Reliability, Realism, and Relativism

by

Kevin T. Kelly, Cory Juhl, and Clark Glymour

May 1992

Report CMU-PHIL-28

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

Reliability, Realism and Relativism

Kevin T. Kelly
Department of Philosophy
Carnegie Mellon University

Cory Juhl
Department of History and Philosophy of Science
University of Pittsburgh

Clark Glymour
Department of Philosophy
Carnegie Mellon University
Department of History and Philosophy of Science
University of Pittsburgh

Three Putnamian Theses

At one time or another Hilary Putnam has promoted each of the following theses:

> (1) *Limiting reiiabiiism:* A scientific method is better insofar as it is guaranteed to arrive eventually at the truth in more possible circumstances.
>
> (2) *Truth as ideaiized justification:* Truth is idealized rational acceptability.
>
> (3) ***Moderate*** *reiativism:* Truth is dependent, in part, upon the concepts, belief system, etc. of agent x.

Thesis (1) appears in two papers published in 1963.[1] Theses (2) and (3) appear in later works under the joint rubric of *internal realism.*[2] Putnam has not explained how the semantic theses that constitute internal realism fit together with his earlier, reiiabilist conception of scientific inquiry. Nor have his followers. Barrels of philosophical ink have been spilled on (3) in complete isolation from (1). Computer scientists, on the other hand, have furthered the study of (1) over the past three decades with no consideration of (2) or (3). So there remains an interesting and

[1] [Putnam 63] and [Putnam 63a].

[2] E.g., [Putnam 90].

obvious question. Can the conception of method characteristic of Putnam's earlier work be squared in a precise and fruitful way with his later semantic views? In this paper, we undertake to answer this question.

In Section I, we discuss Putnam's early methodological work in the context of thesis (1). In Section II, we adapt the techniques discussed in Section I to the analysis of the notion of idealized rational acceptability involved in thesis (2). Finally, we show in Section III how to extend the limiting reliabilist standards discussed in Section I to settings in which evidence and truth can both depend upon the scientists conceptual scheme and beliefs.

I. **Limiting Reliability**

Putnam's concern with the limiting reliability of scientific method is evident in his critique of Carnap's inductive logic[3], a critique informed by Kemeny's reflections on the role of simplicity in inductive inference.[4]

> I shall argue that one can show that no definition of degree of confirmation can be adequate or can attain what any reasonably good inductive judge might attain without using such a concept. To do this it will be necessary (a) to state precisely the condition of adequacy that will be in question; (b) to show that no inductive method based on a 'measure function' can satisfy it; and (c) to show that some methods (which can be precisely stated) can satisfy it.[5]

We will fill in points (a) (b) and (c) in order to illustrate the role played by limiting reliability.

I. A. Reliable Extrapolation in the Limit

Consider the game of guessing the next item in a sequence of zeros and ones[6]. When shown the sequence (0, 0, 0) one might guess that the next entry will be 0. Of course, the data might

[3][Putnam 63].

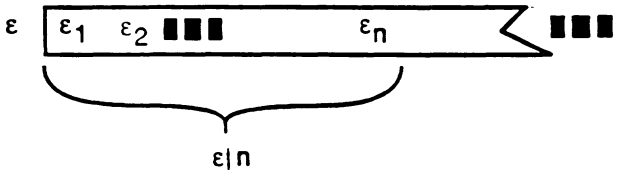
[4][Kemeny 53].

[5][Putnam 63], p. 270.

[6]instead of treating data as binary sequences, we could think of observations as being drawn from a recursively enumerable set E of mutually exclusive and exhaustive, possible observations. But binary data streams suffice for Putnam's argument, so we will assume that E is {0,1).

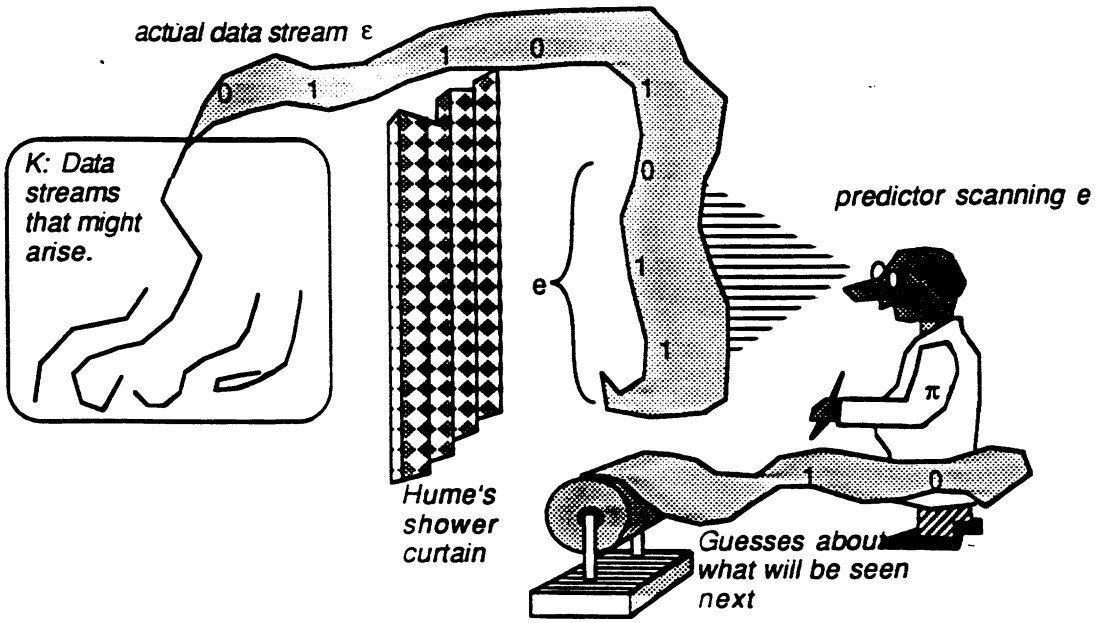
continue (0, 0, 0, 1, 1, 1), suggesting 0 as the next entry. In general, a rule that outputs a guess about what will happen next from the finite data sequence observed so far will be referred to as an ***extrapolator*** or ***predictor***.

In this situation, the extrapolator gets to see more and more of the infinite data stream ε through time. At each stage n, ε_n is received, and by time n all of initial segment ε|n is available for inspection.



We may know something in advance about what the data stream we are observing will be like. We might be told that all the sequences consist of a repeated pattern, or that they all converge to some value. Let K represent the space of all data streams that may arise for all we know or *care*.

The predictor, and his situation can now be depicted as follows. Some ε ∈ K is the actual data stream that π will face in the future. π reads larger and larger initial segments of ε and produces an increasing sequence of guesses about what will happen next. "Hume's shower curtain" prevents π from seeing the future, which would, of course, make prediction a bit too easy.



There are infinitely many possible methods for producing predictions. It remains to say what would count as a good method. It would be unreasonable to expect even the brightest extrapolator to be right *always*. It would also seem unreasonable to expect it to be right after some fixed time that can be specified *a priori*. Whatever that time is, there might be two possible extensions of the data that diverge only after that time. But we might hope at least that for each infinite data stream, there is some time (which may differ from one data stream to another) after which the extrapolator eventually "gets the gist" of the sequence and locks onto it, producing only correct predictions thereafter. We can think of this as a criterion of success for extrapolation methods in general. Let π be an extrapolator and let K be a specified set of data streams that we care about. Then

> π ***reliably extrapolates*** K ***in the limit*** ⇔
> *for each* possible data stream ε in K
> *there is* a time n such that
> *for each* later time m,
> π's prediction is correct (i.e. $\pi(\varepsilon|m) = \varepsilon_{m+1}$).

This criterion of success reflects ***limiting reliability***, since what is required of the extrapolator is *convergence* to the state of producing correct predictions, and this convergence is *guaranteed* over all of K, the space of data streams we care about. Whether or not reliable extrapolation is possible will depend heavily on the structure of K. If K contains only the everywhere 0 data stream, then we succeed over K by always guessing 0, without even looking at the data. If K includes all logically possible data streams, it is not hard to show that reliable extrapolation in the

limit is impossible, no matter how clever our method might be. Then there are intermediate cases, as when K = ***Rec***, the set of all data sequences that can be generated by a computer. That is just the example involved in Putnam's condition of adequacy for extrapolation methods:

(a) extrapolator π is ***adequate*** $\Leftrightarrow$ π reliably extrapolates Rec in the limit

Putnam didn't assume that this notion of adequacy would stand on its own. He was careful to explain that if no possible method is "adequate", then it is this condition of "adequacy" rather than Carnap's methods that must go. The condition is supposed to derive its force from the fact that it is both desirable and *achievable*.

I. B. Putnam's Diagonal Argument

Now we turn to the second step of Putnam's argument

(b) no inductive method based on a 'measure function' is adequate;

Before we can review Putnam's proof of (b), we must clarify what is meant by an inductive method *based on* a 'measure function'. Carnap's c-functions, or logical probability measures, are (in our set-up) conditional probability measures with special symmetry properties on the infinite product space E^{ω}, where E is an effectively enumerable set of atomic (mutually exclusive and exhaustive) possible observations. Such measures may be *turned into* extrapolation methods as follows. Let x be a particular observation and let e*x be the result of concatenating x to finite sequence e. Let c(e*x, e) be the probability that the next observation is x , given that e has been observed so far. Assume a fixed, effective enumeration $x_0, x_1, ..., x_n, ...$ of E. Then we can define predictor π_c so that π_c outputs the unique prediction assigned probability greater than 0.5 by c if there is one, and some "stall" character '#' otherwise.

$$\pi_c(e) = \begin{cases} \text{the first prediction } x \in E \text{ s.t. } c(e^*x, e) > 0.5, \text{ if there is one} \\ \# \text{ otherwise} \end{cases}$$

Observe that if c(e*x, e) is recursive in e and in x (as Carnap's methods are), then π_c is also recursive.

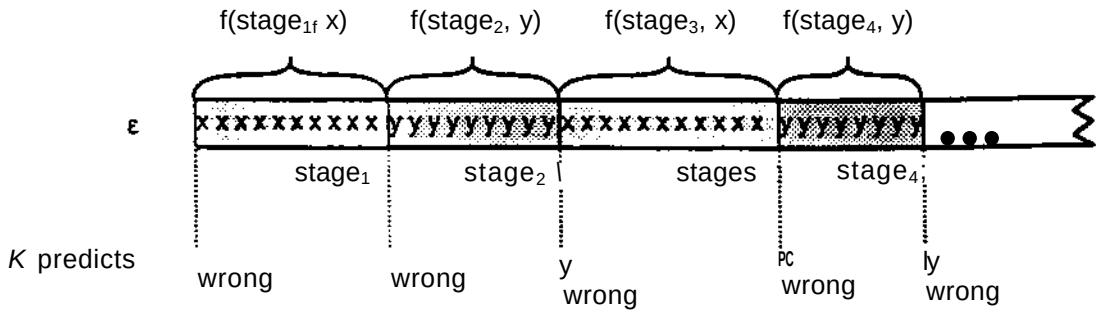
It turns out that when c is one of Camap's methods, the extrapolation method *Kc* has a special property, which is the only property of TCC that is relevant to Putnam's proof. Say that ft is ***gullible*** just in case no matter what has been seen so far, if we feed observation x to *n* often enough, *K* will eventually start to predict that x will occur next, *K* is ***recursively gullible*** just in case there is some effective procedure that enables us to calculate in advance how many x's must be fed to *K* to get a to predict x next, for each finite evidence sequence e and for each x. To be precise, let x^n denote the sequence (x, x,..., x), in which x is repeated n times. Then we have:

K Is ***recursively gullible*** <=>
there exists a computable function f such that
for each finite data segment e
7i predicts x after reading f(e, x) successive x's added to e
(i.e. 7t(e*$x^{f(e,x)}$>) = x)

By an *inductive method based on a 'measure function*[1]*,* we take Putnam to mean any extrapolation method ic_c, where c is one of Camap's c-functions. All such methods are recursively gullible. Thus (b) is implied by

(b^1) If *K* is recursively gullible then *n* does not extrapolate Rec in the limit

Putnam proves (b^1) by means of a diagonal argument. Since *K* is recursively gullible, we let f(e, x) be a computable function that tells us how many x's we would have to add to e to get ft to predict x as the next observation. At each stage, we check what ft predicted at the end of the previous stage (say x). Then we choose some datum y * x and use f to calculate how many y's must be added to the data e presented so far to get *K* to predict y. We add this many y's to e, so that *K* makes a mistake after reading the tast of the y's so added at this stage.



In the limit, *K* is wrong infinitely often. But e is effective since it is defined recursively in terms of the recursive function f. Thus e e Rec, so *K* does not extrapolate Rec.

Putnam mentions as a corollary that (b) remains true if recursive guilibiity is replaced with the more perspicuous condition that *K* be recursive.

(b") If *K* is recursive then *n* does not extrapolate Rec in the limit.

This also implies (b) because c(e*x, e) is recursive in e and x, and hence when c is one of Carnap's methods, He is recursive.

(b^N) follows from the fact that if *n* is recursive and is also able to extrapolate Rec, then *n* is recursively gullible. [7] For if we suppose that some recursive *n* extrapolates Rec, then it follows by (b^1) that *K* does not extrapolate Rec, which is a contradiction. To see that a recursive extrapolator of Rec is recursively gullible, suppose that recursive *n* extrapolates K. Define f(e, x) to be the least number of consecutive x's that, when added to e, leads *n* to predict x. f is clearly computable if *n* is, so the only worry is whether *n* will eventually predict x at all. If *K* doesni eventually predict x, then it is wrong all but finitely often on the data stream in which e is followed by all x's, which is also in Rec. But that contradicts the assumption that *K* extrapolates Rec.

I. C. Hypothetico-Deductivism

Now we come to the last claim of Putnam's argument:

(c) Some 7c extrapolates Rec in the limit.

The method that witnesses this fact is extremely simple. The idea is to enumerate predictive hypotheses, to find the first hypothesis in the enumeration consistent with the current data, and to predict whatever this hypothesis says will happen next. This basic architecture for choosing a hypothesis in light of data is known in the philosophy of science as the *hypothetico-deductive method* or the *method of conjectures and refutations,* in computational learning theory as *the enumeration technique,* in computer science as *generate-and-test search,* and in artificial intelligence as the *British Museum algorithm.* Putnam's interest in such proposals was inspired by an early article by Kemeny [1953] on methods that order hypotheses for test according to simplicity.

[7] The following argument is from [Gold 65]. A more recent version is given in [Osherson *et al.*, 86].

To adapt this venerable idea to the prediction of recursive sequences, we think of computer programs as hypotheses, so that the output of program p on input n is p's prediction about what will be observed at time n in the data stream. For concreteness, let computer programs be written in LISP. A LISP program (hypothesis) is ***correct*** for data stream e just in case it makes a correct prediction for each position in the data stream.

Now we must confront an issue that looks like a mere detail, but that turns out to be the crux of Putnam's argument. Anybody who has written a program in LISP knows that LISP permits the programmer to write programs with "infinite loops". Such a program is incorrect for *every* data stream, since it fails to predict anything when it goes into an infinite loop. If such programs occur in an effective hypothetico-deductivist's hypothesis enumeration, he can never be sure whether his current attempt to derive a prediction is caught in a complex infinite loop or whether it will terminate at the next moment with a correct prediction. If he uses some criterion for cutting off lengthy tests and concluding that he has detected an infinite loop, he might throw out ail the correct programs too early because their predictions take longer to derive than his criterion permits! If he hunkers down and insists on completing each derivation, he will freeze for eternity when he tests a program with an infinite loop. Either way his goose is cooked.

