle\rangle$, f) is some locking sequence for f. Lock(p^*f, g) is some locking sequence for g that extends lock(p, f), if such a sequence exists. If f may extend p mod F, s, then lock(p, f) exists, by the locking sequence lemma and the fact that lock(decr(p), lastworld(p) is sound for f, δ by the definition of "may extend". For each $p \in \mathfrak{T}[n]$ of length n, define

$$\text{ext}(p) = \{p^*\langle f, \text{clue}_\delta(\text{lock}(p, f))\rangle : f \text{ may extend } p \text{ mod } F, s\}.$$

Now define

$$\mathfrak{T}[n+1] = \mathfrak{T}[n] \cup \bigcup_{\substack{p \in \mathfrak{T}[n] \\ \text{length}(p) = n}} \text{ext}(p)$$

We need to know that $\forall n\ \mathfrak{T}[n]$ is a revolution tree. By construction, condition (1) of the definition of revolution tree is satisfied by each added path. Since $D = \text{clue}_\delta(\text{lock}(p, f))$, we have that (3) D is sound for f, δ. So it suffices to show that (2) for each $p^*\langle f, D\rangle \in \mathfrak{T}[n]$, $cs(D) \subseteq az_s(\text{worlds}(p))$. This is done in the following lemma, which proves a bit more.

Lemma A: Let $p^*\langle f, D\rangle \in \mathfrak{T}[n]$. Then

(1) $cs(D) \subseteq az_s(worlds(p))$ and
(2) D is sound for each $f \in$ worlds(p).

Proof: see Appendix. ■

Suppose there is a stage n at which $\mathfrak{T}[n]$ is not safe for F mod s. Then by the definition of safety, $\exists p \in \mathfrak{T}[n]\ \exists f \in F$ such that (1) clues(p) is sound for f and (2) $az_s(worlds(p), f) = \varnothing$ and (3) $\forall c \in az_s(worlds(p))\ ev(f(c)) \subseteq ev(lastworld(p)(c))$. Pick a locking sequence σ for f, δ that extends lock(decr(p), lastworld(p)). There is one by (1), and by the locking sequence lemma. By Lemma A, we know that for each $g \in$ worlds(p), lock(p, lastworld(p)) is sound for g with respect to δ. Hence, until δ sees evidence unsound for some $g \in$ worlds(p), δ produces conjectures only in $az_s(worlds(p))$. But by (3), no data unsound for some $g \in$ worlds(p) is sound for f if it involves only schemes in $az_s(worlds(p))$. So σ is sound for f and for each $g \in$ worlds(p). Hence, by the definition of locking sequence, once σ is read, δ must produce conjectures correct for each $g \in$ worlds(p). But this is impossible, since by (2) we have that $az_s(worlds(p), f) = \varnothing$.

So there is no stage n at which $\mathfrak{T}[n]$ is not safe for F mod s. Define

$$\mathfrak{T} = \bigcup_{i \in N} \mathfrak{T}[i]$$

$\mathfrak{T}$ is safe mod F, s, because each $\mathfrak{T}[i]$ is. And $\mathfrak{T}$ is complete mod F, s, because every f that may extend a path in $\mathfrak{T}$ is added by some stage $\mathfrak{T}[i]$.
So $\mathfrak{T}$ is a complete, safe revolution tree.

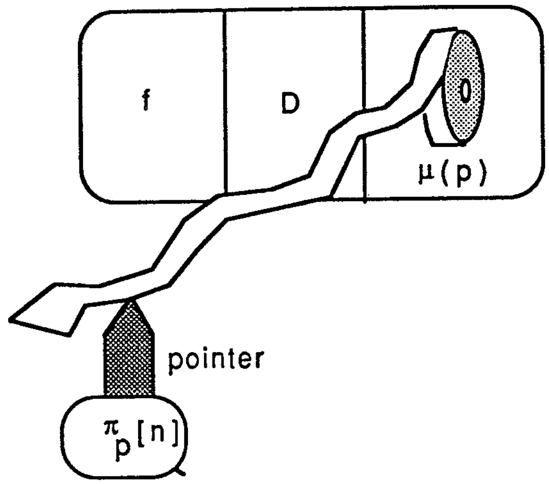
$\Leftarrow$ Suppose that there is a complete, safe, revolution tree $\mathfrak{T}$ mod F, s. We construct a δ that uses $\mathfrak{T}$ to detect s over F.

Let p be a path in $\mathfrak{T}$. Define the *conjecture range* of p, (denoted by cr(p)), as follows:

$$cr(p) = \begin{cases} \{\langle c, 1\rangle : c \in C\} \text{ if p is empty.} \\ \{\langle c, b\rangle : c \in az_s(p)\ \&\ b = tv(s, lastworld(p)(c))\}, \\ \text{otherwise.} \end{cases}$$

Now, for each path p in $\mathfrak{T}$, associate an enumeration $\mu(p)$ of cr(p) in which each element of cr(p) occurs infinitely often. For each such enumeration we supply a pointer whose position in the enumeration at stage n in the operation of our soon to be defined scientist is denoted by $\pi_p[n]$.

The Structure of a Node in $\mathfrak{T}$



Given $\mathfrak{T}$, together with the associated enumerations $\mu(p)$ we define the following method δ, where $\Xi[n]$ is an inductively maintained sequence of clues and $\Omega[n]$ is an inductively maintained sequence of worlds-in-themselves maintained so that $<\Omega[n]_1, \Xi[n]_1>, <\Omega[n]_2, \Xi[n]_2>, ..., <\Omega[n]_k, \Xi[n]_k>$ is always a path in $\mathfrak{T}$. δ may be thought of as maintaining $\Omega[n]$ and $\Xi[n]$ as stacks, so as to perform a depth-first search of $\mathfrak{T}$. Accordingly, define

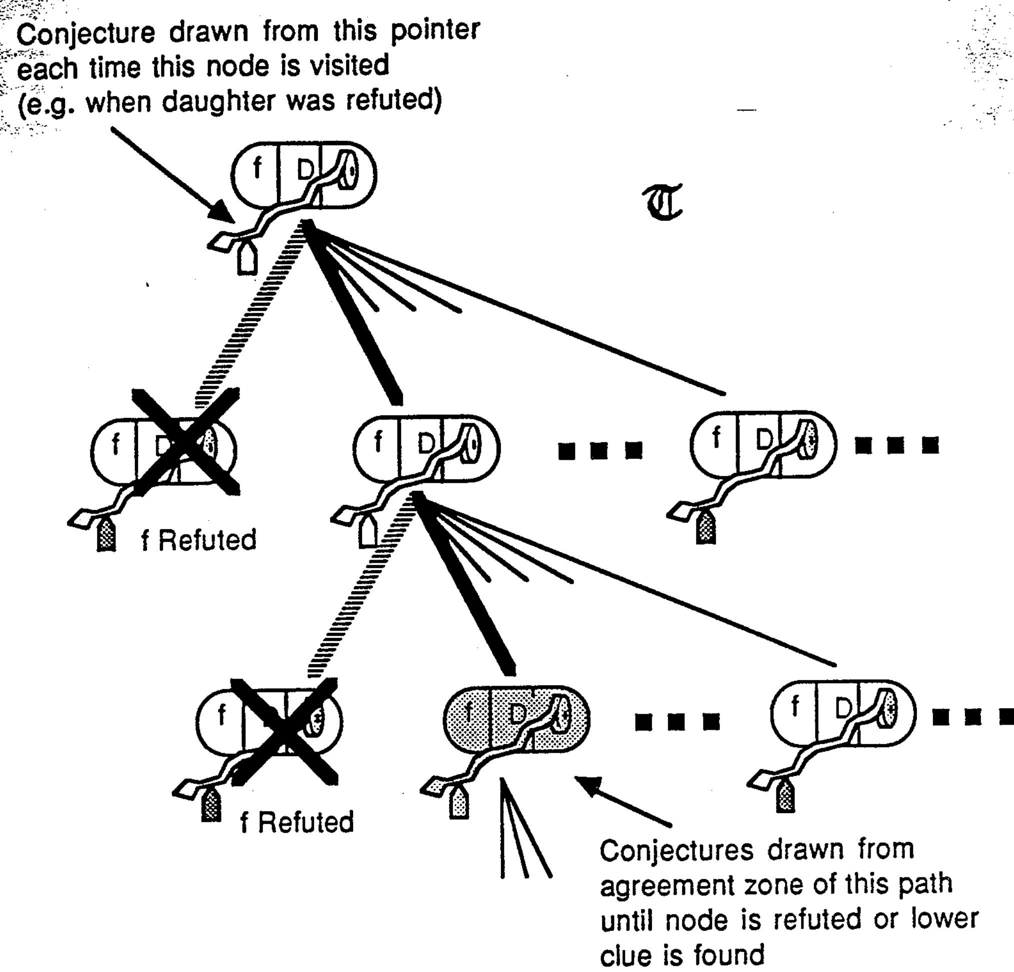
path[n] = $<\Omega[n]_1, \Xi[n]_1>, <\Omega[n]_2, \Xi[n]_2>, ..., <\Omega[n]_k, \Xi[n]_k>$, where k = length(Ξ[n]).

We may think of δ as always "visiting" the end-point of path[n] at stage n. The conjecture of δ at stage n is defined to be the conjecture at position $\pi_{path[n]}[n]$ of enumeration μ(path[n]) of cr(path[n]). So δ produces conjectures from the conjecture tape associated with a node in $\mathfrak{T}$ until either data unsound for the world in its current node is read (in which case δ "pops" up to an ancestor of its current node) or until a clue deeper in $\mathfrak{T}$ is read, (in which case δ "pushes" down below its current node). Each time δ "pops" a node in $\mathfrak{T}$, it produces conjectures according to the parent node it pops to, picking up at the position where the pointer was left when conjectures were last made from the parent node. It is very important that δ start producing conjectures at a parent node from the point where it left off in $\mu(p)$ when it was last producing conjectures from that node. This is because we will need to argue that if infinitely many daughters of a parent <f, D> are considered and rejected, then complete data for f is eventually received. If δ always starts at the beginning of $\mu(p)$ each time <f, D> is visited, then δ might never see any data past the first few entries on $\mu(p)$. The pointers are required to mark the place in $\mu(p)$ last visited by δ, so that δ can remember where he was last time the node was visited.