So the effective hypothetico-deductivist must eliminate all hypotheses with infinite loops from his hypothesis enumeration if he is to be a successful predictor. A program that never goes into an infinite loop is said to be *total.* An enumeration TI of LISP programs is ***total*** if each program occurring in the enumeration is. On the other hand, the enumeration must be *complete* as well, in the sense that it includes a correct program for each recursive data stream. Otherwise, the hypothetico-deductivist will clearly fail if the unrepresented data stream is in fact the one we receive. But it is perhaps the most basic fact of recursion theory that:

Fact I.C.1: An enumeration TI of programs is either *non-recursive* or *incomplete* or *non-total?*

In fact, this is not a special problem for the hypothetico-deductivist. It can be shown[9] that if K c Rec is reliably extrapolable in the limit by some recursive rc, then there is some recursive

[8][Rogers 87].

[9]The idea is due to [Barzdin and Freivaldsi972] and is applied in [Blum and Blum 75]. Let *n* be **a** recursive prediction method. Define program p_e as follows: If k < length(e), return ei<. Otherwise, feed e to *K* and thereafter feed the successive predictions of *K* back to *K.* Halt this process and output the kth prediction as soon as it is reached. If *K* extrapolates e, then after some time n, rc's

hypothesis enumeration that is total and complete for K. So in this sense *every* recursive extrapolator must somehow cope with Fact I.C.1.

Putnam's response to this inescapable limitation on computable extrapolation is to assume as "given" some non-computable oracle producing a complete and total hypothesis enumeration η. When a computable process is provided with a non-computable oracle, it is said to be ***recursive in*** the oracle. Accordingly, let η be a hypothesis enumeration. Now we construct an extrapolation method π_η that uses the given enumeration η as follows.

> $\pi_\eta(e)$:
> find the first hypothesis p in η that eventually returns correct predictions for all of e;
> predict whatever p says the next datum will be.

It is easy to see that

> **(c2)** If η is total then π_η is recursive in an oracle for η.

That's because the program for π can call for successive programs from the oracle η and then simulate each program received from the oracle for its agreement with e, the evidence received so far, producing its next prediction from the first such program found to agree with e. Since η is total, all consistency tests terminate. It is also easy to see that

> **(c3)** If η is total and complete then π_η extrapolates Rec in the limit.

Let ε be in Rec. Since η is complete and $\varepsilon \in$ Rec, some p occurring in η is correct for ε. Let p' be the first one (since there may be many correct programs). Then each preceding program p" is incorrect in at least one of its predictions. Eventually, the data exposes each of these errors, and p' is the first program consistent with the data. It is never rejected, and its predictions are all correct. So once p' is at the head of the list, π_η never makes another mistaken prediction.

---

predictions are all correct. Thus $p_{\varepsilon|n}$ computes ε. On the other hand, for each finite data sequence e, π extrapolates the sequence computed by p_e. Let $e_0, e_1, ..., e_n, ...$ be a computable enumeration of all possible finite data sequences. Then the enumeration of programs $p_{e_0}, p_{e_1}, ..., p_{e_2}, ...$ is computable, total, and complete over the set K of data streams extrapolated by π.

(c3) implies (c) and completes Putnam's argument. Putnam extends his argument by announcing one more fact, namely:

> (d) Any recursive extrapolator that extrapolates K *a* Rec can be improved to extrapolate one more data stream in Rec.

It's easy to see how. If *K* is recursive, then, as we saw in the proof of (b"), *n* is recursively gullible. By Putnam's diagonal argument we can use the gullibility function to produce a recursive data stream e missed by TC. Let p be correct for e. Now we can "patch" *n* by having it predict according p until p is refuted, and thereafter switch to its own predictions. The resulting, recursive predictor 7tp clearly succeeds wherever *n* does, but also succeeds on e. Since n_p is recursive, we can diagonalize again, patch again, and so forth to form an infinite sequence ?c, TC_{P1}, Jtp_2>... of ever better extrapolators of recursive data streams.

I. D. Hypothetico-Obscurantism

Putnam concludes his argument as follows:

> This completes the case for the statement made at the beginning of the paper: namely, that a good inductive judge can do things, provided he does not use 'degree of confirmation', that he could not in principle accomplish if he did use 'degree of confirmation'.[1] [10]

That is, if your extrapolator is $7r_c$, where c is one of Carnap's c-functions, it will be recursively gullible, and no recursively gullible extrapolator extrapolates each data stream in Rec. But the hypothetico-deductive method *K^* can extrapolate ail of Rec, so presumably we should *use K^* instead.

This sounds good— until we recall that the same objection applies to *all* computable predictors. By parity of reasoning, Putnam must also recommend that we use no *computable* methods because we could *do* more if we *used* the non-computable method rc^. The rub, of course, is that computability is the operative explication of what can be done *using* an explicit method. If Church's thesis is correct, then we cannot *use* TU^ in the sense in which a method is ordinarily taken to be used or followed, so we cannot *do* more using rc^ than we could using some

[10] [Putnam 63], p. 282, our emphasis.

computable method *n.* It would seem that Putnam agrees, since he asserts in the same article that an inductive method that is "never computable"[11] is "of no *use* to anybody".[12] So the apparent dominance argument against Carnap's methods fails. Carnap's methods are better insofar as they can be used. The uncomputable methods are better insofar as they would do more if they *could* be used.

So why should c-functions be branded as inadequate because, as computable methods, they cannot do as much as some useless, non-computable method? Putnam's answer is that JC^, though not recursive, is recursive *in an oracle iorr\.* And science should be viewed as having such an oracle "available" through some unanalyzed process of hypothesis *proposal.*

> ....1 suggest that we should take the view that science is a method or possibly a collection of methods for *selecting a hypothesis,* assuming languages to be given *and hypotheses to be proposed.* Such a view seems better to accord with the importance of the hypothetico-deductive method in science, which all investigators come to stress more and more in recent years.[13]

But now a different difficulty arises. If Putnam's favorite method is provided access to a powerful oracle, then why are Carnap's methods denied the same privilege? The real question is whether there is a method 7tc based on a conditional probability measure c that can effectively extrapolate all of Rec when provided with an oracle for TI. But of course there is! Just define (^(x, e) = 1 if x • *n^e)* and c^x, e) = 0 otherwise. These constraints induce a joint measure on the infinite product space co^{03} by the usual measure extension lemmas, and (^(x, e) is clearly computable in TJ.

Putnam seems to have recognized this weakness in his original argument, for in his Radio Free Europe address he proposed the possibility of constructing a measure c for each hypothesis stream *r* so that ^(e) « $7t_c$(e), for each finite data sequence e.[14] Then one may think of the c-function so constructed as implicitly "using" TI. NOW the objection to Carnap becomes more subtle. Carnap, together with contemporary Bayesians, insists that once a c-function is chosen,

---

[11] No rational valued c-function is "nowhere computable" in the sense that no program can find any value of it, since for every particular pair « x , e>, c(x, e)>, there is a program that waits for <x, e> and that produces c(x, e) on this particular input. Indeed, this can be done for any finite number of inputs. The puzzling expression "never computable" must therefore mean "uncomputable" if Putnam's claim is not to be vacuously true.

[12] Ibid., p. 276.

[13] [Putnam 63], p. 292, our emphasis.

[14] [Putnam 63a], p. 302.

all changes in degrees of belief should be regulated by c as more data is read. But if c is a fixed, computable method then some hypotheses will be missing from its tacit η. Putnam's point is that it would be *crazy*, upon being shown that some total program p is missing from η, not to add p to η to form η' and to switch from c to $c_{\eta'}$. But a Carnapian committed to c from birth is not free to do so. In modern Bayesian jargon, such a move would be *diachronically incoherent*.

But again, the same argument applies to *all* computable predictors.[15] When it is pointed out that one's method π fails to "consider" total program p, one ought (by parity with Putnam's argument against Carnap's methods) to switch to some method π' that does everything π does, but that also "considers" program p.

This sounds good, but we must again be careful. We already know that for any fixed, effective method π we choose, we could have chosen a more reliable one, so every effective method is "inadequate" in the sense that using it prevents us from being as reliable as we might have been. But from this point of view, Putnam's objection against following a fixed, computable method is weak. It's like saying of a man who asks a genie for a huge amount of money that he was *crazy* not to ask for twice that amount. If every chosen amount could have been larger, then he cannot be crazy for choosing a particular amount. So by analogy, unless the process of modifying our current method to account for newly "suggested" hypotheses leads to limiting performance better than that of any *possible* recursive extrapolator, we should not conclude that using a fixed, recursive method π is crazy, even in the face of a second-guesser who points out how we could have chosen better as soon as we choose π.

But whether a methodological patcher relying on a hypothesis oracle η can do more than any fixed, recursive method will depend heaviliy on the nature of η. Where could η come from? Putnam's own examples arise from diagonalization. Consider some hypothetico-deductive method π_η, where η is effective. We can effectively recover some total p missing from η. Instead of adding p to the front of η, we insert p into η after the first program consistent with the data at the previous stage, to form η'. Now suppose we keep doing this over and over. The added programs won't interfere with convergence, since they are always inserted behind our current conjecture, so it is a lot better to proceed this way than to stick with η. But everything done here is effective, so all of this Camus-esque striving to escape the bourgeois rut of following a fixed extrapolation algorithm for life amounts to nothing more than *following a fixed, recursive extrapolation algorithm*

[15] C.f. footnote 9 and (b") above.

for life, an algorithm whose own incompleteness can be effectively exposed by Putnam's diagonal argument!

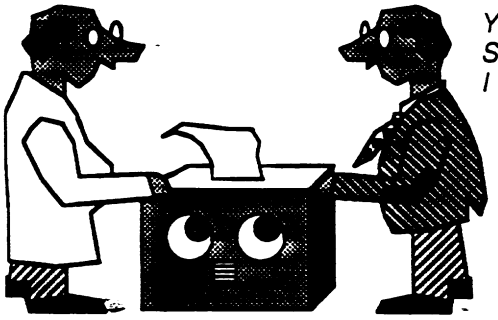
Let us then turn to Putnam's suggestion that η be *given*, rather than formed effectively by diagonalization. This raises the *possibility* that π_η can do more than any effective method, as when η is complete and total for some $K \subseteq Rec$ for which no effective η is total and complete. Let us refer to such an η as ***effectively transcendent***. But now a new problem arises: how are we to know that the enumeration η produced by a given source of hypotheses is, in fact, effectively transcendent? If we cast our lot with some π_η such that η is not effectively transcendent, then some fixed, recursive method π can do *better* than us!

It is hardly clear *a priori* that "creative intuition", "informal rationality", "common sense", or any other alleged source of hypotheses is effectively transcendent. Nor will it do for Putnam to appeal to scientific inquiry to determine the effective transcendence of a given hypothesis source, for it can be shown that in a very weak sense of reliable success to be introduced shortly, *no method*, whether *computable or not*, can figure out whether a given hypothesis source (e.g. Uncle Fred) will in fact produce an effectively transcendent hypothesis stream η by studying its past performance.[16] Thus, the proposal to inductively investigate the status of a given hypothesis source is a cure far worse than the disease, because at least some non-computable method can extrapolate Rec. Moreover, it can be shown that *no effective method* can figure out whether the oracle is even total, let alone transcendent.[17]

---

[16]Effective transcendency is not *verifiable in the limit* over ω^ω, in the sense of Section I.E. below. Let α aspire to this task, with no preconceived ideas about what Uncle Fred will suggest (after all, we are *relying* on Uncle Fred for our new ideas). Then we may diagonalize as follows. Suppose α succeeds. Let η be a fixed, effectively transcendent hypothesis stream. We present η until α says 1, which must happen eventually, since α succeeds and η is effectively transcendent. Now we continue in an effective manner until α says 0, which must happen eventually, and so forth, for each finite sequence of total programs can be extended in an effectively transcendent way or in an effective way. Each time we switch back to η we make sure then next item in η is presented. In the limit, all of η is presented (perhaps with some more total programs) and hence the hypothesis stream presented is effectively transcendent, since η is. But α changes its mind infinitely often, and hence fails, contrary to assumption. Thus no α succeeds, whether computable or not. This is much worse than the original problem of extrapolating Rec, which is impossible only for *effective* methods.

[17]Totality is not verifiable in the limit by an effective agent in the sense of Section I.E. below. Suppose recursive α succeeds. We fool α infinitely often by feeding it programs that simulate α and go into loops until α concludes that they will loop forever, at which point the loops all terminate. Specifically, suppose the finite sequence e of total programs has been presented to α so far. Let η be a fixed, effective, total hypothesis stream. We write a program that on input y simulates α on $e^*X^*\eta$ and that goes into a loop, waiting for α to conjecture some value ≤ 0.5 about the hypothesis of totality. As soon as such a value is produced by α, our program terminates the loop and outputs some arbitrary value, say 0. Note that X is a free parameter. By the Kleene



You need hypotheses?
Switch this stupid thing off.
I have informal rationality!

Scientist *Philosopher*

We can add uncertainty about the hypothesis source to the definition of inductive success so that its significance for successful prediction explicit. Our knowledge about the hypothesis source consists of some set S of infinite hypothesis streams, any one of which the source might generate, for all we know or care. Then reliable extrapolation using the source as an oracle can be defined as follows, where π now takes a finite, initial segment of η as well as a finite, initial segment of ε as input.

π ***reliably extrapolates*** K ***in the limit with a hypothesis oracle in*** S $\Leftrightarrow$
for each possible data stream ε in K
for each possible hypothesis stream η in S
there is a time n such that
for each later time m,
π's prediction is correct (i.e. $\pi(\varepsilon|m, \eta|m) = \varepsilon_{m+1}$).

Without going into a detailed analysis of this intriguing success criterion, it suffices to observe that just as extrapolability *simpliciter* depends essentially upon the space of possible data streams K, extrapolability with a hypothesis oracle depends crucially on the space of possible hypothesis streams S. So we have the intuitive conclusion that we had better *know* a lot more than we do

---

recursion theorem, X can be specified as the program p we have just defined, so it can feed itself to α in the simulation! Now if p stays in its loop forever, this is because α returns values greater than 0.5 forever, so α fails on $e^*p^*\eta$ which is not total, due to p, and we are done. If p exits its loop, this is because α's confidence in totality sagged at some point n in reading η during p's simulation. Then we really feed data $e^*p^*\eta$ to α until α really says 0, as p already determined in its simulation. Then we repeat the whole construction for a new internal simulation program p', and so forth, forever. Thus α's confidence in totality sags infinitely often on a total hypothesis stream, so α fails.

about our alleged oracles (e.g. "creative intuition", Informal rationality", "common sense") before we complain about how "inadequate" all the computable extrapolators are.[18]

There is yet another possible interpretation of Putnam's dissatisfaction with effective extrapolators. We have already seen that no recursive extrapolator succeeds over all of Rec, and that each such method can be improved, so there is no best. So no particular, recursive extrapoiator is ***universal,*** in the sense that it succeeds over all of Rec. On the other hand, for each effective extrapolator *, there is some choice of an effective *r* so that the hypothetico-deductive extrapolator *n^* succeeds on the same space K e Rec of data streams that *K* succeeds over. We may think of hypothetico-deductivism as a recipe or ***architecture*** for building extrapolators in a particular way, using an effective hypothesis stream and a test procedure. Since for every effectively extrapolable K, some effective hypothetico-deductive method *n^* extrapolates it, we may say that hypothetico-deductivism is a ***universal*** architecture for effective extrapolation. Universal architectures have the following, desirable property. While no method built in the specified way is guaranteed to succeed over all data streams in Rec, at least the architecture doesni stand in the way by preventing us from being as reliable as we could have been. \n particular, a scientist wedded to a universal architecture is shielded from Putnam's charges of inadequacy, since completeness implies that there is nothing he could have done by violating the strictures of the architecture that he could not have done by honoring them. This is something of a let-down from Reichenbach's grand ambition of a universal method for science, but it provides the basis for a well-motivated "negative methodology" in which methodological principles ought at least not stand in the way of progress.[19]

Perhaps Putnam inherited from Reichenbach the notion that methodology proper should consist only of universal principles. From this it would follow that methodology should consist only of maxims or architectures, rather than of particular, concrete algorithms for extrapolation since the

[18]Sirrtiar considerations can be raised against the various arguments based on Goedel's theorem [e.g. Lucas 61] which are intended to show that minds are not computers. These arguments assume that the "genius" who (effectively) constructs the Goedel sentence and who "sees" (by following the proof) that the sentence constructed is true *knows* that the system for which he constructs the sentence is sound. But this presupposes a reliable soundness oracle for arithmetic systems, and we do not *know* that the geniuses in question are reliable oracles for soundness. Nor could we reliably *find out* by watching what these geniuses do whether they are reliable soundness oracles if in fact we are computable, by diagonal arguments similar to those rehearsed in the preceding footnotes.

[19]Osherson, Stob and Weinstein refer to maxims or architectures as *strategies,* and refer to case in which maxims stand in the way of reliability as cases of *restrictiveness.* For a raft of restrictiveness results for recursive methods, c.f. [Osherson et. ai. 86], chapter 4.

former can be universal and the latter cannot. On this reading, the "informal rationality" and "common sense" required to apply methodological maxims is a matter of making a maxim into a concrete plan for action, whether explicitly, or implicitly, in light of one's cognitive dispositions to produce conjectures on the basis of data in light of the maxim. This interpretation fits well with Putnam's hypothetico-deductivism, for as we have seen, hypothetico-deductivism is a universal architecture for extrapolation that yields a concrete method only when the hypothesis stream η is specified. But it still doesn't account entirely for Putnam's antagonism toward explicit, recursive methods. To say that a universal maxim is good when every explicit instance of it is "ridiculous"[20] makes about as much sense as the 1992 presidential polls, which placed a Democrat in the lead, but each *particular* Democrat behind.[21]

In short, Putnam's negative philosophical morals about "mindlessly" following computable extrapolation methods are not very persuasive. But ultimately, that isn't so important. The lasting contribution of his argument was to illustrate how rich the study of effective discovery methods can be when approached from a logically sophisticated, recursion theoretic perspective. Putnam exposed intrinsic, formal structure in the problem of effective extrapolation that his predecessors never dreamed of, and that most philosophers of science and statisticians still know nothing about. He set the stage for an approach to inductive methodology in which dogmas and preaching give way to formal facts about what is possible, regardless of what we insist we "must have". His analysis was tied squarely to the objective of getting correct answers eventually, a refreshingly straightforward alternative to evasive epistemologies based on coherence or some primitive notion of "theory justification", whatever that might be; epistemologies that make the *point* of following a reliable method obscure.

The special strength of Putnam's analysis, which is almost entirely missing from the methodological work of his contemporaries, is its exploitation of the strong analogy between the fundamental concepts of computation and induction. Our inductive abilities are fundamentally limited because the data stream can only be scanned *locally*. If it could be seen all at once, extrapolation would be trivial. But a computing machine's computational abilities are limited by its inability to write upon or scan an infinite memory store at once. This fact is manifested in our

---

[20]'There is no logic of discovery'--- in that sense, there is no logic of testing, either; all the formal algorithms proposed for testing, by Carnap, by Popper, by Chomsky, etc., are, to speak impolitely, ridiculous: if you don't believe this, program a computer to employ one of these algorithms and see how well it does at testing theories! [Putnam 74], p. 268.

[21]In Section II we will see that Putnam's internal realist semantics provides a model for this absurdity, so perhaps he would endorse it.

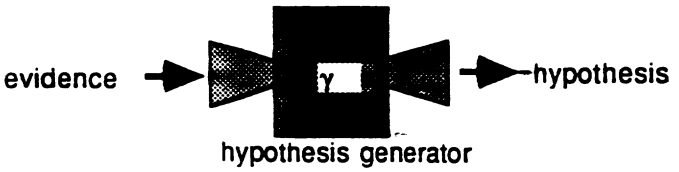
discussion of the hypothetico-deductivist's uncertainty regarding whether or not a given computation will halt. From a computer's point of view, the *formal* question of consistency becomes an internalized *empirical* question, and the machine's potentially infinite, ever extendable memory can be viewed as a second, internalized data presentation, only some finite segment of which can be scanned at a given time. This strong analogy between the leading metaphors of induction and computation poses the prospect for a logical study of induction entirely parallel to modern mathematical logic, both in style and in content. As in mathematical logic, the fundamental objects of inductive logic become problems and relations of relative difficulty between them.

Finally, Putnam's recursion-theoretic methodological results make no recourse to measures and probability, the traditional mainstays of inductive logic. Putnam's attitude is that extrapolation over some range K of possible data streams is a fixed problem, and that methods like Carnap's, that update probability measures over possible observations, are just one proposed solution among many. Whether or not such methods are good solutions to the prediction problem should drop out of a careful, comparative analysis. Nor does Putnam exempt probabilistic methods from failure over sets of data streams the *method* is "almost sure" wont arise, as in Bayesian, "almost sure-convergence theorems. Such theorems show only that each countably additive probability measure is willing to "bet its life" that it will get to the truth in the limit. In Putnam's analysis, the method is required to *really* succeed over all of K in virtue of its computational structure: he does not take the method's own word for its future success. Thirty years later, these proposals are still both radical and exciting.

I.E. Hypothesis Discovery and Hypothesis Assessment

The preceding discussion was about extrapolation methods, since Putnam's article puts matters this way, but his limiting reliabilist analysis applies just as well to other types of methodological problems. Sometimes a scientist desires to produce some correct hypothesis in light of the available data. Methods for doing this are called *logics of discovery* by philosophers, *estimators* by statisticians, and *learning machines* by cognitive and computer scientists. We will refer to methods that produce hypotheses on the basis of evidence as *hypothesis generators.*[22]

[22]The restriction to rational values eliminates problems about effectiveness later on.



Hypothetico-deductive method is most simply conceived of as a hypothesis generator constructed out of a hypothesis enumeration η and a special criterion of hypothesis test, namely, consistency with the data:

$\gamma_\eta^K(e)$ = the first hypothesis in η consistent with e in K.

Recall the criterion of reliable extrapolation in the limit. The idea was that the method should eventually "get it right" after some time, no matter which data stream in K is actual. We can also hope that our methods for discovery will eventually "get it right". In hypothesis generation, "getting it right" means producing a correct hypothesis. Recall, for example, that a LISP program is correct for infinite data stream ε just in case it correctly predicts each observation in ε. Generalizing that example, let ***hypotheses*** be objects in some countable set H, and let ***correctness*** be some relation R between infinite data streams in K and hypotheses in H.[23] Now we can be precise about the notion of "consistency" assumed in our definition of hypothetico-deductive method: e ***is consistent with h in*** K ***with respect to*** R just in case there is some ε in K that extends e and that makes h correct with respect to R.

An ***inductive setting*** is a triple (K, R, H), where K is a set of infinite data streams over some recursive observation space E, H is a decidable set of objects called hypotheses, and R is an arbitrary relation between K and H. For example Putnam's critique of Carnap's methodology essentially assumes that K = Rec, H = LISP, and R = Computes, where ***Computes***$(\varepsilon, p) \Leftrightarrow$ for each n, p outputs ε_n on input n. Putnam's setting (Rec, LISP, Computes) will be referred to as the ***computer modeling*** setting.

[23]Correctness is a very general notion. Correctness might be empirical adequacy (i.e. consistency with the total data to occur in ε) if hypotheses and observations are both drawn from some logical language. As in Putnam's discussion, we might also require that a correct hypothesis predict what will happen in ε. Or correctness might tolerate some number of mistaken predictions, or even an arbitrary, finite number. Simplicity and other properties of hypotheses may also be involved.

Now we can define what it means for a hypothesis generator to "get it right eventually" in an arbitrary inductive setting (K, R, H).

Y *makes reliable discoveries* *in* *the limit given* K (with respect to R) ⇔
for each infinite data stream e in K
there is a time n such that
for each later time m
y produces a hypothesis correct for e
(i.e.R(e,Y(e|m)))

This definition entitles Y to vacillate forever among correct hypotheses on some data stream in K. When Y does not so vacillate, we say that y makes ***stable*** discoveries in the limit.[24] Plato, among others, took stability to be a key feature of knowledge, distinguishing it from mere true belief,[25] so stable discovery is an idea of some epistemoiogical interest.

In the computer modeling setting (Rec, LISP, Computes), the extrapolability of K *Q* Rec in the limit implies reliable discovery in the limit is possible given K, and this claim holds both for effective and for ideal agents.[26] The converse is also true for uncomputable methods but is false for computable ones.[27]. Both directions are false for effective and for ineffective methods when we move from computer modeling to arbitrary inductive settings.

The results for reliable discovery in the computer modeling setting mirror Putnam's results for extrapolation. No recursive y makes reliable discoveries in the limit over all of Rec (this result due to E. M. Gold[28] strengthens Putnam's result), but the hypothetico-deductive method Y# succeeds when provided with a total and complete enumeration TI of LISP programs. There is no best, recursive, limiting discoverer of correct LISP programs. For each recursive generator, it is trivial to make an improved recursive generator that succeeds on one more data stream[29]. One difference between extrapolation and discovery in the computer modeling setting is that there is

[24]Add the condition that 7(e|m) = **y(e|n)** to the expression in parentheses.

[25][Plato 49].

[26]af. [Blum 75].

[27]Ibid.

[28][Gold 65].

[29][Osherson, *et al.* 86].

no guarantee of a recursive, total and complete hypothesis enumeration for K when effective discovery is possible over K in the limit.[30]

A method that assigns degrees of warrant or support in light of the data is called a ***confirmation theory*** or ***inductive logic.*** A method that outputs 1 or 0, for "pass" or "fail", respectively, is called a ***hypothesis test*** We shall refer to all these methods collectively as ***hypothesis assessors.***

Some confirmation theorists view the assignment of a degree of warrant or inductive support to a hypothesis as an end in itself[31]. But assessment is more often viewed as a means for reliably determining whether or not a given hypothesis is correct. A natural analogue of the preceding success criteria suggests itself, where we view H as the set of hypotheses we might possibly be asked to assess:

a *verifies* H *in the iimit given* K (with respect to R) <=>
for each hypothesis h in H
for each possible data stream e in K
R(e, h) »
there is a time n such that
for each later time m > n, $a(h_t\ e|m) > 0.5$.

Verification in the limit requires that a eventually place only values greater than 0.5 on h if and only if h is correct. When h is incorrect, a is free to vacillate forever between values above and below 0.5. When a eventually produces only values less than 0.5 if and only if h is *incorrect,* we say that a ***refutes*** H *in* ***the*** *limit given* K. When a both verifies and refutes H in the limit given K, we say that a ***decides*** H *in the limit given* K.

[3a]This result was presented by Mark Fulk in a Symposium at Carnegie Mellon University in 1989.

[31] E.g. [Horwich 91] and [Lycan 88].

We have discussed hypothetico-deductivism as an architecture for building a prediction method π_η from a given hypothesis enumeration η. Hypothetico-deductivism can be viewed more directly as a proposal for building a hypothesis generation method γ_η from η as follows. Say that h ***is consistent with*** e, K $\Leftrightarrow$ for some $\varepsilon \in K$ e $\subseteq \varepsilon$. Now define the hypothetico-deductive method γ_η determined by η as follows:

γ_η(e) = the first hypothesis h in η that is consistent[32] with e, K.

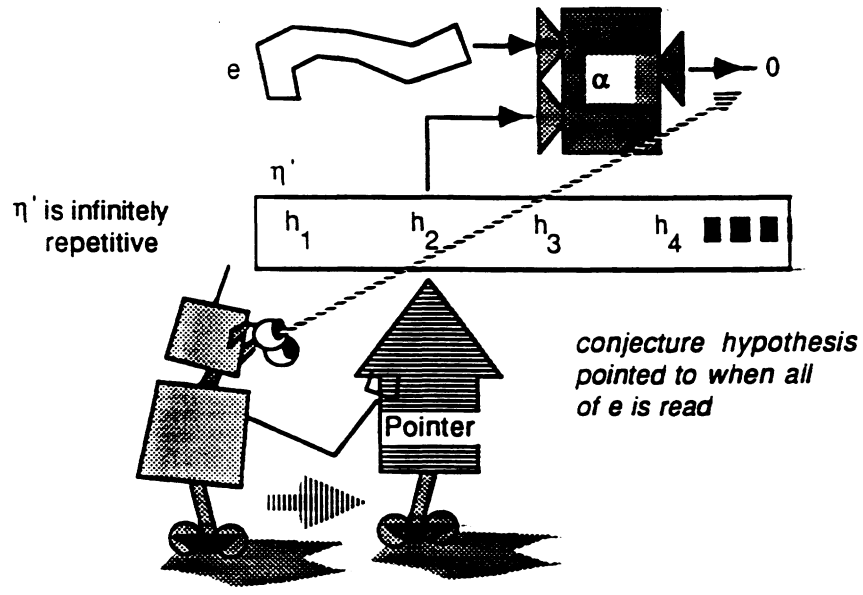
Consider the hypothesis h that at most finitely many 1's will occur in the data stream. This hypothesis is consistent with all finite data sequences, so hypothetico-deductivism is bound to fail. If, for example, h occurs prior to the first correct hypothesis h' in η, no evidence will refute h, and the hypothetico-deductive method will conjecture h forever, even if in fact the number of 1's in the resulting data stream is infinite. On the other hand, the trivial method that returns test result 1 if the last entry of e is 0 and 1 otherwise is an effective method that verifies h in the limit. So perhaps science would benefit by replacing the hypothetico-deductivist's narrow *consistency* test with a more reliable, limiting verification procedure α. Then, in our exemple, the hypothetico-deductivist would recommend producing the first hypothesis in η whose limiting verifier returns an assessment greater than 0.5. But this won't work if α vacillates between values above and below 0.5 forever on the first hypothesis, for then the hypothetico-deductivist will incorrectly conjecture the first hypothesis infinitely often.

A better proposal is this. Let η enumerate H, and let α be an assessment method. Let e be a finite data sequence of observations drawn from E. The ***bumping pointer method***[33] γ^{α}_{η}(e) works as follows. First, construct an infinitely repetitive version η' of η as follows: $\eta' = \eta_0, \eta_0, \eta_1, \eta_0, \eta_1, \eta_2, \ldots, \eta_0, \ldots, \eta_n, \ldots$. This can be done recursively in η. Initialize a pointer to position 0 in η'. The pointer will move as initial segments of e are considered. If the pointer is at position i on initial segment $\varepsilon|n$ of ε, then on segment $\varepsilon|n+1$, we leave the pointer where it is if $\alpha(\eta_i, \varepsilon|n+1) = 1$, and move it to position i+1 otherwise. γ^{α}_{η}(e) returns η_k, where k is the last pointer position upon reading all of e.