δ may be thought of as trying to balance two different and somewhat opposed strategies for success. The first strategy is to throw out of contention any world-in-itself that makes the currently observed data false. The second strategy is to find a restricted range of conceptual schemes over which many worlds-in-themselves agree about the semantic status of s. Unfortunately, the former strategy suggests looking at lots of different conceptual schemes to find possible data unsound for a given world-in-itself, and the latter strategy suggests looking at as narrow a range of conceptual schemes as possible to maintain as much agreement among worlds-in-themselves as possible. δ implements the "refutation" strategy by producing conjectures over *all* conceptual schemes in cr(path[n]) while considering path[n]. If there is data unsound for f that can be read over this range of conceptual schemes, δ will find it. δ implements the "agreement maintenance" strategy by producing conjectures *only* in cr(path[n]) while considering path[n] (recall that each world-in-itself in path[n] agrees about the semantic status of s over the conceptual schemes occurring in conjectures in cr(path[n])). Perhaps the most striking result of this paper is that this curious mixture of the two strategies yields an optimally reliable, universal method for inductive inference through conceptual revolutions.

These considerations, together with the following diagram, should help to provide a basic understanding of the method before the details of the formal definition are consulted.

Operation of method δ



Definition of method δ:

Stage 0:
set $\Xi[0]$ = <>;
set $\Omega[0]$ = <>;
for each path p in 𝔗, set $\pi_p[0] = 0$;
set DATA = <>.

Stage n+1:

define path($\Omega[n]$, $\Xi[n]$) = <<$\Omega[n]_1$, $\Xi[n]_1$>, ..., <$\Omega[n]_k$, $\Xi[n]_k$>>, where k = length($\Omega[n]$)

Consider the following two situations:

(a) DATA[n] is not sound for last($\Omega[n]$) with respect to δ.

(b) ∃ pair <f', D'> in 𝔗 such that

(i) D' is contained in DATA[n] with respect to δ and
(ii) path(Ω[n], Ξ[n])*<f', D'> is a path in $\mathfrak{T}$.
(iii) DATA[n] is sound for f' with respect to δ.

If (a) is satisfied, then
set Ω[n+1] = decr(Ω[n]);
set Ξ[n+1] = decr(Ξ[n]);
for each path p in $\mathfrak{T}$, set π_p[n+1] = π_p[n];
set DATA[n + 1] = DATA[n];
go to stage n+2.

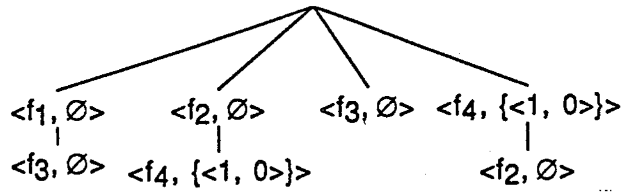
If (b) is satisfied, but (a) is not satisfied, then
let <f', D'> be the least pair (in a fixed enumeration) whose existence is guaranteed by (b);
set Ω[n+1] = Ω[n]*f';
set Ξ[n+1] = Ξ[n]*D';
for each path p in $\mathfrak{T}$, set π_p[n+1] = π_p[n];
set DATA[n + 1] = DATA[n];
go to stage n+2.

If neither (a) nor (b) is satisfied, then
set Ω[n+1] = Ω[n];
set Ξ[n+1] = Ξ[n];
set k = length(Ω[n])
set p = path(Ω[n], Ξ[n]);
conjecture the pair <c, b> at position π_p[n] of μ(p);
set π_p[n+1] = π_p[n] + 1;
$\forall$p' $\neq$ p, set $\pi_{p'}$[n+1] = $\pi_{p'}$[n];
set DATA[n + 1] = DATA[n]*(read next datum);
go to stage n+2.

Let us consider what δ does in a particular example. Recall the following system <F, C, W>.

	c1	c2	c3	c4
f1	F / N-{0}	T / N-{0}	T / N-{0}	T / N-{0}
f2	T / N	F / N-{1}	T / N	F / N-{1}
f3	T / N-{0}	F / N-{0,1}	T / N-{0}	F / N-{0,1}
f4	T / N	T / N-{1}	F / N	T / N-{1}

We have seen that the following tree $\mathfrak{T}$ is a complete, safe, revolution tree for F, s.



Now, let us examine how δ behaves when we hand it $\mathfrak{T}$ and turn it loose in an arbitrary world-in-itself $f \in F$.

Suppose f_1 is actual: Then on no evidence, δ sinks to the bottom of the first path in $\mathfrak{T}$ and sits there producing conjectures that agree for f_1 and f_3 until f_3 is refuted by new data. In fact, the only such conjecture is $<3, T>$. Since no data refuting f_3 will ever be seen under scheme 3, δ converges to a correct conjecture for f_1.

Suppose f_2 is actual: As before, δ sinks to the bottom of the first path in $\mathfrak{T}$ on no evidence. But since f_2 is actual, and since f_2 has total data N under scheme 3, eventually f_3 is refuted by a 0 under scheme 3. At this point, δ pops to the root of the tree and sinks to world f_2 in the second path on no evidence, producing conjectures over all four conceptual schemes according to f_2. Since f_2 is the actual world-in-itself, f_2 will never be refuted. So datum 0 will eventually be read under scheme 1, and δ will then drop to f_4 in the second path, since $\{<1, 0>\}$ is the clue for doing so. The agreement zone of f_2, f_4 is just $\{1\}$. Since f_2 and f_4 have exactly the same data under scheme 1, f_4 is never refuted as δ produces conjecture $\{<1, T>\}$ forever after. So δ succeeds.

Suppose f_3 is actual: As usual, δ plunges blindly to the bottom of path $<f_1, f_3>$ on no evidence, and conjectures $<3, T>$ until data refuting f_3 is seen. But no such data is ever seen, so δ succeeds.

Suppose f_4 is actual: Again, δ plunges to the bottom of path $<f_1, f_3>$ on no evidence and conjectures $<3, T>$ until data refuting f_3 is seen. But f_4 eventually yields such data, at which point δ pops to the tree root and plunges down to f_2 on the second path on the basis of no data. Here, δ produces conjectures agreeing with f_2 over all conceptual schemes until either 0 is read under scheme 1 (causing δ to plunge to f_4), or data unsound for f_2 is read (causing δ to pop to the tree root and plunge to f_3 on no data). The latter will never occur, since f_4 has data everywhere identical to f_2. But eventually, a 0 is read under scheme 1 and δ plunges to f_4. Since no data unsound for f_4 can be read, δ converges to conjectures correct for f_4 at this point.

So by exhaustion, we see that δ succeeds.

Some observations are in order, that will be useful in following the general proof that δ succeeds. First, δ sometimes succeeds by overshooting the actual world-in-itself along a path (e.g. when f_1 or f_2 is actual). This causes no harm, since all subsequent conjectures along a path must still agree with earlier worlds along the path (c.f. Lemma 3).

Second, the tree serves as a sort of computer program branch instruction that tells δ what to conjecture, when to wait for refuting data, and when to wait for a sign that δ should narrow the range of conceptual schemes to be visited infinitely often. Together, the first two paths of $\mathfrak{T}$ amount to the following instruction:

δ's procedure:

Until data unsound for f_3 is read, do repeat conjecture <3, T>.
Until clue {<1, 0>} is read do repeat conjectures agreeing with f_2.
Repeat conjecture <1, T>.

This is a sensible way to handle problem F, and it emerges effectively from our procedure for constructing $\mathfrak{T}$.

The order in which δ considers paths in $\mathfrak{T}$ is irrelevant to the correctness of δ's performance. But the program read off of $\mathfrak{T}$ may be very different under a different ordering of paths. In our example, δ only has to examine the first two branches of $\mathfrak{T}$ to succeed. But the unused branches may be necessary to ensure δ's success under some other ordering.

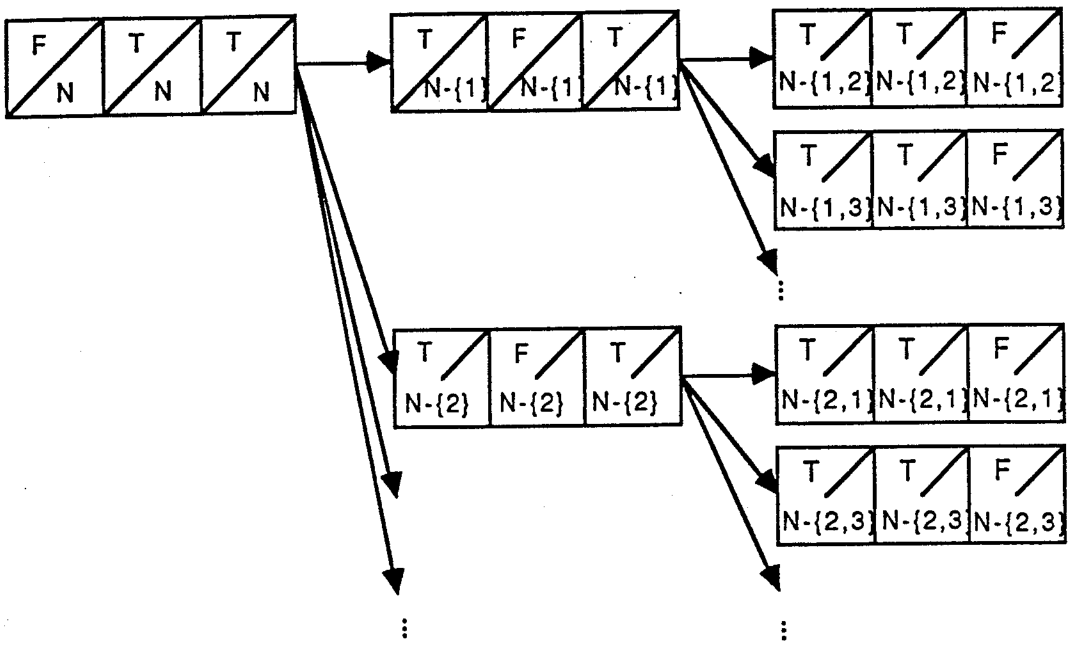
To complete the proof of the theorem, it suffices to show:

Lemma B: If $\mathfrak{T}$ is a complete, safe revolution tree mod F, s then δ detects s over F.

Proof: see Appendix. ■

4.1.4 An Infinite Example

In this section we prove a negative result about an infinite relativistic system by applying Theorem 4.1. Consider problem G depicted in the following figure[4]. G is presented as a tree, because its infinite structure would be obscured if depicted in a finite fragment of a matrix. G involves three conceptual schemes. Each daughter world-in-itself has its total data sets included in the total data sets of its parent. Finally, each daughter differs in its truth assignment for s in exactly one place where all its parents agree.



Fact 4.1.4: s is not detectable over G.