---

[32]In our setting, say that h ***is consistent with*** e, K $\Leftrightarrow$ for some $\varepsilon \in K$ e $\subseteq \varepsilon$.

[33]This method and the following argument were introduced in [Osherson and Weinstein 91] in the first-order hypothesis setting.



Say that H ***covers*** K ***according to*** R just in case each data stream in K bears R to some hypothesis in H. Then we have:

Fact I.E.1

(a) if rng(η) covers K according to R and α verifies rng(η) in the limit given K (w.r.t. R) then γ_{η}^{α} stably identifies R- correct hypotheses in the limit.[34]

(b) γ_{η}^{α} is recursive in α, η.

(b) is immediate. To see (a), let ε ∈ K. Since rng(η) covers K, we may choose h ∈ rng(η) so that R(ε, h). Then for some n, we have that ∀ m ≥ n, α(h, ε|m) =1. Moreover, h occurs in η' (the infinitely repetitive version of η constructed by γ_{η}^{α}) infinitely often, and hence at some position m' ≥ n in η'. Thus, the pointer cannot move beyond position m'. Now suppose h' occurs at some position m" prior to m' in η' and ¬R(ε, h'). Then α returns 0 infinitely often for h' on ε. Thus, the pointer cannot remain at position m" forever. So the pointer must remain forever at some correct hypothesis, and γ_{η}^{α} converges to this hypothesis. ■

Recall Putnam's recommendation that we view science as a set of universally applicable maxims that require ingenuity to apply. We may think of the bumping pointer proposal γ_{η}^{α} and of hypothetico-deductive proposal γ_{η} as general maxims or architectures that yield different,

[34] rng(ε) denotes the range of ε, or the set of all items that occur in ε.

concrete hypothesis generators for different specifications of η and of α. Recall that an architecture or maxim is *universal* or *complete* just in case it yields a successful method whenever there is one. What we have just seen is that the hypothetico-deductive architecture is not complete since *some* method can find the truth even when all hypotheses are all unfalsifiable, but *no* hypothetico-deductive method can succeed in such cases even in principle. This refutes Popper's bold conjecture[35] that the method of bold conjectures and refutations is our best and only means for finding the truth. But it remains to consider whether or not bumping pointer architecture is in fact universal. What is required is nothing less than a formal proof of completeness for a discovery architecture. In the next section we will see that in spite of the long-standing dogma that there is no logic of discovery, Putnam developed just such proof techniques to handle a purely logical problem posed by Mostowski.

I. F. Transcendental Deductions of Convergent Reliabilism

A ***transcendental deduction*** is a demonstration that some condition is necessary for knowledge of a certain kind. A transcendental deduction is ***complete*** if it yields both a necessary and a sufficient condition. If we think of stable, reliably inferred, correct hypotheses as knowledge (a view with a philosophical pedigree extending from Plato to the present day) then a complete transcendental deduction would be a necessary and sufficient condition for reliable inference in an arbitrary inductive setting (K, R, H). Among computational learning theorists, such results are known as ***characterization theorems***. In 1965, Putnam and E. Mark Gold published in the same journal the same characterizations for reliable verification, refutation, and decision in the limit.[36] Complete architectures for assessment and discovery drop out as a by-product.

The Gold-Putnam result characterizes verifiability in the limit by computable methods in terms of the ***computational complexity*** of the correctness relation relative to background knowledge K. The relevant scale of computational complexity is known as the ***Kleene*** or ***arithmetical hierarchy***. The basic idea is to define correctness in terms of quantifiers over a decidable

---

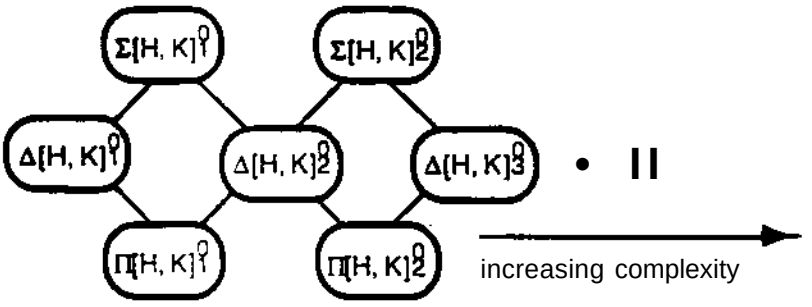
[35][Popper 68].

[36][Putnam 65], [Gold 65]. In personal correspondence, Putnam insists that his paper on Mostowski's problem was conceived quite separately from the paper on Carnap's methods. He does use inductive inquiry as a metaphor to motivate his construction, however. Gold clearly took the characterization result to have methodological significance, as did Putnam's student Peter Kugel [Kugel 77].

relation, and to measure complexity in terms of the number of blocks of adjacent quantifiers of the same kind in the resulting expression. For example, suppose that some correctness relation R can be defined *relative to K and H* as follows:

Ve € K, Vh e H, R(e, h) « 5w3xVyVz P(e, h, x, y, w, z)

where P(E, h, x, y, w, z) is required to be ***recursive*** in the sense that it can be mechanically decided by a LISP program armed with an oracle for the data stream e. Then we count the number of distinct blocks of quantifiers of the same type in the definition. Here we have the blocks 33, W, so the count is two. Then we say that R is in the complexity class $3^{H} * {}^{K}$$ just in case the first block of quantifiers in its definition consists of 3's and there are n blocks in all. R is in the dual complexity class $^{n}(^{H}$» Kin^{0} just in case there are n blocks of quantifiers in its definition starting with a block of V's, in which case $\overline{R}$ e **Z{H,** K]S. Finally, R is in complexity class A(H, K]8 just in case R is in both qH, K]? and n[H, K]S. A standard fact of mathematical logic is that these classes form a hierarchy. Links in the foiiowing diagram indicate proper inclusion, with smaller classes to the left and larger ones to the right.



What Putnam and Gold showed was the following:

Theorem I.F.I (Gold, Putnam):

(a) H *Is effectively verifiable in the limit given* K (w.r.t. **R) « R e** £[H, **K]2.**

(b) H *Is effectively refutable in the limit given* K (w.r.t. **R) « R e** n[H, K&

(c) H *is effectively decidable in the limit given* K (w.r.t. **R) « R G A[H,** $K]^{0}_{2}$.

It suffices to consider only the first case, the other two following readily. (=*) Suppose recursive a verifies H in the limit over K. Then by definition,

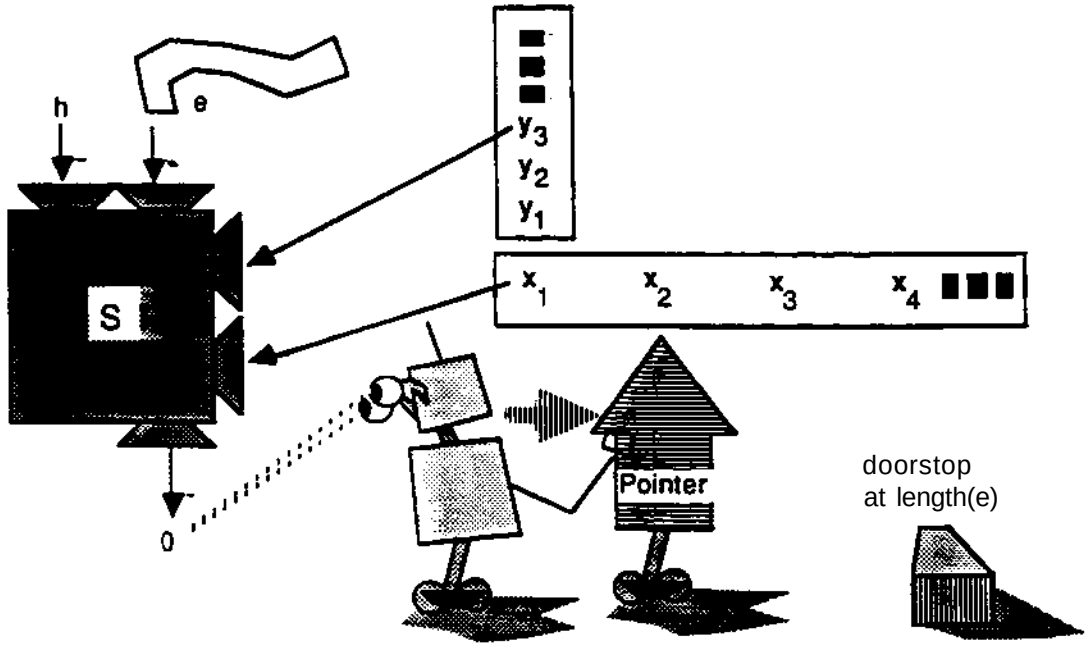
Ve € K Vh € H, R(e, h) *c*=> 5n Vm≥ n a(h, e|m) > 0.5.

where the relation "a(h$_f$ e|m) > 0.5" is recursive since **a** is. So **R € 5H,** K]$_2^0$.

(<=) Let R e ^ H . K fe^0. Then for some recursive relation S we have

Ve e K Vh € H, R(e, h) <=> 5x Vy S(e, h, **x, y),**

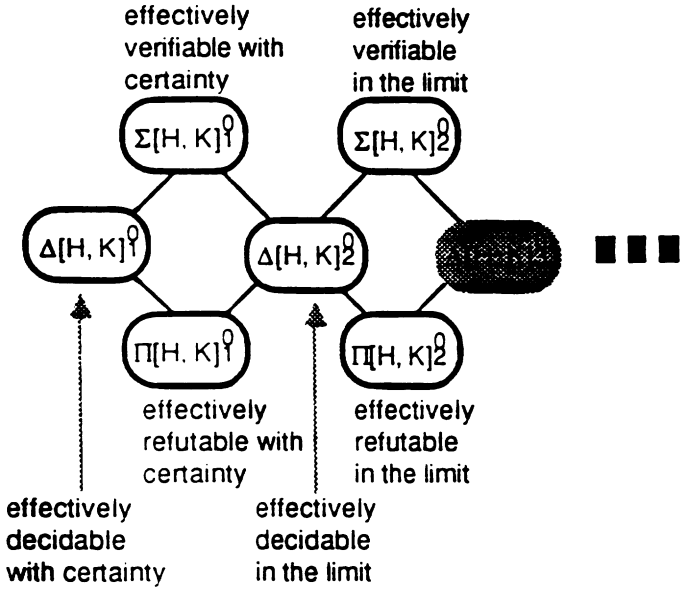
where x, y are **vectors of** variables. Assume **some** fixed, **effective enumeration x<i, X2, ·... x$_n$,...** of **the** possible **values for x** and another effective **enumeration y i , y 2 , . . yn. •• of the possible** values of **y. We now define** an effective **mechanism for managing a pointer on the enumeration xi , X2 , .. , x$_n$,... as follows:** on data e of length n$_t$ the mechanism **maintaining the pointer** seeks the first position k ≤ n such that for each j < n, the decision procedure **for** S(e, h, Xk, yj) does not return 0. If there is **no** such k, then the pointer moves all the **way** to position n.



We define the method a(e) to say 0 whenever the pointer moves and to say 1 otherwise. Now we verify that a works. Let h *e* H, *z e* K. Suppose R(e, h). Then 3k Vj S(e, h, XR, yj). Let k' be the least such k. Then the pointer can never be bumped past k\ so a stabilizes correctly to 1. Suppose -«R(e, h). ThenVk3j s.t. -.S(e, h, Xk, yj). Let Xk be arbitrary. Choose yj so that -tSfc h, Xk, yj). e is eventually long enough so that S(e, h, Xk, yj) halts with output 0. So the pointer is eventually bumped past k. Since this is true for arbitrary k, and a produces a 0 whenever the pointer is bumped, a produces infinitely many 0's, as required. I

The method constructed in the proof of this theorem can be thought of as an application of the bumping pointer architecture to hypothesis assessment. Thus, a corollary of the theorem is that bumping pointer assessment architecture is universal for verification in the limit.

The theorem can be extended to mechanical falsifiability and verifiability in the old, logical empiricist sense. Say that H ***is effectively verifiable with certainty given*** K (w.r.t. R) just in case there is an effective α such that for each hypothesis h in H and data stream ε in K, α eventually halts with 1 if and only if h is correct in ε. ***Effective refutability*** or ***"falsifiability" with certainty*** is defined similarly, except that α must halt with 0 just in case h is incorrect in ε. And ***effective decidability with certainty*** requires both verifiability and refutability with certainty. Then by the very definitions of the complexity classes, these cases are characterized, respectively, by $\Sigma[H, K]^0_1$, $\Pi[H, K]^0_1$, and $\Delta[H, K]^0_1$, yielding the following correspondence between success criteria and complexity classes:



We can also characterize stable discovery in a related fashion. The idea is that some effective discovery method works whenever any effective discovery method works, so the architecture for building discovery methods out of effective hypothesis enumerations and effective hypothesis assessors is ***complete*** in the sense of limiting reliability.

But there is an interesting twist. Contrary to the hypothetico-deductivists' intuitions, *effective,* reliable discovery of hypotheses in H is possible in the limit even when no *ideal* assessment method can even verify hypotheses in H in the limit. The following setting provides a simple illustration of this fact[37]

$\mathfrak{P}_0 = (K_0, R_0, H_0)$

$R_0(e, i) \Leftrightarrow e_0 = i$ or i occurs infinitely often in e.

K_0 « a)**.

$H_0 = \omega$

Thus, the successful discovery method cannot succeed by using a reliable assessment procedure, since there may be no such procedure when some method can succeed! The trick is that the method is free to "pretend" that correctness is ***more stringent*** than it really is, where the more stringent correctness relation R* c R is obtained from the original relation R by making some hypotheses incorrect where they were previously deemed by R to be correct. Then the discovery method can employ a reliable assessor a attuned to this "imaginary- notion of correctness, and whatever a judges correct will be correct, but not conversely.

This seems paradoxical! How could pretending that fewer hypotheses are correct for various data streams make it *easier* to find correct hypotheses? The answer is that making a correctness

---

[37]The tnvially effective discovery procedure "tfe) = the last entry in e^M makes reliable discoveries in the limit in this setting. On the other hand, let a be an arbitrary assessor, assumed for reductio to succeed in the limit in $\$_0$- To show that a does not verify co in the limit, assign a hypothesis 0 to investigate. Feed 111.... until a reports some value $\leq$ 0.5. Then fill in with 0's until a reports a value > 0.5, and so forth. Each such time must arise else a fails on the data we continue to feed, waiting for a to change its mind. If we make sure that a 0 is added each time a changes its mind, the result is a data stream for which 0 is correct, but a's confidence drops below 0.5 infinitely often.

Let *y(e)* = ei. y is trivially effective, and makes reliable discoveries in the limit. We show that co is not verifiable in the limit given co^ (w.r.t. R_o)~ even by an effective agent— by means of a simple Gold-Putnam diagonal argument, let a be an arbitrary assessor, assumed for reductio to verify co in the limit given coo (w.r.t. R_o). Pick hypothesis 2. Start out with 0, so that if 2 ends up being correct (according to Ro), it is not for the trivial reason that the data stream we feed has 0 in position 0. Now feed 2 until a's confidence rises above 0.5. This must eventually happen, else a's confidence always remains below 0.5 on a data stream with infinitely many 2's. When it happens, feed ail 0's until a's confidence falls to or below 0.5, which again must happen else a is eventually always more than 50% sure of 2 on a data stream that has only finitely many 2's and that does not have 2 in position 0. Continuing in the fashion, we produce a data stream with infinitely many 2's on which a's conficence vacillates above and below 0.5 infinitely often, so a does not verify 2 in the limit given coco (w.r.t. Ro). Contradiction.

relation more stringent can also make it far less complex, just as planing **wood** from a rough surface can make it smooth. We can choose R^f_o c Ro so that $R^f_o(e, i)$ « e begins with i, for example, thereby reducing the complexity of Ro from a hefty $n[co,K_o]2^9$ to a trivially manageable $A[co,K_o]?$. Thus, contrary to the usual hypothetico-deductive view, reliable discovery in the limit can in be much easier in principle than reliable assessment in the limit.