Proof: Call the root of the tree g. Let $\mathfrak{T}[1]$ = {<>, <g, D>}, where D is sound for g. $\mathfrak{T}$ is not yet complete, for whatever data occurs in D, we can choose a daughter f of g that may extend <g, D> mod G, s. Choose D' sound for f and set $\mathfrak{T}[2]$ = {<>, <<g, D>>, <<g, D>, <f, D'>>}. No matter what data occurs in D, D', some daughter of f is a witness of the fact that $\mathfrak{T}[2]$ is not safe. So there is no complete, safe revolution tree mod G, s. Now apply Theorem 4.1. ∎

It is often easier to give a proof by nested locking sequences directly, without invoking Theorem 4.1. Such proofs also help to illustrate what is going on in the negative side of the proof of Theorem 4.1.

---

[4]System F may be thought of as a relativized, 3-dimensional generalization of problem (b) of Exercise 2.C in [15]. F was instrumental in our isolation of the characterization conditon for truth detectability.

Alternate proof of Fact 4.1.4: Suppose δ succeeds. Choose σ_1 to be locking for δ, f_1, s, where f_1 is the root of the tree. Now choose some daughter f_2 of f_1 such that σ_1 is sound for f_2, δ, s. Choose σ_2 to be locking for δ, f_2, s so that $\sigma_1 \subseteq \sigma_2$. Choose some daughter f_3 of f_2 so that σ_2 is sound for f_3, δ, s. Choose σ_3 to be locking for f_3, δ, s so that $\sigma_2 \subseteq \sigma_3$. $\sigma 3$ is sound for f_1, f_2, δ, s. Hence, $\delta(s, \sigma 3)$ must be correct for f1, f2, f3. But this is impossible, since $az_s(\{f_1, f_2, f_3\}) = \varnothing$. ■

4.1.5 When C is Infinite

The characterization given in Theorem 4.1 covers only relativistic systems in which C is finite. The results of this section establish that the characterization condition is necessary for arbitrary relativistic systems, but is not sufficient when F and C are countably infinite.

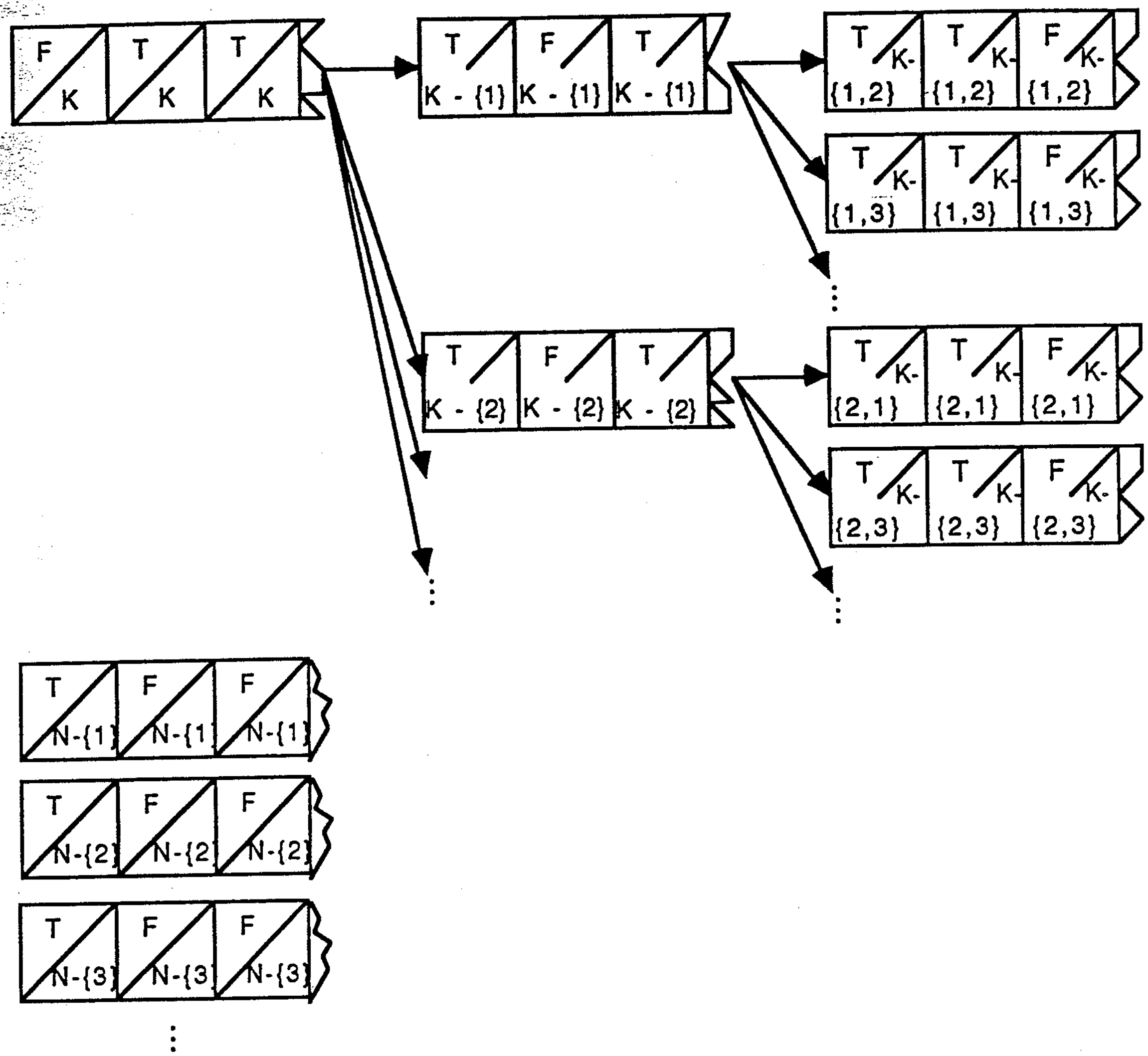
Corollary to Theorem 4.1: Let <F, C, W> be an arbitrary relativistic system, and let $s \in \Sigma^*$. Then

if s is truth detctable over F then there is a complete, safe, revolution tree mod F, s.

Proof: Nothing in the necessity side of the proof of Theorem 4.1 made use of the cardinality of F or of C. As the cardinalities of F and C increase, the branch factor of the tree increases, but its depth is still bounded by ω. ■

The following example shows that our characterization condition is not sufficient for truth detectability when C is infinite.

Let K = N - {0}. Define F to be the following problem, consisting of a tree of worlds-in-themselves together with an additional set of worlds-in-themselves.



All worlds-in-themselves are infinite matrix rows, so only initial segments can be shown. Each world-in-itself in the additional sequence has T in the first place and has F everywhere else. Moreover, for each world in the sequence, the total data is N-{i}, for some i $\in$ N. In the tree, each world occurring at level n of the tree has F only under conceptual scheme n and has total data K - D where 0 $\notin$ D and K = N - {0} and |D| = n. Now we proceed to show that there is a complete, safe revolution tree mod F, s but s is not truth detectable over F.

Fact 4.1.5.a: There is a complete, safe revolution tree mod F, s.

Proof: We will construct a tree $\mathfrak{T}$ in which all clues are empty. For each path $<f_1, f_2, ..., f_n>$ in the "tree part" of F, let $<<f_1, \emptyset>, ..., <f_n, \emptyset>>$ be a path in $\mathfrak{T}$. Each f in the "table part" of F is extended as follows: for each f' in the "tree-part" of F whose data is included in the data of f, add the path

<<f, Ø>, <f', Ø>> to $\mathfrak{T}$. It is straightforward to verify that the resulting $\mathfrak{T}$ is a complete, safe revolution tree. ■

Fact 4.1.5.b: s is not truth-detectable over F.

Proof: Suppose δ can detect s over F. Pick σ_1 to be locking for δ and for the root f_1 of the tree. Now, choose a daughter f_2 of the root of the tree for which σ_1 is sound. There is one, by construction. Pick σ_2 to extend σ1 and to be locking for f_2, δ. Pick f_3 so that σ2 is sound for f', etc. Observe that σ_i will be sound for each f_j, for $j \leq i$, by construction of the tree. Since each σ_i is locking for f_i, and is sound for each fj for $j \leq i$, we know that the conceptual schemes occurring in σ_n are in $az_s(f_1, ..., f_{n-1})$, else δ makes an incorrect conjecture for one of $f_1, ..., f_{n-1}$ after seeing locking sequences for $f_1, ..., f_{n-1}$ and after seeing no data unsound for $f_1, ..., f_{n-1}$. Notice that $az_s(f_1, ..., f_n) = \{i: i > n\}$. Hence, for each conceptual scheme i, there is a time k in reading $\sigma_1, \sigma_2, ..., \sigma_n, ...$ after which i is no longer in $az_s(f_1, ..., f_k)$ or in any further agreement zone. Hence, the data presentation t such that each σ_i is an initial segment of t is for any world-in-itself for which it is sound (recall that completeness of data demands only that all true data be presented when δ stops at a conceptual scheme infinitely often).

Now we will show that t is sound and hence complete for some world in the sequence below the tree. Suppose t is sound for no world in that sequence. Then for each n > 0, n occurs under some conceptual scheme in t. But then t is sound only for f_1, which contradicts the fact that σ_i is sound for each f_j such that $j \leq i$. So we have that t is sound for some f in the list following the tree. But observe that f disagrees with $f_2, ..., f_n, ...$ over $az_s(f_1, f_2)$. Hence on some data presentation t for f, δ makes infinitely many mistakes about f, for we have seen that once δ sees σ_2, δ must produce conjectures correct for f_2 forever after, and that these conjectures must occur over schemes in $az_s(f_1, f_2)$, but f_2 disagrees with f everywhere over this region. ■

Let us refer to the "tree part" of F as F'. F' is an infinite-dimensional generalization of the three dimensional problem G examined in the previous section. It is interesting that F' is solvable, while G is not. To solve F', one may simply conjecture <1, T>, <2, T>, ..., <n, T>, ... without ever looking at the data. The trouble in the finite problem G is that eventually, one must stop choosing new conceptual schemes, because there are only finitely many to choose from.

Problem F is unsolvable for a reason that is not captured in our characterization theorem. It forces a conflict between two strategies for relativistic induction. The first is to visit each conceptual scheme at most finitely often, saying T each time. In the tree this works fine, for after some time you walk to the right of the one scheme in which you should have said F, and you are correct thereafter. But if you are in the sequence, this strategy fails. In the sequence, you can get by if you always say "F" and stay away from conceptual scheme 1. So we know what to do if we know whether we are in the tree or in the sequence. How could we tell which we are in? If we are in the sequence, then if we wait long enough in one place, we will see a zero; something that we would never see in the tree. So it would make sense to assume we are in the tree until we see a zero. But we are only guaranteed to see a zero if we visit some conceptual scheme infinitely often and wait for the zero there. And if the actual world-in-itself is in the tree, and in fact is the one that has "F" under the scheme where we search for a zero, then we are wrong infinitely often. So if we don't stay in one place we can't tell if we are in the tree or in the sequence and if we do wait for data, then we may be wrong in the tree. In more popular language, the example shows a fundamental dilemma between "revolutionary" and "normal science".

The failure of our characterization condition to be sufficient for success when C is infinite is tied to the possibility that two worlds of inquiry under the same conceptual scheme can assign distinct truth values to s, while the complete data for one is included in the complete data for the other. This may happen because s involves unobservable vocabulary under some conceptual scheme, because the evidence language lacks a negation operator under some conceptual scheme, or because the negations of some observable strings are not observable. The following simple result shows that when these circumstances do not obtain under some conceptual scheme, no conceptual revolutions are necessary for truth detection, regardless of the cardinality of C. So under these circumstances, virtually all of the structure discussed in this paper collapses. This is the circumstance assumed, for example, in Gaifman and Snir's Theorem 2.1 concerning the convergence to the truth of conditional measures [5].

Say that s is *data-determined* mod F, c $\Leftrightarrow$

$\forall f, g \in F$, if $tv(s, f(c)) = tv(s, g(c))$ then

$ev(f(c))$ is not a subset of $ev(g(c))$

Fact 4.1.5.c: Suppose that F is countable and there is a $c \in C$ such that s is data-determined mod F. Then s is detectable over F by a method that never switches conceptual scheme.

Proof: For let $f_1, \ldots, f_n, \ldots$ be an enumeration of F. Define $\delta(s, \sigma)$ to conjecture $\langle c, tv(s, f_i(c))\rangle$, where f_i is the first world-in-itself in F for which σ is sound with respect to δ. It is easily verified that δ succeeds in the required sense. ■

The triviality of data-determined truth detection disappears when F is uncountable (c.f. [11], Proposition 12).

4.1.6 Why Simpler Conditions Don't Work

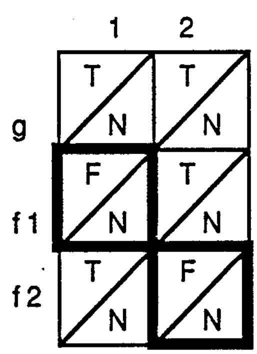
The characterization conditions provided in Theorem 4.1 may seem unduly complex. In this section, we consider two plausible alternatives, and show how they fail.

Fact 4.1.6.a: The following condition is necessary for truth detectability over arbitrary relativistic systems, but is not sufficient even when C and F are both finite:

$\forall f\ \exists$ clue D sound for f s.t.
$\forall f'$ if D is also sound for f' then
$\exists c \in C$ s.t. either
tv(s, f(c)) = tv(s, f'(c)) or
$ev(f(c)) - ev(f'(c)) \neq \varnothing$.

Proof: To show necessity, deny the condition and suppose that δ succeeds. pick a locking sequence σ for f, δ, s, and then choose f' so that σ is sound for f', δ, s. Extend σ to a presentation t for f', δ. δ fails on t. Contradiction.

To see that sufficiency fails for finite relativistic systems, consider the following problem F:



F satisfies the condition (each pair of worlds-in-themselves agrees about the truth value of s in some place), but s is not truth-detectable in F. The evidence is the same no matter what. So suppose that δ succeeds in g. Then he fails to detect s either in f1 or in f2, since δ must make infinitely many conjectures either under scheme 1 or under scheme 2. ■

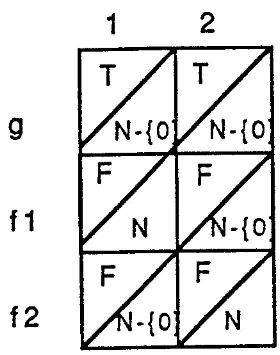
Fact 4.1.6.b: The following condition is sufficient when C is countable and F is countable but is not necessary for truth-detectability even when C is finite:

$\forall f\ \exists$clue D sound for f s.t.

$\forall f'$ if D is also sound for f' then
$\forall c \in C$ either
$tv(s, f(c)) = tv(s, f'(c))$ or
$ev(f(c)) - ev(f'(c)) \neq \emptyset$.

The method δ simply produces conjectures from an infinitely repetitive enumeration μ of $cr = \{<c, T>: c \in C\}$ until it finds a clue D for some unrefuted $f \in F$, and when it finds one, it moves to the last position visited on another infinitely repetitive enumeration $\mu(f)$ of $cr(f) = \{<c, b>: c \in C \;\&\; tv(s, f(c)) = b\}$.

To see that the condition is not necessary, consider again the problem G' in the proof of Lemma 6.2.



Let D be any clue sound for g. This clue is also sound for f_1. But under scheme 2, the total data is the same for f_1 and for g, but the truth values assigned to s by these two worlds differ. Hence the condition fails to hold. But s is truth-stably detectable over G'. ■

4.2 Truth-Stable Truth Detectability Characterized

4.2.1 Definitions

Our characterization of truth-stable detectability differs from the one given for truth detection *simpliciter* primarily in that the latter notion of success requires a stricter notion of agreement zone. In the case of truth detection *simpliciter*, the agreement zone of a collection of worlds-in-themselves is just the set of all conceptual schemes under which no two worlds in the collection assign distict truth values to s. For truth-stable truth detection, we insist, in addition, that no world-in-itself in the collection assign different truth values to s on different conceptual schemes in the agreement zone. Let $b \in \{T, F, U\}$.

$$az_{s,b}(\{f_1, \ldots, f_n\}) = \{c \in C: \forall i \text{ s.t. } 1 \leq i \leq n,\ tv(s, f_i(c)) = b\}$$

The concepts of *may extend*, *revolution tree*, and *safe* are just as before, except for the substitution of the new notion of agreement zone and for their inheritance of the agreement zone's relativization to b. It also turns out to be useful to make completeness relative to worlds-in-themselves, as follows.

$\mathfrak{T}$ *is complete* for $f \in F$ mod $F, s, b \Leftrightarrow$
    $\exists D\ \langle f, D\rangle \in \mathfrak{T}$ and
    $\forall p \in \mathfrak{T}\ \forall f' \in F$
        if p extends $\langle f, D\rangle$ in $\mathfrak{T}$ and
        f may extend p mod F, s, b
        then $\exists D'\ p^{*}\langle f, D'\rangle \in \mathfrak{T}$.

4.2.2 Characterization Theorem for Truth-Stable Detectability

Theorem 4.2: Suppose F is countable and C is finite. Then

s is truth-stably detectable over F $\Leftrightarrow$
$\forall f \in F\ \exists \mathfrak{T}\ \exists b$ s.t.
$\mathfrak{T}$ is a safe, revolution tree complete for f mod F, s, b.

Proof: Parallel to the proof of Theorem 4.1. ■

Corollary to Theorem 4.2: The left-to-right direction of Theorem 4.2 holds for arbitrary relativistic systems.

4.3 Scheme-Stable Truth Detectability Characterized

4.3.1 Definitions

This time we make the notion of agreement zone relative to conceptual scheme. Hence, the appropriate agreement zone for scheme-stable detectability is either empty, or contains exactly one conceptual scheme.

$$az_{s,c}(\langle f_1, \ldots, f_n\rangle) = \begin{cases} \{c\} \text{ if } \exists b\ \forall i\ 1 \le i \le n \Rightarrow\ tv(s, f_i(c)) = b \\ \varnothing \text{ otherwise} \end{cases}$$

The notion of safe revolution tree is modified in light of this change just as in the case of truth-stable identifiability. Just as before, we make completeness relative to f.

4.3.2 Characterization Theorem for Scheme-Stable Detectability

Theorem 4.3: Suppose F, C are countable. Then

> s is truth-stably detectable over F $\Leftrightarrow$
> $\forall f \in F\ \exists \mathfrak{T}\ \exists c$ s.t. $\mathfrak{T}$ is a safe, revolution tree complete for f mod F, s, c.

Proof: Parallel to the proof of Theorem 4.1. ■

The restrictive notion of agreement zone for scheme-stable detectability means that the paths in a revolution trees may have length no greater than 1. Because of this restriction, much of the apparatus required for the characterization of detectability *simpliciter* and truth-stable detectability collapses. This collapse is reflected in the following corollary.

Corollary 4.3.1: If F, C, are both countable then

s is scheme-stably detectable over F $\Leftrightarrow$

(A) $\forall f \in F\ \exists c \in C\ \exists$ clue D s.t.
$\{\langle\rangle, \langle\langle f, D\rangle\rangle\}$ is a safe revolution tree mod F, s, c. $\Leftrightarrow$

(B) $\forall f \in F\ \exists c \in C\ \exists$ clue D s.t.
D is sound for f and
$\forall f' \in F$ if
$tv(s, f(c)) \neq tv(s, f'(c))$ and
D is sound for f'
then $ev(f'(c)) - ev(f(c))) \neq \varnothing$

Proof: The equivalence of (A) with the characterization condition of Theorem 3 is immediate from the fact that $\{\langle\rangle, \langle\langle f, D\rangle\rangle\}$ is complete for f mod F, s, c because no path can have length greater than 1. The equivalence of (A) and (B) is straightforward, by the definitions of completeness and safety. ■

5. Separation Results

In this section, we show that the apparently weaker notions of truth-detection really are weaker. In more contentious language, we show that conceptual revolutions can make a method more reliable at getting to the truth. We take the opportunity to illustrate our characterization theorems in the following proofs, although direct proofs are often easier to provide.

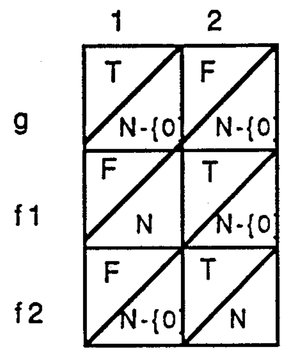
Theorem 5:
scheme-stable detectability $\Rightarrow$
truth-stable detectability $\Rightarrow$

detectability simpliciter

and none of the converses is true.

Lemma 5.1 Truth-stable detectability ⇒ detectability but not conversely.