It is possible to make R too stringent. The simplest, and most stringent notion of correctness makes no hypotheses correct under any circumstances. But of course, such a relation is also a useless guide for what to conjecture. Thus we must insist that in reducing complexity, we retain some correct hypothesis for each data stream in K. Moreover, we will need to enumerate the relevant hypotheses, so we insist that the remaining, correct hypotheses be in some recursively enumerable set.

Theorem I.F.2:

Let (K, R, H) be an arbitrary inductive setting.

Correct hypotheses are effectively stably discoverable in the limit given K (w.r.t. R) »

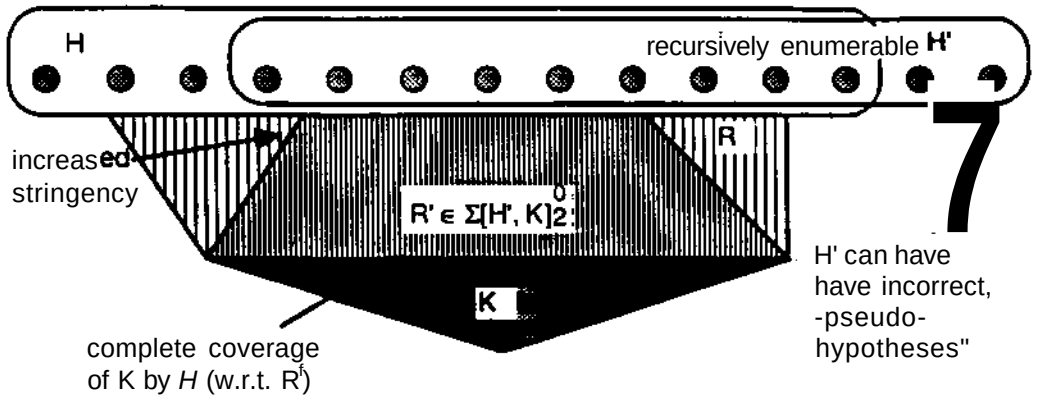
there is some R'cR.H's.t.

(1) H^f is recursively enumerable and

(2) H^1 covers K according to R' and

(3) $R'_6 I[H\backslash K]°_2$

Corollary: The theorem still holds if we add the condition: (4) FT is single valued.



(=*) Let recursive *y* stably discover correct hypotheses in the limit given K according to R. Then we have:

(i) $\forall \varepsilon \in K\ \exists n\ \forall m \geq n\ \gamma(\varepsilon|m) = \gamma(\varepsilon|n)\ \&\ R(\varepsilon, \gamma(\varepsilon|m))$.

Choose H' as follows:

H' = the set of all hypotheses conjectured by γ.

Next, define R' over H' and K as follows:

(ii) $\forall \varepsilon \in K, \forall h \in H', R'(\varepsilon, h) \Leftrightarrow \exists n\ \forall m \geq n\ \gamma(\varepsilon|m) = h$.

R' is a subset of R by (i) and by (ii). (1) follows from the definition of H' and the fact that γ is recursive. (2) follows from (i) and (ii). (3) follows from the form of (ii).

($\Leftarrow$) Suppose conditions (1 - 3) are met for some $R' \subseteq R$, H'. Then let recursive α verify H' in the limit given K with respect to the new relation R', by Theorem I.F.1 and (3). Let η be an effective enumeration of H' by (1). By (2) and by Fact I.E.1, the method γ_{η}^{α} is effective and stably discovers correct hypotheses in the limit given K w.r.t. R'. And if this hypothesis is correct according to R' it is correct according to R since $R' \subseteq R$. ■

This is a proof that the bumping pointer architecture is a universal maxim for discovery, but *not* if we insist in addition that α be a reliable test, for when R is not $\Sigma[H', K]_2^0$, there can be no reliable test (in the sense of verifiability in the limit) even though the bumping pointer method for some less complex $R' \subseteq R$ succeeds at discovery. It is remarkable that Putnam's purely logical work should have such powerful and immediate consequences for his philosophical thesis that scientific method should be viewed as a complete set of maxims.

The theorem also places Putnam's critique of Carnap's methodology in perspective, for the theorem shows that there is *exactly one way* in which effective discovery can be computationally harder than effective hypothesis assessment: it may be that no H' satisfying (3) is recursively enumerable. The computer modeling setting has just this special feature, so that total programs are reliably discoverable but not assessable. In other settings, discovery can be much simpler than assessment, as in the case of the setting $\mathfrak{P}_0$ discussed earlier.

The general point, however, is that Putnam's style of analysis provides a framework in which completeness and other logical concerns arise for inductive methods in just the manner that they arise for proof systems and for algorithms for other tasks, irrespective of whether we consider discovery, assessment, or prediction. The results reviewed are only a small sample of what can be done. We can consider different notions of computability (e.g. finite-state automata rather than LISP programs), different notions of success (e.g. stabilizing to within a finite set of correct hypotheses) and different side constraints (e.g. Bayesian coherence). For each mixture of conditions, we can seek a complete architecture, as well as a classification of concrete inductive settings into intrinsically solvable and unsolvable cases.

Putnam's early work sketches a mathematical edifice for methodology, complete with rooms, unexplored halls, and a partially stocked tool chest. Our question now is whether this edifice harmonizes or clashes with Putnam's more recent views about truth, when truth is viewed as an aim of reliable in inquiry.

II. Convergent Reliabilism and Truth as Ideal Rational Acceptability

We now turn to the second of the theses introduced at the beginning of this paper, which asserts that truth is a kind of idealized, epistemic justification. Putnam describes this conception of truth in his book, *Reason, Truth and History*:

> 'Truth', in an internalist view, is some sort of (idealized) rational acceptability--- some sort of ideal coherence of our beliefs with each other and with our experiences as those experiences are themselves represented in our belief system---- and not correspondence with mind-independent or discourse-independent 'states of affairs'.[38]

The operative standards of coherence and rational acceptability are psychologically based:

> Our conceptions of coherence and acceptability are, on the view I shall develop, deeply interwoven with our psychology. They depend upon our biology and our culture; they are by no means 'value free'.[39]

---

[38][Putnam 90], p. 50.

[39]Ibid., p. 55.

But truth is not *just* coherence or rational acceptability, because truth is stable in time and not a matter of degree, whereas the rational acceptability of a statement changes as the evidence or circumstances change and is also widely regarded to be a matter of degree.[40] To account for stability, Putnam appeals to *idealization*:

> What this shows, in my opinion, is ... that truth is an idealization of rational acceptability. We speak as if there were such things as epistemically ideal conditions, and we call a statement 'true' if it would be justified under such conditions. 'Epistemically ideal conditions', of course, are like 'frictionless planes': we cannot really attain epistemically ideal conditions, or even be absolutely certain that we have come sufficiently close to them. But frictionless planes cannot really be attained either, and yet talk of frictionless planes has 'cash value' because we can approximate them to a very high degree of approximation.[41]

The frictionless plane metaphor is rather vague. Fortunately, we get a bit more:

> The simile of frictionless planes aside, the two key ideas of the idealization theory of truth are (1) that truth is independent of justification here and now, but not independent of all justification. To claim a statement is true is to claim it could be justified. (2) truth is expected to be stable or 'convergent'; if both a statement and its negation could be 'justified', even if conditions were as ideal as one could hope to make them, there is no sense to thinking of the statement as having a truth-value.[42]

Finally, Putnam provides us at the very end of his book with an explicit reference to truth as an ideal *limit*:

> The very fact that we speak of our different conceptions as different conceptions of *rationality* posits a *Grenzbegriff*, a limit-concept of the ideal truth.[43]

It isn't hard to see how such views fit with Putnam's limiting reliabilist past. For we may conceive of ***rational acceptability*** as some hypothesis assessment function α, that somehow results from our cognitive wiring, our culture, and the accidents of our collective past together. Hypothesis h is then said to be ***true*** for a community whose standard of rational acceptability is α just in case α converges in some sense to a high assessment for h as evidence increases and "epistemic conditions" improve. We may think of ε as the data stream that arises for a community committed

---

[40]Ibid.

[41]Ibid.

[42]Ibid., p. 56.

[43]Ibid., p. 216.

to an ideal regimen of ever improving epistemic conditions concerning some hypothesis h.[44] Background knowledge K_{ideal} can be what the community knows ahead of time about how the data would come in if this regimen of improving epistemic conditions were to continue indefinitely.

Putnam claims to provide an "informal elucidation" rather than a formal theory of truth, but vague proposals license the reader to consider precise interpretations, and our discussion of Putnam's early work on induction suggests one. Let h be a hypothesis, and let £ be a data stream that might possibly arise under the assumption that we are committed to the continual improvement of our "epistemic conditions", so e *e* Kjdeai- Then define:

$$True_a(e, h) \Leftrightarrow \exists n \forall m > n\ a(h, e|m) > 0.5.$$
$$False_a(e, h) \Leftrightarrow \exists n \forall m \geq n\ a(h, e|m, h) \leq 0.5$$

This proposal defines truth in terms of what a hypothesis assessment method a does in the limit, a suggestion reminiscent of C.S. Peirce's definition of reality.

> And what do we mean by the real? * * * The real, then, is that which, sooner or later, information and reasoning would finally result in, and which is therefore independent of the vagaries of me and you.[45]

Peirce's motivation, like Putnam's, is to appeal to limits to wash out intuitive disanalogies between truth and rational acceptability. Therefore, we will refer to the general strategy of defining truth in terms of the limiting behavior of some methodological standard as the ***Pelrce reverse.*** In particular, $Truth_a$ is stable and independent of particular assessments by a, where a can be viewed as an arbitrary, socially or psychologically grounded standard of "rational acceptability", as Putnam intends.

$Truth1_a$ is trivially verifiable in the limit by a community whose standard of rational acceptability is a *because* a is the community's standard of rational acceptability. The same triviality does not extend to discovery or to prediction, however. Recall that in Theorem I.F.2 a necessary condition for effective discovery is the existence of an effectively enumerable collection of hypotheses covering K w.r.t. R. This condition may fail even when each hypothesis is verifiable in the limit, as in the computer modelling setting. Moreover, since extrapolation does not depend upon the

[44]This commitment to a *fixed* regimen of improving epistemic conditions will be relaxed considerably in Section III below.

[45][Peirce 58], p. 69.

notion of hypothesis correctness, internal realism does not make extrapolation any easier. Thus, Putnam's analysis of extrapolation stands unaffected by the move to truthi. Finally, verifiability, refutabiiity, and decidability with certainty are not trivialized by this proposal.

Putnam requires (in the above quotations) that a hypothesis and its negation cannot both be true. Clearly, truthi does not satisfy this condition over arbitrary choices of a, for if a assigns 1 to both h and -ih no matter what the data says, both h and -th will be truei. One solution would be to impose "rationality" restrictions on a that guarantee that the requirement in question will be satisfied. A natural such constraint would be that a(h, e) be a conditional probability. In that case, $a(h_t\ e) > 0.5$ implies a(—«h, e) $\leq$ 0.5, so Putnam's requirement that h and —ih not both be true is satisfied by truth1_a.

There are other, equally intuitive conditions that Putnam could have imposed. For example, it would seem odd if both h and h* were true but h & h* were not true. But this is possible under the definition of truthi even when a is a probability measure, since the probability of a conjunction can be much lower than the probabilities of the conjuncts, as in the familiar *lottery* ***paradox.*** Insofar as Putnam's gambit is to wash out the standard, intuitive disanaiogies between confirmation and truth by appeals to idealization and limits, to this extent he is unsuccessful.[46]

As we have already seen, Putnam has been critical of probabilistic methods anyway. Perhaps some different choice of rational acceptability standard a would guarantee that $truthi_a$ is well behaved. But it turns out that no a satisfying a plausible constraint (being able to *count)* can provide such a guarantee. The argument for this fact is of special interest, since it adapts Putnam's methodological diagonal argument to the investigation of internal realist semantics.

Let h_n be the sentence "as we progressively idealize our epistemic conditions in stages, observable outcome x occurs in at least n distinct stages of idealization". In our community, the sentence "there are at least n x's" is intimately connected with the practice of *counting,* in the sense that if we suppose the x's to be easily visible without a lot of effort, we can count and tally the number of x's seen up to the present, returning 0 until the tally reaches n, and returning 1 thereafter. That is what we do when we are asked if there are at least ten beans in a jar. We don't say "yes" until we pull out ten beans. Assume further that when considering -.h_n, the method says 1 until n x's are counted and 0 thereafter. Any method that assesses hypotheses of form h_n,

[46]Teddy Seidenfeld and Wilfried Sieg provided helpful comments concerning these issues.

$\neg h_n$ by means of counting occurrences of x in this manner will be referred to as a ***counting method***.

We needn't say what counting *is*, since social practices are the primitive in the internal realist's elucidation of truth. It suffices that counting somehow return a natural number n (the count of x's) for each finite data segment e presented to the counter. Indeed, we must be careful in our logical analysis of internal realism not to second-guess the practice of counting by asking whether its count of x's is *correct* relative to some independently specified number of x's in the data, since by the Peirce reverse, the truth about the number of x's in the data is fixed by the practice of counting and not the other way around. But our refusal to second-guess the accuracy of counting in this manner does not prevent us from listing some evident properties the practice of counting that make no reference whatsoever to what is "really" in the data:

> (A) No matter what data e has been shown to the counter, we can feed a stream of data that keeps the count of x 's fixed where it is forever.
>
> (B) No matter what data e has been shown to the counter, we can feed a finite chunk of data that makes the count of x's increase.
>
> (C) The count of x's never goes down.

Not all obvious explications of our society's standards of rational acceptabiliy are counting methods. For exemple, if the data stream is thought to be generated by independent tosses of a fair die, then $p(h_n) = 1$ for each n, and hence for each finite sequence e, $p(h_n, e) = 1$. Then $\alpha_0(h_n, e) = p(h_n, e)$ is not a counting method because a counting method cannot, by definition, output 1 for more than a finite number $h_1, \ldots, h_n$ of hypotheses on finite evidence segment e. Now conditional probability measures are thought by many methodologists to provide a good approximation of our practices of rational acceptability, so it would be hard for us to make an absolute case here that our community's actual standard of rational acceptability is a counting method. But we can do something easier; we can make our case conditional on the charitable assumption that internal realism is more or less correct and that the English speaking community is not radically confused about the meaning of "at least n x's will appear". To do so, we put the following question to the English speaking community:

> (Q) Suppose that we never count more than two x's in the data for eternity. Is it true or false that at least a billion x's eventually appear in the data?