Proof: The implication is trivial. To see that the converse fails, consider the following problem.



s is not truth-stably detectable over G, as we will show by an application of Theorem 2. Suppose $\exists \mathfrak{T}$, b such that $\mathfrak{T}$ is a safe revolution tree that is complete for g mod F, s, b. Since $\mathfrak{T}$ is complete for g, $\exists D$ <g, D> $\in \mathfrak{T}$. So long as D is sound for g, D is also sound for f_1 and for f_2. So if b = T then f_2 witnesses the violation of safety for $\mathfrak{T}$, G, s, b. And if b = F then f_1 witnesses the violation of safety for $\mathfrak{T}$, G, s, b. Finally, if b = U then either f_1 or f_2 witnesses the violation of safety for $\mathfrak{T}$, G, s, b, since $az_{s,U}(g) = \varnothing$.

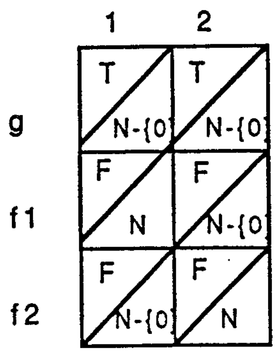
It is possible to truth-detect s in G, however. We apply Theorem 1. Set $\mathfrak{T}$ = {<>, <g, ∅>, <f_1, {<1, 0>}>, <f_2, {<2, 0>}>}. It is easy to verify by cases that $\mathfrak{T}$ is a complete, safe revolution tree mod G, s. ■

An obvious solution to the above problem is to waffle back and forth over schemes 1 and 2 making guesses correct for g until a 0 is seen in the data. If 0 occurs under 1 then you know you are in f_1, and you start producing arbitrary conjectures sound for f_1. If the 0 occurs under f_2, then you know you are in f_2 and you start making guesses sound for f_2.

Now consider what method δ in the proof of theorem 1 would do when given tree $\mathfrak{T}$. In this tree, g is paired at the root level with the empty clue, which is always found in all data. Hence, δ goes down branch <g, ∅> immediately, and stays on this branch, switching back and forth between schemes 1 and 2 until a 0 is seen. In that case, the path <g, ∅> is dropped and the next path taken depends upon where the 0 is seen. So δ duplicates the performance of the obvious method in this case.

Lemma 5.2: Scheme-stable detectability ⇒ truth-stable detectability but not conversely.

Define problem G' as follows:



S is not scheme-stably detectable over G', as we shall show by Corollary 3.1.B. For let D be any clue sound for g, and let c = 1. Then f_2 disagrees with g about s under scheme 1, but there is no data true in $f_1(1)$ but false in g(1). Let c = 2. Then the same can be said of f_1. So by Corollary 3.1.B, s is not scheme-stable detectable over G'.

By Theorem 2, we show that s is truth-stably detectable over G'. Define Set $\mathfrak{C}g$ = {<>, <g, ∅>} Set $\mathfrak{C}f_1$ = {<>, <f1, {<1, 0>}>}. Set $\mathfrak{C}f2$ = {<>, <f1, {<2, 0>}>}. It is readily verified that $\mathfrak{C}g$ is complete for g mod G', s, T, and $\mathfrak{C}f_1$ is complete for f_1 mod G', s, F, and $\mathfrak{C}f_2$ is complete for f_2 mod G', s, F. The result follows by Theorem 2. ∎

It is useful to consider what the truth-stable version of δ does on this example. Suppose g is actual. Then δ succeeds immediately, since δ considers path <g, ∅> on the basis of the empty clue. No data unsound for g will ever be read, and δ produces only conjectures correct for g forever after. Suppose f_1 is actual. Then δ again considers <g, ∅> and produces conjectures out of an infinitely repetitive enumeration of {<1, T>, <2, T>} until a 0 is read somewhere. The crucial point here is that g assigns the same truth value to s under both conceptual schemes, so that both schemes are in the agreement zone for path <g, ∅>. When the 0 is seen, δ drops path <g, ∅>. If the 0 is read under scheme 1, then the clue {<1, 0>} has been seen for path <f1, {<1, 0>}>, and δ converges to conjectures correct for f_1. The case when f_2 is actual is parallel.

6. Conclusion

In this paper we have presented a precise framework for thinking about convergence to the truth when truth, syntax and observability may depend upon what the investigator does. Within this framework, we characterized solvability according to three distinct concepts of convergence to the relative truth. To prove these results, we constructed methods that get to the relative truth by directing conceptual revolutions in which truth, syntax and observability may change. Our techniques for constructing such methods were shown to be complete, in the following sense: if a problem is solvable, then the technique generates a method that solves it. We illustrated the limitations of these results, and proved that the three notions of convergence to the relative truth are indeed distinct.

Several philosophical morals may be drawn from this work. First, we have seen that relativistic systems can model both experimenter effects and conceptual change through "scientific revolutions". In the first case, the dependency is causal, while in the second it is linguistic. Whether or not a system is relativistic depends upon what we aspire to say about it. If we want to discover all the ways in which the system responds to our acts, then we are not involved in a relativistic inquiry, for what we are trying to discover (i.e. the dependency itself) does not depend upon what we do. If, on the other hand, we intend to discover only the laws of a particular state of a system that responds to our actions, then our study is relativistic, for the laws we seek will change as the state of the system changes. So for example, a conceptual historian, who looks at past scientific episodes and tells us when and how conceptual changes occurred is not involved in a relativistic inquiry (at least insofar as these conceptual changes are concerned). On the other hand, scientists at the time of the conceptual change who were *using* the concepts that changed were, indeed, involved in a relativistic inquiry.

Second, we have seen that relativism is a more general thesis than is radical subjectivism. Relativism says only that truth depends in some way or other on the acts of the scientist; perhaps in a way that the scientist does not know *a priori*. Radical subjectivism says that whatever the scientist chooses to believe is true. Inductive inquiry is trivial in subjectivist systems, but it need not be trivial (and may, indeed, be impossible) in highly relativistic systems.

Third, Rorty, Kuhn, Feyerabend, and myriad others are in error when they argue from relativism to the impossibility of general methodological norms that hold across conceptual revolutions. The methods constructed in the proofs of our characterization theorems are demonstrably complete, and as in deductive logic, this property would seem to carry at least some normative weight.

Fourth, our analysis refutes the popular assumption that mathematical work in methodology must *somehow* fail to take relativism into account. Some philosophers and anti-philosophers identify all precise methodological work with logical positivism, and reject it for the same reasons. But logical positivism involves semantic theses that our framework is in no way committed to. These include the existence of meaning postulates, analytic truth, and analytic reduction relations between theory and evidence. The positivists did not propose scientific methods that work across conceptual revolutions, and we do. In short, we hope to have shown that relativism is no excuse for obscurantism in matters methodological. In fact, our results show that getting to the truth takes on a much richer mathematical structure when truth is relative.

This paper is just a small first step into the logic of relativistic inquiry. Many important questions remain open about the present system. For example, we would like to obtain a characterization theorem for truth detectability when C is countably infinite and F is uncountable. We would also like to obtain some general results concerning the scope of computationally bounded, relativistic truth detectors. Results in standard learning theory provide a blue-print for such a project.

Imagine the process of building up a complete, true theory of the world by adding new truths and deleting falsehoods, so that each truth is eventually added and each falsehood is eventually withdrawn forever. In realist settings, the existence of a truth detector for each hypothesis is equivalent to the existence of a theory discovery device in the sense just given [15, Proposition 80]. But in relativistic systems, this equivalence fails, because the truth detectors need not agree about conceptual scheme, so the truth values they return cannot be relied upon jointly. The situation becomes still more complicated when truth depends upon the theory currently conjectured.

There are many sorts of relativism that the framework of this paper does not address. For example the world of inquiry may depend upon the conjectured truth value as well as the chosen conceptual scheme. This kind of relativism can be handled by a slight modification of our techniques.

It is more difficult to handle cases in which the world of inquiry depends upon the whole history of inquiry, rather than simply upon some contemporary choice of conceptual scheme on the part of the scientist. This is the sort of view suggested by Marxists who insist upon "taking history seriously". The formal effect of this proposal is to make it structurally difficult for a scientist to visit a world of inquiry at will, since his past history may prevent him from returning to it.

A third possibility is to relativize the history of inquiry as well as the subject matter under investigation. In our results, this history is held to be objective. No matter what conceptual scheme a scientist chooses, the truth about the conceptual schemes visited and data strings received in the past remains fixed. But if this history also changes with changes in conceptual scheme, the scientist's convergence to the relative truth is itself relative. A philosophically interesting proposal is to combine relativity *of* history with truth dependence *on* history. In this system, the current history determines the world of inquiry from which the next datum is received. The conceptual scheme chosen on the basis of adding the next datum to the current history may then radically alter the current history. The subsequent datum is chosen with respect to the altered history. Hence, there is no vicious circularity, but there is a much more free-wheeling form of relativism than the sort studied in this paper.

It would be a mistake to infer that these added sources of relativism re-open a Pandora's box of obscurity in methodology. Now that it has been shown how to study convergence to the truth in a simple sort of relativistic framework, it becomes clear that similar, albeit more sophisticated techniques will suffice in the study of additional sources of relativism.

Acknowledgments

Kevin Kelly was supported during the Spring and Summer of 1989 by a grant from Danish Research Academy, issued through the cognitive science group at Risø National Laboratory and Roskilde University Center, Denmark. In particular, he appreciates many lively discussions on the topic with Stig Andur Pedersen, Arne Thing Mortensen, and Henning B. Andersen. The core ideas of this paper were presented in talks at Roskilde University Center, the University of Aalborg, Denmark, and the University of Konstanz, West Germany. Comments at these lectures were duly noted. Special thanks are due to the Danish state medical system for patching Kelly's broken knee free of charge.

Clark Glymour was supported by the John Simon Guggenheim Foundation. Together, we would like to thank Cory Juhl for discussions at the beginning of the project, and Cristina Bicchieri for useful suggestions near its conclusion, including the example of the stock market predictor. Teddy Seidenfeld provided useful comments.

Appendix

Proof of Lemma 3.2: Let δ detect s in f. Now let γ be sound for f. Suppose that no σ such that $\gamma \subseteq \sigma$ is locking for δ, f, s. So we have

(*) $\forall \sigma$
if $\gamma \subseteq \sigma$ and
σ is sound for δ, f, s and
$tv(s, f(\delta(s, \sigma)_1)) = \delta(s, \sigma)_2$ then
$\exists \tau \in SEQ$
  (a) $\sigma \subseteq \tau$ and
  (b) τ is sound for δ, f, s and
  (c) $\delta(s, \tau)_2 = tv(s, f(\delta(s, \tau)_1))$

We construct a data presentation t complete and sound for f, δ, on which δ makes infinitely many mistakes, which is a contradiction. We construct t in alternating stages. In even-numbered, or "fooling" stages, we add a τ of the sort guaranteed by (*) above to force δ to make a mistake. In odd-numbered, or "data completion" stages, we add a new datum, if possible, to each conceptual scheme visited by δ when we added the previous τ and in the previous odd stage. We ensure soundness for f, δ by adding only chunks of data that are sound for f, δ. We ensure completeness for f, δ because any conceptual scheme visited infinitely often by δ will have complete data presented during odd-numbered stages. Now we make the construction precise:

Stage 1: $t\{1\} = \zeta$, where

  (i) ζ is sound for δ, f, s and
  (ii) $\gamma \subseteq \zeta$ and
  (iii) $\delta(s, \zeta)_2 = tv(s, f(\delta(s, \zeta)_1))$.