We suspect that almost everybody would say "false", and that most would agree that a method like α_0 that says 1 no matter what is observed *could* lead in the limit to a different truth1 assignment on

hypotheses of form h_n than counting methods would lead to. In light **of** this **and** similar inquiries with similar results, we infer roughly, that either (1) the community **of English speakers doesnt** *understand* the simple English statement "at **least** n x's will **appear, or (2) it is internal realism rather than the** community's understanding that is **defective, or (3) internal realism is correct, the** English **speaking** community understands [M]at **least** n x's will **appear, and the standard of rational acceptability that grounds** usage for such **hypotheses** in **our community is a counting method. Respect both for Putnam** and for the English **speaking community dictates that we conclude:**

(3) Our society's standard of rational **acceptability** is **a counting method.**

Let ho be the sentence "as we progressively idealize our epistemic conditions in stages, observable outcome x occurs in infinitely many distinct stages **of idealization". Say that truth1$_a$ can** be ***oincomptete*** just in case for some data stream e, each h_n is $true1_a$, but h© is not $true1_a$. And if there is some e in which h^ is $truei_a$ and some -«h_n is $truei_a$ then **we** say that $truth1_a$ can be ***inconsistent.*** Now we may construct a Putnam-style diagonal agrument to show that:

Theorem 11.1:

If a is a counting method, then $Truthi_a$ can be either inconsistent or co-incomplete.

Let a be a counting method. Suppose for *reductio* that $Truthi_a$ can be neither ©-incomplete nor co-inconsistent. We present data as follows. By axiom (B), show a successive chunks of data that continue to make the count rise repeadedly, until a starts to return some value greater than 0.5 for h®. Such a time must come, else by axiom (C), we have that for all n, h_n is $truei_a$ but h© is not, so $truth1_a$ can be co-incomplete, contrary to assumption. As soon as a outputs a value greater than 0.5 for h^ we start presenting data in a way that will prevent the count from ever rising higher (by axiom (A)), until a's confidence in h^ drops below 0.5. This must happen, else there is an n such that -ih_n is $truei_a$ but h^ is also $truei_a$, so $truth1_a$ can be inconsistent, which is a contradiction. By repeating this procedure over and over (we don't have to *count* how many times we have done it) we end up (by axiom (C)) with a situation in which each h_n is $truei_a$ but h^ is not, so $truth1_a$ can be co-incomplete. Contradiction. I

By analogy, Goedel's first incompleteness theorem shows that a system of arithmetic is either co-inconsistent or incomplete. Although the "co" has switched sides (curiously enough), the import is the same: a methodological substitute for truth (e.g. mechanical proof, limiting rational acceptability) does not measure up to our intuition that truth should be complete and consistent.

Putnam might very well embrace the result. He has expressed a guarded admiration for intuitionism, and intuitionistic truth can be gappy when no construction is available either for p or for —«p. Nor has he claimed that internal realism's appeal to limits is supposed to "plug[11] these gaps. Thus, he may well have expected something much stronger than what we have shown, namely, that truth is *actually incomplete.* But whether or not the result was expected, it is interesting that it can be proved using the same sort of diagonal construction Putnam employed in his critique of Carnap's methodology. By interposing practice between the data stream and the method, Putnam's limiting reliabilist methodology is transformed into internal realist semantic theory. By playing the same game with different sets of sentences and output constraints on practices, one could no doubt prove more subtle and impressive things about internal realist truth.

Despite this potential for agreement, it is still worthwhile to consider some possible objections that an internal realist might raise, both to dispel them and to illustrate further how diagonal arguments bear on the internal realist's conception of truth. We have already addressed the objection that our argument assumes a "god's eye view" of the data to second-guess a's performance. In our argument, truth is entirely fixed by the practice of counting, and the social practice of counting is a primitive characterized entirely by three axioms that make no reference to what "really" occurs in the data. No external "semantics" of any sort is imposed on the hypotheses in question aside from internal realist truth.

An internal realist might also object, in light of passages like the following, that our diagonal argument bases internal realist truth on a *fixed* method a:

> I agree with Dummett in rejecting the correspondence theory of truth. But I do not agree with Dummett's view that the justification conditions for sentences are fixed once and for all by a recursive definition. * * * In my view, as in Quine's, the justification conditions for sentences change as our total body of knowledge changes, and cannot be taken as fixed once and for all. Not only may we find out that statements we now regard as justified are false, but we may even find out that procedures we now regard as justificatory are not, and that different justification procedures are better.[47]

But this is no objection to our argument, for if truth is defined as an ideal limit of rational acceptability, then truth is relative to the standard a of rational acceptability operative at the time,

[47] [Putnam 89], p. 85. The context of this passage is an attack on Dummetfs proposal that rational acceptability be defined in terms of a simple, Tarski-style recursion on formula complexity, according to intuitionist semantics. We treat the passage, out of context, merely as a potential source of objections among those familiar with Putnam's expressed views.

so truth changes as the standard α changes, but is fixed for eternity relative to a fixed standard, as Putnam requires. The diagonal argument shows that *none* of these successively adopted standards of rational acceptability grounds a notion of truth guaranteed to be both ω-complete and consistent, so long as it is a counting method. And we have already argued that our current α is a counting method.

Nor is it relevant to object that we assume α to be some *recursive* function of the data. We require only that α be a counting method satisfying the weak axioms (A), (B) and (C), which entails neither that α is recursive, nor even that α is *definable*.

A more promising response to our diagonal argument would be to adopt a less stringent convergence criterion for truth while retaining the basic spirit of the Peirce reverse, as follows:

$$\textbf{\textit{True2}}_\alpha(\varepsilon, h) \Leftrightarrow \forall s > 0, \exists n\, \forall m \geq n \;\; 1 - \alpha(h, \varepsilon|m) < s.$$
$$\textbf{\textit{False2}}_\alpha(\varepsilon, h) \Leftrightarrow \forall s > 0, \exists n\, \forall m \geq n \;\; \alpha(h, \varepsilon|m) < s.$$

A hypothesis is true2 if and only if the society's standard α of rational acceptability produces assessment values that move ever closer to 1, possibly not ever arriving there. Unlike truth1, truth2 does not require that the assessment value drop below some fixed threshold (0.5) infinitely often when h is not true. This laxity makes it possible to construct a community standard α of rational acceptability that guarantees that truth2$_\alpha$ is both ω-complete and consistent.[48] A method β that ensures falsity2$_\beta$ to be both ω-complete and consistent can also be constructed.[49]

But the internal realist is not home free, for he cannot find an adequate δ that *jointly* guarantees that truth2$_\delta$ and falsity2$_\delta$ will be ω-complete and consistent.

Fact II.2: If α is a counting method then

either truth2$_\alpha$ or falsity2$_\alpha$ can be either ω-incomplete or inconsistent.

---

[48]We are indebted to Jeff Paris for suggesting the following construction: define $\alpha(h_\omega, e) = 1 - (1/\#_x(e))$, where $\#_x(e)$ is the current *count* of x's in e, and for each n, let $\alpha(h_n, e) = 1$ if $\#x(e) \geq n$ and = 0 otherwise. Now if the count goes up forever in ε, then h_ω is true2$_\alpha$ and each h_n is true2$_\alpha$; and if exactly n x's are counted in ε, then for each $k > n$, h_k is false2$_\alpha$ and for each $k' \leq n$, k' is true2$_\alpha$.

[49]Define $\alpha(h_\omega, e) = 1/\#_x(e)$. If the count of x's goes up forever in ε, α goes to 0 and if only finitely many x's are ever counted, α stops short of 0 forever.

To see why this should be so, we may extend our transcendental deductions to cover the notion of reliable success in the limit implicit in the definitions of truth2 and falsity2:

α ***weakly verifies*** H ***in the limit given*** K (with respect to R) $\Leftrightarrow$
for each hypothesis h in H
for each possible data stream ε in K
$R(\varepsilon, h) \Leftrightarrow$
for each real number $r > 0$
there is a time n such that
for each later time $m > n$, $1 - \alpha(h, \varepsilon|m) < r$.

Of course, truth2 trivializes weak verifiability in the limit just the way truth1 trivializes verifiability in the limit. Weak refutation in the limit is defined dually, and weak decision in the limit requires that a single method both weakly verify and weakly refute H in the limit given K. Following the strategy of Theorem I.F.1 above, we arrive at a characterization of weak hypothesis assessment in terms of arithmetic complexity.

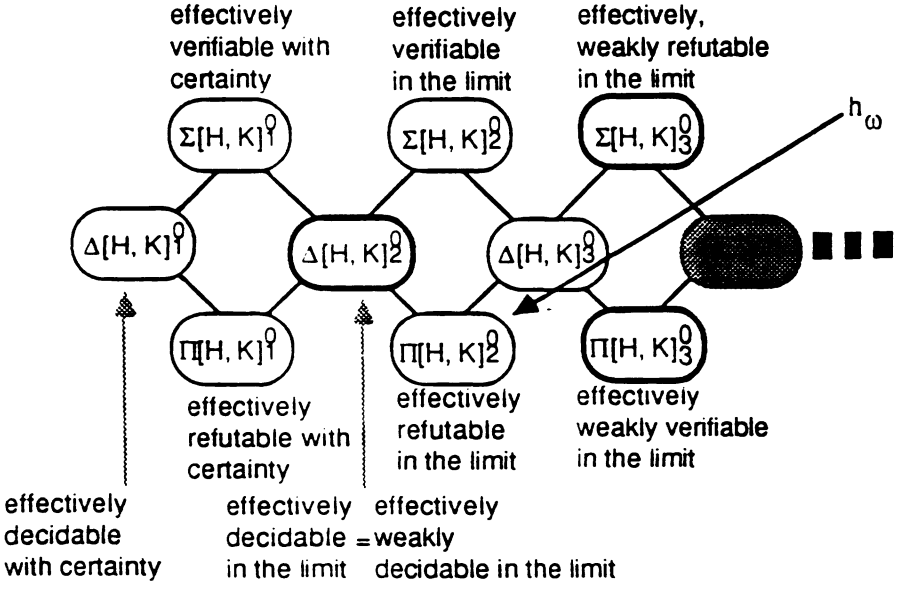
Theorem II.3:

(a) H ***is effectively weakly verifiable in the limit given*** K (w.r.t. R) $\Leftrightarrow R \in \Pi[H, K]^0_3$.

(b) H ***is effectively weakly refutable in the limit given*** K (w.r.t. R) $\Leftrightarrow R \in \Sigma[H, K]^0_3$.

(c) H ***is effectively weakly decidable in the limit given*** K (w.r.t. R) $\Leftrightarrow R \in \Delta[H, K]^0_2$.

Combining this result with Theorem I.F.1, we have the following complexity classification of our various notions of reliable hypothesis assessment:



Consider clause (c) of the Theorem II.3, which states that weak decidability in the limit is characterized by $\Delta[H, K]^0_2$ rather than by $\Delta[H, K]^0_3$, as might be expected by analogy with the other cases. (c) follows from the trivial fact that a weak, limiting decision procedure just *is* a limiting decision procedure, and decidability in the limit is characterized by $\Delta[H, K]^0_2$. This is what underlies Fact II.2. From a realist's point of view, the diagonal argument for Fact II.1 shows that h_ω (under the "obvious" realist interpretation) is decidable in the limit over arbitrary data streams in E^ω.[50] In light of (c), if both truth2$_\alpha$ and falsity2$_\alpha$ were guaranteed to be ω-complete and consistent, then α would weakly decide h_ω in the limit (according to the "obvious" realist interpretation of the hypotheses h_ω, h_n). But as we have just remarked, this is equivalent to demanding that α decide h_ω in the limit, and this is impossible, since no method (effective or otherwise) can even verify h_ω in the limit (by the proof of Theorem II.1).

Parts (a) and (b) of the Theorem II.3 explain, further, why truth2$_\alpha$ and falsity2$_\beta$ can both be guaranteed separately to be ω-complete and consistent so long as they are based on distinct methods α and β. Correctness for h_ω, together with each h_n, $\neg h_n$, is a $\Pi[H, K]^0_2$ relation, and hence is both a $\Sigma[H, K]^0_3$ relation (so falsity2$_\beta$ works out) and a $\Pi[H, K]^0_3$ relation (so true2$_\alpha$ works out). But for both truth2$_\delta$ and falsity2$_\delta$ to work properly when based on a single method δ, correctness would have to be a $\Delta[H, K]^0_2$ relation, which it is not. Thus, a fellow possessed of both a weak, limiting verifier and a weak limiting refuter of h_ω is in a curious position. The first method will converge to 1 on h_ω if some x occurs infinitely often in ε and the second machine will converge to 0 on h_ω if no x occurs infinitely often in ε, but there is no way to assemble these two methods into a single method that has both properties.[51]

We finish this section with a proof of Theorem II.3. (a). (b) follows by duality. The proof is of independent interest, since it extends Putnam's program of transcendental deductions to a weaker criterion of success, and it also exhibits a universal architecture for this notion of success.

---

[50]For brevity and mathematical clarity, we revert to the realist mode of expression, but the discussion can be reworked into a more "internally realistic" version parallel to the proof of Theorem II.1.

[51]By similar reasoning, it can be seen that weak, limiting verification does not guarantee the existence of a reliable discovery procedure, either. This is why Reichenbach's straight rule estimates only rational-bounded intervals around limiting relative frequencies, rather than limiting relative frequencies themselves. The exact value of a limiting relative frequency hypothesis is weakly verifiable in the limit, but that fact doesn't help one to discover them.

(=*) Suppose that a effectively, weakly verifies H in the limit given K (with respect to R). Then by definition

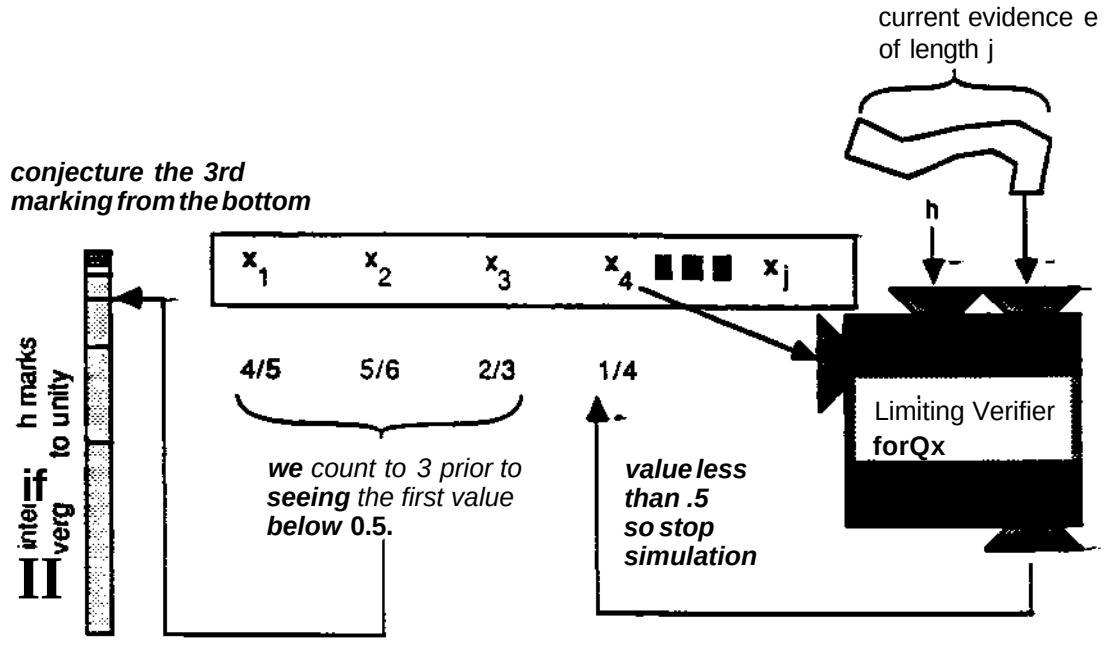
$$Ve\ e\ K\ Vh\ e\ H_t\ R(e, h) \Leftrightarrow^* Vreai\ r > 0\ 3n\ Vm \geq n\ 1 - a(h_f\ e|m) < r.$$

We can replace the quantifier over reals with a quantifier over rationats since for every real greater than 0 there is a smaller rational greater than 0. And we can effectively encode rationais with natural numbers, so that the quantifier over reals greater than 0 can be replaced with a quantifier over natural numbers. Since i-a(h, e|m) < r is recursive if a is recursive and r is a rational effectively encoded as a natural number, we have that R e n[H, K]£.