There is such a ζ, for let t be a data presentation sound and complete for δ, f such that $\gamma \subseteq t$. If δ never produces a correct response on t after seeing γ, then δ fails to detect s in f, which is a contradiction. So eventually, $\exists m$ $\delta(s, t\{m\})_2 = tv(s, f(\delta(s, t\{m\})_2))$ and $\gamma \subseteq t\{m\}$. Let $\zeta = t\{m\}$.

Stage 2n (fooling stage):
If
  (1) $t\{2n-1\}$ is sound for δ, f and
  (2) $\gamma \subseteq t\{2n-1\}$ and
  (3) $\delta(s, \sigma)_2 = tv(s, f(\delta(s, \sigma)_1))$

then we are free to choose τ according to (*) with the following properties:

  (a) $t\{2n-1\} \subseteq \tau$ and
  (b) τ is sound for δ, f, s and
  (c) $\delta(s, \tau)_2 \neq tv(s, f(\delta(s, \tau)_1))$.

Otherwise let τ = some arbitrary default.

Define $t\{2n\} = \tau$.

Stage 2n+1 (data completion stage). Define

$$J_{2n+1} = \{c \in C: \exists\kappa \text{ such that } t\{2n\text{-}2\} \subset \kappa \subseteq t\{2n\} \text{ and } \delta(s, \kappa)_1 = c\}.$$

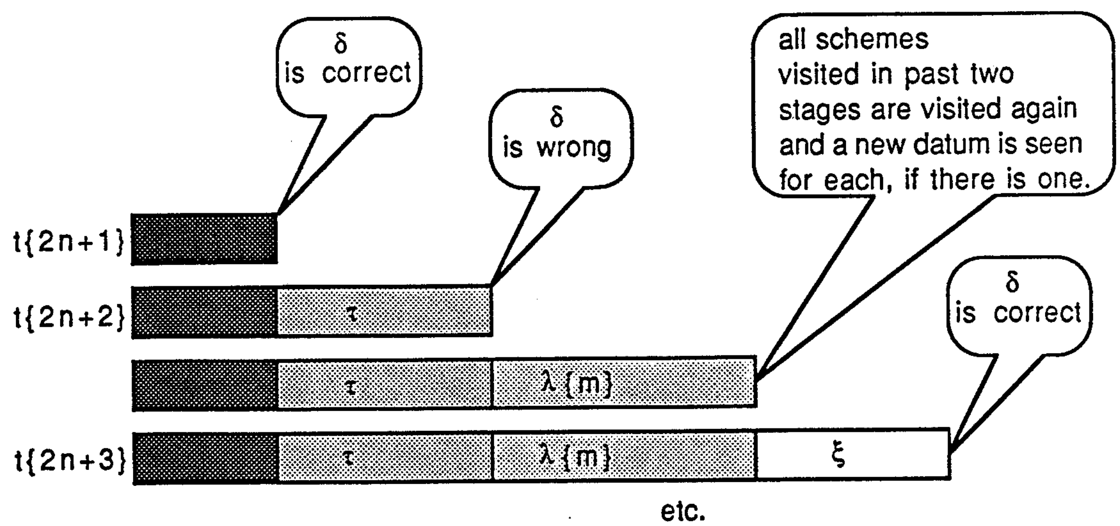
Think of J_{2n+1} as the set of all conceptual schemes visited by δ during the previous cycle of fooling and completion. J_{2n+1} is finite. Let $m = |J_{2n+1}|$. Choose some fixed enumeration $c_1, c_2, c_3, \ldots, c_m$ of J_{2n+1}, and some fixed enumeration $e_1, e_2, \ldots, e_n, \ldots$ of $ev(f(c_i))$, for each c_i. Set $\lambda\{0\} = t\{2n\}$. $\forall i$ s.t. $0 < i < m$, Define $\lambda\{i+1\}$ to be some choice of ε such that

(i) $\lambda\{i\} \subseteq \varepsilon$ and
(ii) ε is sound for δ and f and
(iii) $\delta(s, decr(\varepsilon)) = c_i$, and
(iv) the last entry in ε is the first $e \in ev(f(c_i)) - rng(decr(\varepsilon)[\delta, c_i, s])$ if there is one, and is some arbitrary $e \in ev(f(c_i))$ otherwise.

if there is such an ε, and $\lambda\{i+1\} = \lambda\{i\}$ otherwise.

Now find some finite data ξ sound for δ, f, s, such that $\lambda\{m\} \subseteq \xi$ and $\delta(s, \xi)_2 = tv(s, \delta(s, \xi)_1)$. There must be one. For extend $\lambda\{m\}$ to a presentation t' for δ, f, s. There is a first time r at which $\delta(s, t'[r])_2 = tv(s, \delta(s,t'[r])_1)$, else δ fails to detect s on some presentation t for f, which contradicts the Lemma's assumption. Let $\xi = t[r]$.

Define $t\{2n+1\} = \xi$.



By a straightforward inductive argument, we have that

(A) $\forall n \geq 0$, $t\{2n+1\}$ satisfies conditions

(1) t{2n+1} is sound for δ, f and
(2) $\gamma \subseteq$ t{2n+1} and
(3) tv(s, f(δ(s, t{2n+1})$_1$)) = δ(s, t{2n+1})$_2$.

It follows from (A) and from the construction that

(B) ∀n ≥ 1, t{2n} satisfies conditions

(b) t{2n} is sound for δ, f, s and
(c) δ(s, t{2n})$_2$ ≠ tv(s, δ(s, t{2n})$_1$)

Now define t = the unique data presentation such that for each n, t{n} ⊆ t. (A) and (B) show that t is sound for δ, f, s and δ makes infinitely many errors on t (because δ makes an error after reading each even stage of t).

It remains only to show that t is complete for δ, f, s. That is, we need to show that if there are infinitely many j such that δ(s, t[j])$_1$ = c, then ev(f(c)) ⊆ rng(t[δ, c, s]).

Accordingly, suppose that there are infinitely many j such that δ(s, t[j])$_1$ = c. Suppose further that e ∈ ev(f(c)). Choose k so that δ(s, t[k])$_1$ = c. There is an n such that t{n-1} ⊆ t[k] ⊆ t{n}. Let t{m} be the first odd ("data completion") stage extending t{n}. Let c = c_i in the enumeration of J_m. Now consider the formation of λ[i] during stage m. Recall, there are infinitely many j such that δ(s, t[j])$_1$ = c_i. So ∃j' ∃e' ∈ ev(f(c)) such that δ(s, t[j'])1 = c_i and t[j']*e' = λ[i] ⊆ t{m}, by the definition of λ[i]. If e ∈ t[j'][δ, c, s] then we are done. So assume that e ∉ t[j'][δ, c, s]. Then e' ∉ t[j'][δ, c, s], by the definition of λ[i].

We can repeat this argument until we reach an odd stage m' such that each e" prior to e in ev(f(c)) is in rng(t{m'}[δ, c, s]). Either e ∈ rng(t{m'}[δ, c, s]) already or e ∈ rng(t{m'+2}[δ, c, s]). ■

Proof of Lemma A: By induction on length of p.

Suppose the length of p is 0. Then (1) az_s(worlds(p)) = C so cs(D) ⊆ C and (2) D = lock(<>, f) is sound for f.

Suppose the lemma for each path of length m or shorter. Suppose p = <<f_1, D_1>, ..., <f_{m+1}, D_{m+1}>> is a path of length m+1 and p*<f, D> ∈ $\mathfrak{T}$[n]. For each position i in p, D_i = $clue_\delta$(lock(p[i-1], f_i)). Let σ_i = lock(p[i-1], f_i). Applying the induction hypothesis, σ_{m+1} is sound for each f_i ∈ worlds(p), since D_{m+1} = $clue_\delta(\sigma_{m+1})$. Hence, δ(s, σ_{m+1}) is a conjecture correct for each f_i in p, by the definition of locking sequence and the fact that for each position i in p, σ_i is locking for f_i. But δ(s, σ_{m+1}) can be correct for each f_i ∈ worlds(p) only if δ(s, σ_{m+1})$_1$ ∈ az_s(p), by the definition of agreement zone. Hence, (*) until δ sees data unsound for some f_i ∈ worlds(p), δ produces conjectures involving only conceptual schemes in az_s(worlds(p)).

Since $p^*\langle f, D\rangle \in \mathfrak{T}[n]$, f may extend p mod F, s (by the definition of $\mathfrak{T}[n]$). By clause (2) of the definition of "may extend", $\forall c \in az_s(worlds(p))$, $ev(f(c)) \subseteq ev(lastworld(p)(c))$. Indeed, since this relation is maintained throughout the construction of p, we have by transitivity of set inclusion that (**) for each $f_i \in worlds(p)$, $\forall c \in az_s(worlds(p))$, $ev(f(c)) \subseteq ev(f_i(c))$. Hence, data unsound for some $f_i \in worlds(p)$ cannot be read from f under schemes in $az_s(worlds(p))$. So by (*), δ never examines conceptual schemes outside of $az_s(worlds(p))$ after reading σ_{m+1}. Since lock(p, f) exists by the locking sequence lemma and the fact that δ detects s over F, and since δ can only examine schemes in $az_s(worlds(p))$ after reading σ_{m+1}, we have that (1) $cs(clue_\delta(lock(p, f))) \subseteq az_s(worlds(p))$. By (1) and (**) we have (2) $clue_\delta(lock(p, f))$ is sound for each $f \in worlds(p)$. ■

Proof of lemma B: Let $f \in F$. Let t be a complete, sound data presentation for δ, f.

Say that path $\langle f_1, D_1\rangle, \ldots, \langle f_n, D_n\rangle$ *extends* path $\langle g_1, H_1\rangle, \ldots, \langle g_m, H_m\rangle$ in $\mathfrak{T}$ $\Leftrightarrow n \geq m$ and $\forall 1 \leq i \leq m$, $f_i = g_i$ and $D_i = H_i$.