(«=) Suppose that R € n[H, K& Then for some recursive relation S we have

$$Ve \in K\ Vh\ e\ H, R(e, h)\ c^*\ Vx\ 3y\ Vz\ S(e, h, x, y, z),$$

where x, y, z are vectors of variables. Let k be the length of x. Then for each a € co^k, $Q_a(e, h)$ » 3y Vz S(e, h, a, y, z) is a I(H, K]2 relation. An examination of the proof of Theorem I.F.1 reveals that we can construct a (rational valued) assessor $a_a(h, e)$ recursive in a, h, and e so that for each a € co^k, aa verifies H in the limit given K w.r.t. Q_a. Now we use ota to construct an effective, weak limiting verifier a of H given K (w.r.t. R) as follows. First, we effectively construct an infinite sequence 0of rationais that converges to unity (e.g. 1-1/2, 1-1/4, ..., $1-1/2^n$, ...) and we effectively enumerate (o^k as (a<f, a2,..., a_n, ...). so that each a; is a k-vector of natural numbers. Now let h *e* H and finite data segment e of length j be given. We calculate b = (aai(^{h}» ^{e}K 0^2^' e). . . . .cc_aj(h, e)), and set w = the greatest x ≤ j such that bi, b2,..., b_x are all greater than 0.5. Then output 0_w. Observe that each of these operations is effective.



Now we show that a works. Let h *e* H, e e K. Suppose R(e, h). Then for each a», 3y Vz S(e, h, a;, y, z), so Qaj(e. h). Then for each i, a_{aj}eventually produces only values strictly greater than 0.5. Let k be given. Then there is a time n such that for all m ≥ n, cta^h, e|m) > 0.5, $cxa_2(^{h}.\ el^{m}) > °\text{-}^{5}$» ..., and aa_k(h, e|m) > 0.5. Thus, for all m ≥ n, a(h, e) > 6k. Since k is arbitrary, we have that for all r > 0 3n Vm ≥ n 1 - ot(h, e|m) < r, as required. Now suppose that -IR(E, h). Then for some a;, we **have** that -»Qaj(e, h). Then for infinitely many m, ^ (h , e|m) ≤ 0.5. But for each such m, a(h, e|m) ≤ 9j. Thus a weakly verifies H in the limit given K (w.r.t. R), as required. I

III. Reliability **and** Relativism

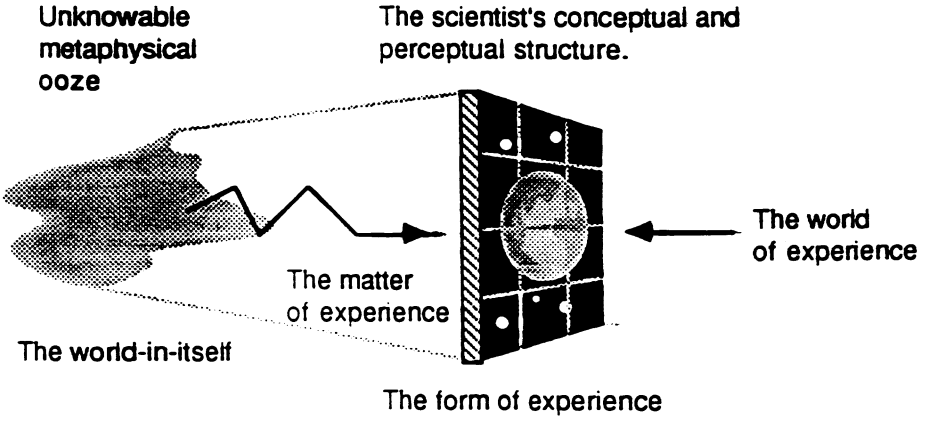
We now turn to the third of Putnam's theses, moderate relativism. Putnam puts the matter in **terms of *reference.***

> "What does the world consist of is a question that it only makes sense to ask within a theory or description".[52]
>
> ... a sign that is actually employed in a particular way by a particular community of users can correspond to particular objects *within the conceptual scheme of those users.* Objects do not exist independently of conceptual schemes.[53]

[52][Putnam 90], p. 49.

[53][Putnam 90], p. 50.

The passage reflects Kant's distinction between the world-in-itself and the world of experience. The world-in-itself does not come pre-partitioned into distinct individuals and relations so reference and truth make no sense in relation to it. But relative to the conceptual and perceptual apparatus of the perceiver, truth and reference are possible.



This sort of thing admits of degrees. ***Coherentists*** exclude the role of the world-in-itself altogether, insisting that any coherent set of beliefs is true *because* we believe it.[54] ***Naive realists*** insist that the conceptual scheme is irrelevant, so that experience is a direct apprehension of things in themselves. ***Moderate relativism*** covers the interesting ground between these extremal positions. Putnam pursues this moderate course, in which truth and evidence depend both upon the world-in-itself and upon our conceptual contribution:

> Internalism does not deny that there are experiential inputs to knowledge; knowledge is not a story with no constraints except internal coherence; but it does deny that there are any inputs which are not themselves shaped by our concepts....[55]

In traditional philosophy, the world-in-itself has been taken as the objective, external, or mind-independent component of truth, while the conceptual scheme has been identified with the subjective, internal, or mental component of truth. In methodology, we don't care about the purely metaphysical distinction between objectivity and subjectivity. We care about methods and *strategies*, where the crucial issue is *control*. For our purposes, the world-in-itself is the

---

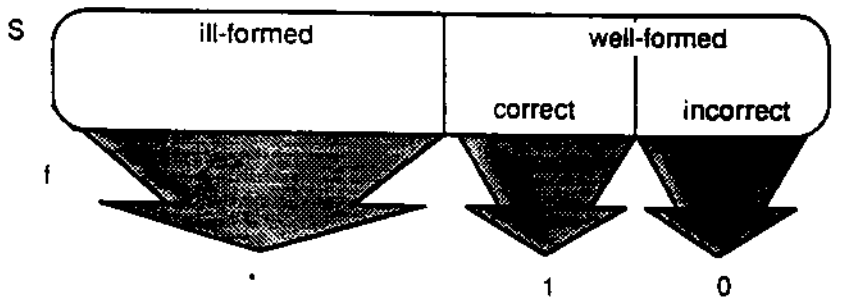
[54]Isaac Levi's position is exactly this [Levi 83].

[55][Putnam 90], p. 54.

component of truth that cannot be manipulated at will by us. The conceptual scheme is the component that can. In Putnam's terms, we adopt a ***functionalist*** perspective on the metaphysics of relativism, since that is all that matters for the purposes of reliabilist methodology. So for example, Kant, who admitted a strong, subjective component in truth, was nonetheless a functional realist, since he took the conceptual scheme to be fixed for ail humanity (though it might be different for other creatures).

III. A. **Worlds-in-Themselves**

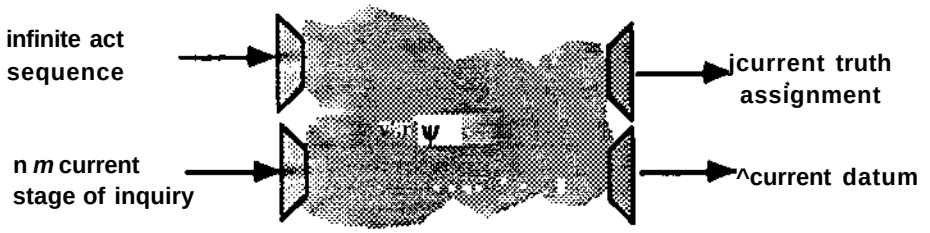
in general, a functionalist world-in-itself is some *dependence* of truth and evidence upon voluntary acts by the scientist. These acts may include belief change, conceptual change, experimental set-ups, changing what is meant by correctness, changing the color of your tie (in the case of social constructivist theories of truth), and finally, the act of making a particular conjecture. More precisely, let E once again be a set of possible observations. Let S be the set of finite strings over some suitably large alphabet. A ***correctness assignment*** lor H will be some map f: S -» {1, 0, *}, where * indicates "ill formed", 1 indicates correctness, and 0 indicates incorrectness. Correctness might be truth, something stronger, or something weaker.[56] Let C be the space of all correctness assignments.



A hypothesis generator produces some hypothesis h and may perform some other semantically relevant acts summarized by a e A. Thus the pair <h, a> *e* H X A summarizes ail of the scientist's semanticaliy relevant acts at a given time. A *wortd-in-itself* is then a map that takes an infinite sequence of complete semantic acts <hi, ai>, <h_2, a_2>, ... together with some specified moment

[56]We might have added other kinds of semantic status, including well-formed but no truth value, and so forth. Nothing of interest depends on these choices in what follows.

n of inquiry, and returns to the scientist the datum x for stage n and the current truth assignment f for stage n. That is, a world-in-itseif is most generally a map y: ((H X A)°> X co) -»(E X C).



It is often assumed by relativist philosophers of science that only the scientist's *current* conceptual scheme (act) is relevant to the current truth assignment to hypotheses. Such worids-in-themselves will be called *semanticatly immediate.* Some Marxists and feminists claim that one's entire ideological history is relevant to truth in the present. Worlds of this sort are *semantlcally local* but not immediate. The limiting conception of truth promoted by Peirce illustrates that the situation can be much more general, for in that case, the truth depends upon what is conjectured by the scientist *forever* Such dependencies (reminiscent of the medieval problem of God's foreknowledge and future contingents) are *semantlcally global.* There are still other possibilities. Truth can depend only on the current time, quite independently of the acts of the scientist. Then we say that the wortd-in-itself is *semanticatly spontaneous.* This would be the case if the laws of nature slowly evolve through time independently of anything we do, as some cosmologists have entertained, or if hypotheses involve indexicals such as "now" or 'this". If the truth assignment is constant, we say that the world-in-itself is *semantlcally fixed.* This is the position of an extremely naive realist.

All of these distinctions make sense for evidence as well, but we know of no published version of relativism that cannot be modeled with evidentially spontaneous or local worlds. For example, Kuhn's proposal is evidentially and semantically local. Quine's holism is semantically immediate (only the current "web" of belief matters) and evidentially spontaneous (if "evidence" is taken to be "surface irritation"). Naive convergent realism assumes that the world-in-itself is evidentially spontaneous and semantically fixed. This is the sort of setting assumed in Putnam's critique of Carnap, as reviewed in Part I of this paper. A wide range of philosophical positions can be parametrized in terms of various constraints on evidentially local worlds-in-themselves.

Finally, some worids-in-themselves are functions whose values do not depend upon what the scientist conjectures, but only upon acts independent of conjecturing. If we model a logical

positivist setting by interpreting a to be a conventional choice of analytic truths or meaning postulates, then truth depends only on this choice, and the conjecture of an empirical (non-analytic) h would have no bearing on meaning or truth. Such worlds-in-themselves will be called ***conjecture-independent***. By making truth depend upon all of one's beliefs, Quine insists that the world-in-itself is ***conjecture-dependent***, if we take the conjecture to be added to the current set of beliefs. An economic prognosticator who brings down the market by predicting doom provides a more vivid example of conjecture dependency.

III. B. Transcendental Background Knowledge

In relativistic settings, background knowledge is a set K of possible worlds-in-themselves. It may be objected that such transcendental knowledge is hopeless to obtain. But whatever our own views on the possibility of such knowledge may be, literary theorists, metaphysicians and philosophers of language regularly propound transcendental theses as a matter of course. The naive realist considers only semantically fixed and evidentially spontaneous worlds-in-themselves so that, the world-in-itself collapses to a single data stream and a single notion of truth. Coherentists, at the other extreme, are sure that all that matters to truth or to data is which coherent set we decide to believe. The positivists knew that truth depends only upon the free act of adopting some special set of conventional, analytic truths. Quine knows the negation of this claim.[57] Kuhn knows that the scientist's training, colleagues, and stock of solved examples are what determine truth and evidence. He even speaks of scientists with different backgrounds as inhabiting different "worlds".[58] Hanson knew that evidence depends upon perceptual and conceptual "gestalts".[59] Social constructivists know that truth is a matter of community assent, so the world-in-itself is the causal disposition of the community to assent in response to how one interacts with the community. Many Marxists know that truth is a matter of one's political role.

We can also consider the transcendental knowledge implied by Putnam's internal realism. His moderate relativism is characterized by at least the following principles:

---

[57][Quine 51].

[58][Kuhn 70] pp. 111-112.

[59][Hanson 58], chap i.

(P1) *Reference* depends upon the community's current *conceptual scheme*.[60]

(P2) *Individuation* depends upon the community's current *theory*.[61]

(P3) *Experience* depends upon the community's *belief system*.[62]

(P4) *Truth* is an ideal limit of rational acceptability under increasingly idealized *epistemic conditions*.[63]

(P5) Whether or not a given sequence of epistemic conditions is *increasingly idealized* depends on our *knowledge*.[64]

For simplicity, we will identify *conceptual schemes*, *knowledge*, *theories* and *belief systems*, and we will also identify *experience* with *evidence*. We expect that *truth*, *individuation*, and *reference* are closely enough related to be assimilated to *truth* in our discussion. *Epistemic conditions* are something else again. We think of them as alterations to our local environment that affect our powers of observation (e.g. building larger and larger radio telescopes, peering more closely, etc.).[65]

Then Putnam's proposal seems to be this. Semantically relevant acts are triples <B, α, c>, where $B \subseteq H$ is the current belief system, α is the current standard of rational acceptability, and c is the current attempt to improve our epistemic conditions. The evidence at stage n depends immediately (as opposed to historically) upon the current belief system B and the current experimental set-up c, but is presumably not directly dependent upon the current α (though α may be a factor underlying *our* choice of B).[66] Accordingly, let $e_\psi(B, c)$ be the datum received from ψ in response to B and c.

---

[60][Putnam 90], p. 50.

[61] *Ibid.*, p. 49.

[62] *Ibid.*, p. 50.

[63] *Ibid.*, pp. 55-56.

[64][Putnam 89], p. xvii.

[65]"Consider the sentence 'There is a chair in my office right now.' Under sufficiently good epistemic conditions any normal person could verify this, where sufficiently good epistemic conditions might, for example, consist in one's having good vision, being in my office now with the light on, not having taken a hallucinogenic agent, etc." [Putnam 89], p. xvii.