Lemma 1: $\forall$ path $p \in \mathfrak{T}$, $\forall f \in F$,
if
 (1) clues(p) is sound for f and
 (2) $\forall D \in$ Clue, $p^*\langle f, D\rangle \notin \mathfrak{T}$,
then either
 (a) $\exists c \in az_s(worlds(p))$ s.t. $ev(f(c)) - ev(lastworld(p)(c)) \neq \varnothing$
 or
 (b) $az_s(worlds(p)) - az_s(worlds(p), f) = \varnothing$.

Proof: Let $p \in \mathfrak{T}$. Suppose $\forall D \in$ Clue, $p^*\langle f, D\rangle \notin \mathfrak{T}$. Then since $\mathfrak{T}$ is complete, we have that f may not extend p mod F, s. So we have that either

(i) clues(p) is not sound for f or
(ii) $\exists c \in az_s(worlds(p))$ s.t. $ev(f(c)) - ev(lastworld(p)(c)) \neq \varnothing$ or
(iii) $az_s(worlds(p)) - az_s(worlds(p), f) = \varnothing$ or
(iv) $azs(worlds(p), f) = \varnothing$

(i) contradicts assumption (1). (ii) is just (a). (iii) is just (b). Suppose for reductio that (ii) and (iii) are false. Then since (i) is false by assumption, we may deduce (iv). But by ¬(i), ¬(ii), ¬(iii) and (iv), f bears witness that $\mathfrak{T}$ is not safe, which contradicts the theorem's hypothesis. ■

Lemma 2: $\forall$ path $p \in \mathfrak{T}$, $\forall f \in F$, either

(α) $\forall$ path p' extending p in $\mathfrak{T}$ either
 (1) clues(p') is not sound for f or
 (2) $\exists k$ $length(p) \leq k \leq length(p')$
  $\exists c \in az_s(worlds(p'[k]))$
   $ev(f(c)) - ev(lastworld(p'[k](c)) \neq \varnothing$.

or

(β) $\exists$ path p' extending p in $\mathfrak{T}$ such that

(1) clues(p') is sound for f and

(2) $\forall k$ s.t. length(p) $\leq k \leq$ length(p'),

$\forall c \in az_s(p'[k])$,

$$ev(f(c)) - ev(lastworld(p'[k](c)) = \varnothing \text{ and}$$

(3) $\forall c \in az_s(worlds(p'))$, $tv(s, f(c)) = tv(s, lastworld(p')(c))$.

Proof: Suppose $\neg(\alpha)$. So $\exists$ path p' extending p in $\mathfrak{T}$ such that clues(p') is sound for f and $\forall k$ s.t. length(p) $\leq k \leq$ length(p'), $\forall c \in az_s(worlds(p'[k]))$, $ev(f(c)) - ev(lastworld(p')(c)) = \varnothing$.

Case (i): Suppose that $\exists D\ \exists k$ s.t. length(p) $\leq k \leq$ length(p') s.t. $p'[k]^{*}<f, D>$ is a path in $\mathfrak{T}$. Path $p'[k]^{*}<f, D>$ witnesses the truth of condition (β.1), since clues(p'[k])is sound for f, and since D is sound for f by condition (3) of the definition of revolution tree, together with the fact that $p'[k]^{*}<f, D> \in \mathfrak{T}$ and $\mathfrak{T}$ is a revolution tree. Path $p'[k]^{*}<f, D>$ also witnesses the truth of condition (β. 2), since p'[k] has property (β. 2) and adding $<f, D>$ to this path cannot violate property (β. 2). Finally, the path $p'[k]^{*}<f, D>$ trivially witnesses the truth of (β. 3).

Case (ii): Suppose that $\forall D\ \forall k$ if length(p) $\leq k \leq$ length(p') then $p'[k]^{*}<f, D>$ is not a path in $\mathfrak{T}$. Then in particular, (*) $\forall D\ p'^{*}<f, D>$ is not a path in $\mathfrak{T}$. So by Lemma 1 and the fact that clues(p') is sound for f, we may conclude that (a) $\exists c \in az_s(worlds(p'))$ s.t. $ev(f(c)) - ev(lastworld(p')(c)) \neq \varnothing$ or (b) $\forall c \in az_s(worlds(p'))$ $tv(s, f(c)) = tv(s, lastworld(p')(c))$. But by $\neg(\alpha)$, (a) is false. Hence (b) obtains, which establishes (β. 3) for path p'. Path p' satisfies (β. 2) by the assumption of $\neg(\alpha)$. And p' satisfies (β. 1) because clues(p') are sound for f. ■

Lemma 3: Let path $p \in \mathfrak{T}$, and let $f \in F$.

If $\forall c \in az_s(worlds(p))$, $tv(s, f(c)) = tv(s, lastworld(p)(c))$
then for any extension p' of p in $\mathfrak{T}$, $\forall c \in az_s(worlds(p'))$,

$$tv(s, f(c)) = tv(s, lastworld(p')(c)).$$

(I.e. the property of agreement with f over the agreement zone of one's ancestors is closed downward in $\mathfrak{T}$).

Proof: Let p' extend p in $\mathfrak{T}$. Then $az_s(worlds(p')) \subseteq az_s(worlds(p))$ and $\forall c \in az_s(worlds(p'))$, $tv(s, lastworld(p)(c)) = tv(s, lastworld(p')(c))$, by the definition of agreement zone. So by the lemma's hypothesis, $\forall c \in az_s(worlds(p'))$, $tv(s, f(c)) = tv(s, lastworld(p')(c))$. ■

Say that δ *considers* path $<f_1, D_1>, ..., <f_n, D_n>$ at stage m $\Leftrightarrow$ $<f_1, ..., f_n>$ is an initial segment of $\Omega[m]$ and $<D_1, ..., D_n>$ is an initial segment of $\Xi[m]$.

Say that δ *considers exactly* path $<f_1, D_1>, ..., <f_n, D_n>$ at stage m $\Leftrightarrow$ $<f_1, ..., f_n> = \Omega[m]$ and $<D_1, ..., D_n> = \Xi[m]$.

Lemma 4: If data unsound for lastworld(p) is read in t, then p is never again considered by δ on presentation t.

Proof: Suppose that data unsound for lastworld(p) is read by δ. Then by clause (b. iii) of the definition of δ, δ never again considers any path p' such that lastworld(p') = lastworld(p). ■

Lemma 5: Let $t \in$ PRES(f, δ, s). Then

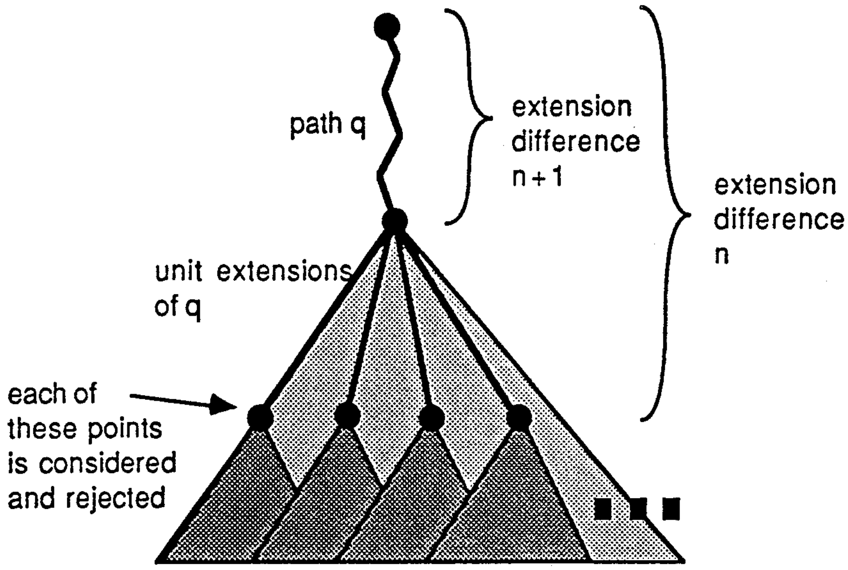
$\forall$ path $p \in \mathfrak{T}$,
if
(α) $\forall$ path p' extending p in $\mathfrak{T}$
(1) clues(p') is not sound for f or
(2) $\exists k$ s.t. length(p) $\leq k \leq$ length(p')
$\exists c \in az_s(p'[k])$
$ev(f(c))$ - $ev(lastworld(p)(c)) \neq \varnothing$ and
(γ) $\exists m$ s.t. δ considers p at some stage m
then
(1) there is data unsound for lastworld(p) in data presentation t with respect to δ and
(2) after this data is read by δ, δ never again considers p on presentation t.

Proof: Assume the Lemma's antecedent. By Lemma 4, it suffices to establish (1). Define the *extension difference* of a path q in $\mathfrak{T}$ to be the difference in length between q and the longest extension of q in $\mathfrak{T}$. The extension difference is well-defined, since no path in $\mathfrak{T}$ is longer than |C|. We establish (1) by induction on the extension difference of p .

Base case: If the extension difference of p in $\mathfrak{T}$ is 0, then there is no extension of p in $\mathfrak{T}$. So once δ considers path p, condition (b) of the definition of δ will never again be satisfied until data unsound for lastworld(p) w.r.t. δ is read from t (i.e. until condition (a) of the definition of δ occurs). So once δ considers p then δ sticks with conjectures in $cr(p) = \{<c, b>: c \in az_s(worlds(p))$ and $b = tv(s, lastworld(p)(c))\}$ until data unsound for lastworld(p) is read. Since p is considered by δ at some stage (by (γ) of the Lemma's antecedent), we know that clues(p) is sound for f, else condition (b) of the definition of δ would never be satisfied as δ reads t, and p would never be considered, contrary to assumption. Since p is an extension of itself, we may use the fact that clues(p) is sound for f, together with condition (α) of the lemma's antecedent, to infer that $\exists c \in az_s(worlds(p))$ s.t. $ev(f(c))$ - $ev(lastworld(p)(c)) \neq \varnothing$. This data will be found by δ as pointer $\pi_p[n]$ is incremented forever in $\mu(p)$ of cr(p).

Induction: Now suppose the lemma for each path p in $\mathfrak{T}$ whose extension difference is $\leq$ n. Suppose also that the extension difference of path q in $\mathfrak{T}$ is n+1, and q satisfies conditions (α) and (γ) of the lemma's antecedent. Consider an arbitrary unit extension $q' = q^*<g, H>$, of q in $\mathfrak{T}$. Path q' has extension difference n in $\mathfrak{T}$. Condition (α) of the antecedent of the induction hypothesis is satisfied by q', because the paths extending q' in $\mathfrak{T}$ are a subset of the paths extending q in $\mathfrak{T}$. So the antecedent of the induction hypothesis applies to each unit extension q' of q that is eventually considered by δ (and hence that also satisfies condition (γ)). So we may apply the induction hypothesis to conclude that for each such extension q'= q*<g, H> of q considered by δ, δ eventually sees data unsound for g.

Suppose for reductio that infinitely many distinct unit extensions q' = q*<g, H> of q are eventually considered by δ. Then all the data for lastworld(q) over the schemes az_s(worlds(q)) is seen because by the definition of δ, each time n a unit extension q' of q is refuted, pointer $\pi_q[n]$ is incremented by one and δ conjectures the pair in position $\pi_q[n]$ of fixed, infinitely repetitive enumeration μ(q) of cr(q) = {<c, b>: c ∈ az_s(worlds(q)), b = tv(s, lastworld(q)(c))} (see the diagram).



as each point is rejected, more data from the agreement zone of q is read.

The complete data for f over schemes in az_s(worlds(q)) includes data unsound for lastworld(q), by the same argument given in the base case. But once data refuting lastworld(q) is read, we have by Lemma 4 that no path involving lastworld(q) is ever again considered. Hence, only finitely many distinct unit extensions q' of q are considered by δ, which is a contradiction. So we may conclude that only finitely many distinct unit extensions q' of q are considered by δ in reading t. Eventually all these finitely many considered extensions are refuted and δ considers *exactly* path q thereafter, until data unsound for lastworld(q) is seen. That data unsound for lastworld(q) will appear in t under conceptual schemes in az_s(q) is guaranteed by condition (1) of the lemma's antecedent and the fact that each path is an extension of itself. So data unsound for lastworld(q) is eventually seen, and q is never again considered after this point, in accordance with the definition of δ. ■

Lemma 6: Let t ∈ PRES(δ, f, s).

if ∃ a path p in $\mathfrak{T}$ such that,

(a) δ considers p all but finitely often on presentation t for f and

(b) p satisfies condition (β) of Lemma 2

then either

(1) $\forall c \in az_S(\text{worlds}(p))$, tv(s, f(c)) = tv(s, lastworld(p)(c)) or
(2) ∃ unit extension p*<g, H> of p in 𝕿 that satisfies property (β) of Lemma 2 and that is considered all but finitely often by δ on t.

Proof: First we must show

Lemma 6.1: If p satisfies (a), (b) and ¬(1), then ∃D such that p*<f, D> extends p in 𝕿.

Suppose p satisfies (a), (b) and ¬(1). Since 𝕿 is complete mod F, s, if f may extend p mod F, s, then ∃D p*<f, D> ∈ 𝕿. So it suffices to show that f may extend p mod F, s. First, (1) clues(p) is sound for f, else p is never considered by δ on t for f (by clause (b) in the definition of δ). Second, we have that $\forall c \in az_S(\text{worlds}(p))$, $ev(f(c)) \subseteq ev(\text{lastworld}(p)(c))$ by condition (β.2) of Lemma 2 and by the fact that p is an extension of itself. Finally, we have by assumption ¬(1) that (3) $az_S(\text{worlds}(p))$ - $az_S(\text{worlds}(p), f) \neq \varnothing$. ∎

Lemma 6.2: δ considers path p infinitely often on t ⇔ δ considers p all but finitely often

Proof: ⇐ Trivial. ⇒ In the definition of δ, a path p is dropped from consideration by δ only if data unsound for lastworld(p), δ, is read (condition (a)). But a path p is considered only if no data unsound for lastworld(p) has been read (condition (b)). So once p is considered and dropped from consideration, it is never considered again. ∎

Proof of Lemma 6 continued: Suppose there is a path p in 𝕿 such that (a) δ considers p all but finitely often on presentation t for f, δ, s, and (b) p satisfies (β). Suppose further that p does not satisfy condition (1). We will establish that p satisfies (2).

Suppose for reductio that no unit extension p*<g, H> of p is considered infinitely often on t. Then either at most finitely unit extensions of p are considered and rejected by δ on t, s, or infinitely many unit extensions of p are eventually considered and rejected by p on t, s. In the first case, path p (and no proper extension of p) is visited each time one of the infinitely many considered unit extensions is dropped from consideration. In the second case, path p (and no proper extension of p) is considered all but finitely often since after some time no more extensions are considered. Either way, exactly path p is considered infinitely often by δ on t, s. So δ produces conjectures that visit each conceptual scheme in $az_S(\text{worlds}(p))$ infinitely often (by the definition of δ, δ produces the next conjecture in an infinitely repetitive enumeration μ(p) of the conjectures correct for lastworld(p) involving conceptual schemes in $az_S(\text{worlds}(p))$ each time exactly path p is considered).

But by Lemma 6.1, ∃D such that p*<f, D> extends p in $\mathfrak{T}$. Hence, cs(D) ⊆ az_s(worlds(p)), by condition (2) in the definition of revolution tree. Hence, by some time n, D is contained in t[n] mod δ, s. Since there are at most finitely many unit extensions of p preceding p*<f, D> in $\mathfrak{T}$, and since all of these are eventually rejected by δ according to our case hypothesis, it follows that eventually p*<f, D> is considered by δ on t, s. But p*<f, D> cannot be rejected by δ on f, s, for p is never rejected (by hypothesis) and no data unsound for f can occur in t with respect to δ, s, because t ∈ PRES(f, δ, s). Contradiction.

So some unit extension p*<g, H> of p is considered infinitely often by δ on t, s. Then p*<g, H> does not satisfy (α) of Lemma 2, by Lemma 5. Hence, by Lemma 2, p*<g, H> satisfies (β) of Lemma 2. By Lemma 6.2, p*<g, H> is considered all but finitely often by δ on t, s. ■

Note that the unit extension p*<g, H> of p considered all but finitely often does not have to be p*<f, D>. It may be some unit extension p*<g, H>, where g differs from f both in evidence and in its truth assignments to s. Indeed, δ may make many errors with respect to f while considering p*<g, H>. But since p*<g, H> satisfies condition (β) of Lemma 2, we know that δ can continue to extend p*<g, H>, and thus to narrow the agreement zone until conjectures correct for g are also correct for f. That this in fact happens is the point of the next lemma.

Lemma 7: ∃ a path p ∈ $\mathfrak{T}$ such that

(a) δ considers p all but finitely often on presentation t for f and

(b) ∀c ∈ az_s(worlds(p)), tv(s, f(c)) = tv(s, lastworld(p)(c))

Proof: We construct p inductively, in stages.

Stage 0: First <> satisfies property (β) of Lemma 2, since <> is extended by some path <f, D> (by completeness of $\mathfrak{T}$ and the fact that f may extend <> mod F, s). Second, <> is always considered by δ. Set p[0] = <>.

Stage *n+1:* Now suppose that p[n] that has property (β) and is considered all but finitely often by δ. By Lemma 6, either (1) ∀c ∈ az_s(p[n]) tv(s, f(c)) = tv(s, lastworld(p[n]) or (2) ∃ extension p[n]*<g, H> of p[n] in $\mathfrak{T}$ that satisfies property (β) and that is considered all but finitely often by δ on t. In case (1), we are done and p[n] is the desired path. In case 2, set p[n+1] = p[n]*<g, H>, which satisfies (β) and is considered all but finitely often by δ on t, as promised by case (2).

We continue to build up p until at some stage n, case (2) is no longer satisfied. There is such an n, for $\mathfrak{T}$ is uniformly bounded in depth by |C|. By Lemma 6, we know that (1) must be satisfied by stage n. Then p[n] is the desired path. ■

Lemma B is an immediate consequence of Lemma 7 and Lemma 3. Lemma 7 says that δ considers a path p leading to conjectures correct for f all but finitely often on t, s. Lemma 3 says that it doesn't matter what extensions of p are considered by δ, since they also lead to conjectures correct for f. ■

References

[1] Angluin, D., "Inductive Inference of Formal Languages from Positive Data", *Information and Control* 45 (1980), pp. 117-135.

[2] Angluin, D. and C. Smith, "A Survey of Inductive Inference Methods", Technical Report 250, Yale University, October (1982).

[3] Blumer, A., A. Ehrenfeucht, D. Haussler and M. Warmuth, *Learnability and the Vapnik-Chervonenkis Dimension.* Research Report UCSC-CRL-87-20, Dept. of Computer Science, University of California, Santa Cruz (1987).

[4] Feyerabend, P. K., "How to be a Good Empiricist" in *Challenges to Empiricism*, Harold Morick, ed. (1980), pp. 164-193.

[5] Gaifman, H. and M. Snir, "Probabilities over Rich Languages, Testing and Randomness", *Journal of Symbolic Logic*, 47: 3 (1982).

[6] Gold, E. M. , "Limiting Recursion", *Journal of Symbolic Logic* 30 (1965) pp. 27-48.

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

[8] Goodman, N., *Ways of Worldmaking.* Indianapolis: Hackett (1978).

[9] Halmos, P. R., *Measure Theory*, New York: Springer Verlag, (1970).

[10] Hanson, N. R., *Patterns of Discovery*, Cambridge: Cambridge University Press (1965).

[11] Kelly, K. and C. Glymour, "Convergence to the Truth and Nothing but the Truth", *Philosophy of Science* 56:2 (1989).

[12] Kelly, K. and C. Glymour, "Theory Discovery from Data with Mixed Quantifiers", *Journal of Philosophical Logic*, Forthcoming.

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

[14] Laudan, L., "Why was the Logic of Discovery Abandoned?", in *Scientific Discovery, Logic, and Rationality*, T. Nickles, (ed) (1980).

[15] Osherson, D., and S. Weinstein, "Paradigms of Truth Detection", *Journal of Philosophical Logic*, 18 (1989), pp. 1-42.

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

[17] Putnam, H. "'Degree of Confirmation' and Inductive Logic", in *The Philosophy of Rudolph Carnap*, A. Schilpp (ed), Lasalle, Illinois: Open Court (1963).

[18] Putnam, H. "Trial and Error Predicates and a Solution to a Problem of Mostowski". *Journal of Symbolic Logic*, 30: 1 (1965). pp. 49-57.

[19] Quine, W. V. O., "Two Dogmas of Empiricism", in *Challenges to Empiricism*, Harold Morick, (ed) (1980).

[20] Rorty, R., *Philosophy and the Mirror of Nature*, Princeton: Princeton University Press (1979).

[21] Sellars, W. "Empiricism and the Philosophy of Mind", in *Challenges to Empiricism*, Harold Morick, (ed) (1980), pp. 95-107.

[22] Valiant, L. "A theory of the Learnable", *Communications of the Association for Computing Machinery*, vol. 27, (1984) pp. 1134-1142.