[66]Notice that the statement "evidence depends upon α" is metaphysically ambiguous. On the one hand, it could mean that we have no control over how our choice of α affects the evidence, in which case the dependency would belong to the world-in-itself. On the other hand, it could mean that we happen to use α in our process for generating B, and that it is the dependency between B

Truth at stage n depends immediately upon α (as in our discussion of $truth1_\alpha$ and $truth2_\alpha$) and immediately upon K (insofar as the data stream that would be produced under increasingly idealized choices of c will depend immediately upon K). Truth does not depend upon the actual c chosen by the scientist, however. It depends only on what the current α would do if c *were*, counterfactually, to be progressively idealized for eternity. Let P be the set of all programs or procedures for enumerating an infinite sequence χ of epistemic conditions, and let $p \in P$. Let p[n] be the epistemic condition output by p on input n. Let Improve(p) be the hypothesis "p enumerates an infinite sequence of improving epistemic conditions". Assume that Improve(p) $\in$ H. Let $f_{\psi(B, \alpha)}$ be the truth assignment for hypotheses in ψ relative to <B, α, c> (recall that truth does not depend upon the *actual* choice of c). Then a rough reconstruction (along the lines of truth1)[67] of Putnam's transcendental knowledge is:

(a) $f_{\psi(B, \alpha)}(h) = 1 \Leftrightarrow$
$\forall p \in P$, if $f_{\psi(B, \alpha)}(\text{Improve}(p)) = 1$,
then $\exists n\ \forall m \geq n$ such that $\alpha(h, \langle e_\psi(B, p[0]), \ldots, e_\psi(B, p[m])\rangle) > 0.5$

(b) $f_{\psi(B, \alpha)}(h) = 0 \Leftrightarrow$
$\forall p \in P$, if $f_{\psi(B, \alpha)}(\text{Improve}(p)) = 1$,
then $\exists n\ \forall m \geq n$ such that $\alpha(h, \langle e_\psi(B, p[0]), \ldots, e_\psi(B, p[m])\rangle) \leq 0.5$

We take Putnam at his word that truth is stable, rational acceptability under increasingly idealized epistemic conditions, not stable rational acceptability under conditions we *believe* to be ideal or that we would be *justified in asserting* to be ideal. Putnam makes a great point of distinguishing internal realism from redundancy theories of truth and from Dummett's identification of truth with actual verification. But then, the truth about claims concerning epistemic conditions should *also* be internal realist truth, not actual warranted assertibility or willingness to believe.[68] Thus, we run

---

and the current evidence that is out of our control. Then the dependency in question results from our own choices (i.e. we could choose B *without* using α), and is not properly considered to reflect the structure of the world-in-itself. The matter can be subtle. Have we *chosen* α if it has no role in shaping B? Is B *really* our belief system if α had no role in shaping it? Answers to such questions can incline us to cast dependencies either to the side of the world or to the side of the scientist.

[67]It is a simple matter to rework the account along the lines of truth2.

[68]Putnam seems to fudge this point in [Putnam 89], p. xvii. Picking up from the quote in footnote 65, we have:

> How do I know these are better conditions for this sort of judgment than conditions under which one does not have very good vision, or in which one is looking into the

the claim Improve(p) back through internal realist truth in the antecedents of (a) and (b). Since internal realist truth is relative to B, we still save (P5). But true to internal realism, idealization is more, on our account, than what we believe or are currently justified in believing about idealization.

Putnam insists that he does not intend internal realism to be a definition of truth. As a matter of fact, this reservation is warranted, since different truth assignments are consistent with (a) and (b) under a fixed choice of \|f, B and a.[69] The problem is that the recursive call to $f_v(B_f a)^{in}$ the antecedents of (a) and (b) may or may not "bottom out", depending on the structure of y and of a. Since examples can be found in which the alternative truth assignments are entirely symmetrical (up to permutation of O's and Vs), it is hard to see what further restrictions of an internal realist sort could be added to (a) and (b) to pick out one of them in a non-arbitrary manner.

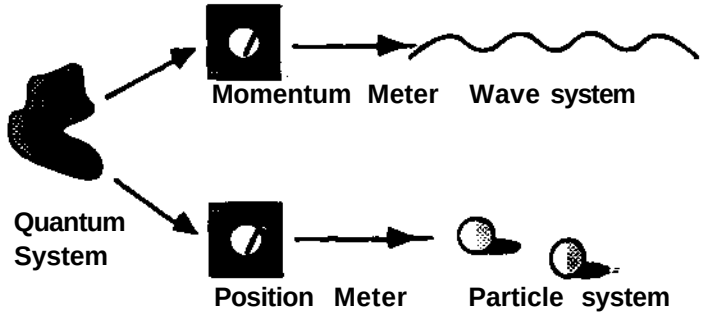
Relativistic dependencies can be much more mundane than in these philosophical examples. In the case of the economic prognosticator who crashes the market, the truth-dependency is entirely causal, but the logic of the situation is the same: truth depends on what the prognosticator predicts. The power of suggestion of a psychoanalyst on the personality under study provides another such example. In a quantum system, the truth can depend essentially upon the scientist's free choice of an experimental act. Thus, believers in quantum mechanics think they know quite a bit about the structures of possible worlds-in-themselves.

---

room through a telescope from a great distance, or conditions in which one has takenLSD? Partly by knowing how talk of this sort operates (what the 'language game[1] is, in Wittensteing's sense), and partly by having a lot of empirical information. There is no single genera! rule or universal method for knowing what conditions are better or worse for justifying an arbitrary empirical judgment.

Here, Putnam shifts from the internal realist truth about increasing idealization to what we know about increasing idealization. Those are two very different things.

[69]To see that multiple truth assignments are consistent with (a) and (b) for a fixed choice of y, a, and B, suppose that P = {po, Pi}, where po[n] = 2n and pi[n] = 2n+1. Let B be a fixed belief system. Let e^B, x) = 1 if x is odd, and 0 otherwise. Thus, $<e_v$(B, Po[O]),.... e\j/(B, Po[^{m}l). •••>$^{is\ a}$ sequence of O's and $<e_v$(B, pi[0]), ... e^B, pi[m]), ...> is a sequence of Vs. Let a(Improve(pi), e) = 1 if e contains only Vs, and let a(Improve(pi), e) = 0 otherwise. Dually, let a(Improve(po). e) « 1 if e contains only O's and let a(Improve(po), e) = 0 otherwise. Define S(p\ p) <=> 3n Vm ⩾ n such that a(Improve(p^{l}), $<e_{yi/}$(B, p[0]), ..., e^B, p[m])>) > 0.5. Then we have S(pi, Pi), S(p2» P2). -iS(pi, P2), and —iS(P2, Pi). Thus, (a) is satisfied when Improve(po) is true (relative to B, a) and Improve(pi) is not, and when improve(pi) is true (relative to B, a) and Improve(po) is not.



Indeed, it can be argued that Bohr conceived of **quantum mechanics as a relativistic system in just** this **manner.**[70] **Relativity** theory can be **viewed as a** relativistic **system in which the free decision to** ignite rocket fuel is the act relevant to the **truth about simultaneity. And to make the matter as** concrete as possible, whenever the hypothesis under investigation is contingent, the scientist's acts can help determine whether or not it is true (e.g. the scientist can **murder the last non-black** raven). So ordinary experimental investigation of contingent hypotheses can **also pose a** relativistic problem in the functional sense. In each of these cases, **we take our** "transcendental knowledge" K about functional wortds-in-themselves (truth dependencies) to be quite strong.

III.C. Convergent Relativism

The philosophy of science is still reeling from the relativistic and holistic blows levelled by Quine and by Kuhn. According to to the usual story, scientific objectivity, and hence scientific rationality has been undercut. Since the evidence depends upon the background and pet theories of the perceiver, different scientists following the same method can end up with different, justified conclusions.

Philosophy journals are stuffed with articles that accept the implication from non-trivial relativism to methodological nihilism, but which deny the antecedent. The story is that relativism happens to be benign in real historical cases so rationality is saved by accident.[71] On the other side are the anarchists eager to exploit the barest hint of relativity to reject all prospects for general methodological norms and principles. But neither party challenges the *inference* from relativity to "anything goes".[7]* Our strategy is just the opposite: to concede the possibility of strong

[70] [Faye91], p. 194.

[71] E.g. [Toretti 90], p. 81.

[72] Isaac Levi is a notable exception to this rule. His views are quite different from ours, however.

versions of relativism and incommensurability while undermining the usual inference from relativity to methodological nihilism.

The first question is why the inference from relativism to "anything goes" is so popular. The logical positivists, like the phenomenalists and empiricists before them, started out as reliabilists. The point of formulating analytic reduction relations was to skirt the global underdetermination[73] of physical theory by data, for underdetermination precludes the reliability even of a god who can see all the data that will ever arrive in the future. But once the reductionist program was pursued in earnest, it became clear that the highly ambitious aim of "logically" reducing all knowledge to sense data was not going to pan out. The reductionist program had too much momentum to be halted by the mere rejection of its fundamental motivation, however. Instead, the language of sense data was exchanged for the *fallible* language of macroscopic physical objects in order to make reductioneasier, and reliability was replaced with a new rationale for reductionism: *intersubjective agreement.*

> The condition thus imposed upon the observational vocabulary of science is of a pragmatic character; it demands that each term included in that vocabulary be of such a kind that under suitable conditions, different observers can, by means of direct observation, arrive at a high degree of certainty whether the term applies to a given situation. *** That human beings are capable of developing observational vocabularies that satisfy the given requirement is a fortunate circumstance: without it, science as an *intersubjective enterprise* would be impossible.[74]

Notice the striking absence of concern that science be a *reliable* enterprise in the last sentence of this passage. Kuhn and Quine undercut this *particular* rationale for reductionism. The difficulty wasn't relativism *per se.* The positivists were relativists, but the relativism was contrived to be *conjecture-independent* since meaning was pinned to conventionally selected, analytic meaning postulates independent of the scientist's election to believe this or that contingent, empirical hypothesis. Analytic truths were taken to "frame" the language of science, and inductive methods were conceived as operating *within* this framework, leaving it undisturbed by their changing conjectures. Quine's holism made truth depend upon all beliefs, and Kuhn's evidential relativism made data depend upon paradigms partly constituted by shared beliefs and assumptions. Thus both proposals imply conjecture dependency.[75]

---

[73]A hypothesis is ***globally underdetermined*** just in case the infinite data stream ε can be the same whether or not the hypothesis is true.

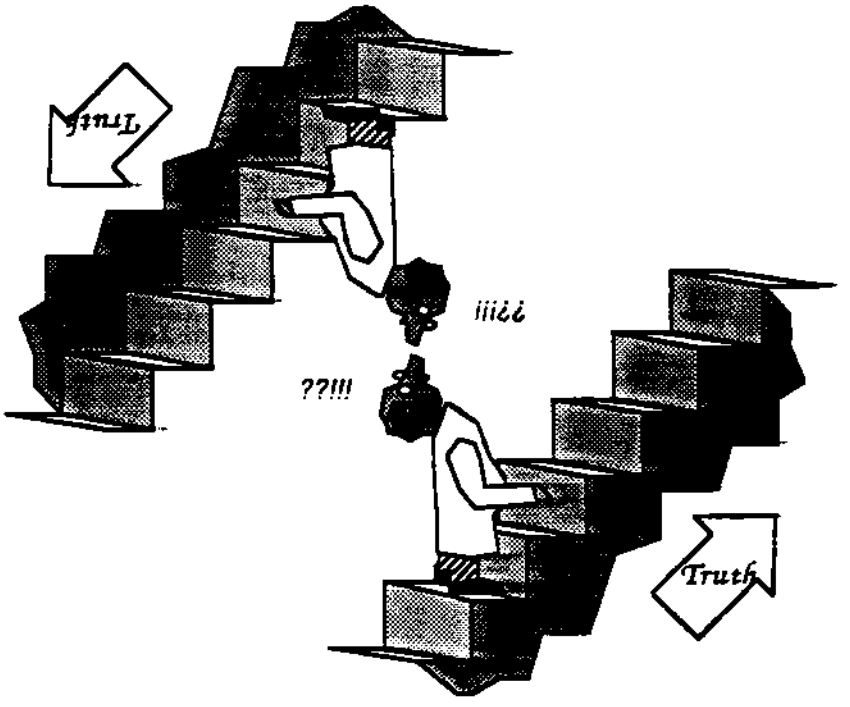
[74][Hempel 65], p. 127, n. 10, our emphasis.

[75]Rorty seems to agree with this interpretation:

In a conjecture-dependent wortd-in-itself, a shared method of hypothesis assessment and shared "conventions" do not guarantee shared conclusions, even in **the long run. For suppose we were** to follow something like hypothetico-deductive method or its improvement, the bumping pointer method (introduced in Section I above). If our hypothesis **enumerations differ ever** so slightly, or if our data streams differ in the slightest respect (e.g. from **different optical perspectives** on **the** same thing) we may **end** up with different conclusions in the short run. **But in** light of conjecture-dependent data, these slight differences of conjecture can lead **to slight differences** of **data,** which will lead to greater differences in conjecture, and so forth, until **truth, evidence,** and **hence** "justified conclusions" become very different for the two former colleagues. If precisely the same discovery methods **were** used, and precisely the same data were collected through time, even conjecture dependency would not undermine the late-positivistic aim of intersubjective science, but these exacting conditions are too fragile to carry much philosophical significance.

So relativism is a special problem for confirmation theorists who ground scientific rationality on inter-subjective agreement. Is it a problem also for the limiting reliabilist? It may seem that the proposal to unite relativism and convergent reliabilism is a non-starter, for if truth changes through time, convergence to *the* truth makes no sense. But we can still attempt to isolate the circumstances under which a scientist can converge to his *own* truth, even though others may converge to their truths which differ from his. It cannot be objected that science is concerned with real truth rather than truth-for-me, since, under the hypothesis of relativism, there is no truth but truth-for-me. The real losers in relativistic science are not those who disagree with one another. The losers are those who are fooled for eternity according to their own, *internal* versions of falsehood. The proposal that scientific methods converge reliably to the relative truth when truth is not unique will be referred to as *convergent relativism.* Thus, we can unify Putnam's early, limiting reliabilism with his recent, moderate relativism.

To say that we have to assign referents to terms and truth-values to sentences in the light of our best notions of what there is in the world is a platitude. To say that truth and reference are relative to a conceptual scheme" sounds as if it were saying something more than this, but it is not. as long as "our conceptual scheme" is taken as simply a reference to what we believe now - the collection of views which make up our present-day culture. This is all that any argument offered by Quine, Sellars, Kuhn. or Feyerabend would license one to mean by "conceptual scheme". [Rorty 79], p. 276.



There are perhaps two main reasons why convergent relativism has not captured the imagination of philosophers. First, the goal seems too difficult. 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 conceptual frameworks to find one that is suitable. This observation is doubly flawed, however. It is by no means always trivial to find the truth in a fixed system, as the many negative results presented in the first section of this paper 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. This sort of stratagem is familiar from the old discussions of *conventionalism.*

In light of these comments, getting to the relative truth may now appear too *easy.* If truth depends upon you, then what is the point of inquiry? Just *make* your present beliefs true by an act of will, and be done with it! But this triviality holds only in the most extremal forms of coherentism, and does not follow for moderate relativism. If truth depends not only upon the investigator but also upon some independent reality over which the investigator has no control, then the scientist *may not know just how tmth actually depends upon what he does.* This is the case in each of the intuitive examples of relativism discussed above. Under such circumstances, finding the truth about a given hypothesis may be difficult or even impossible.

III.D. Reliable Relativistic Discovery in the Limit

A relativistic hypothesis generator γ is just like its non-relativistic counterpart, except that it produces both a hypothesis h and some act $a \in A$ on the basis of the data provided, rather than just a hypothesis h.[76] Assuming that the world-in-itself ψ is evidentially local, the interaction between the world-in-itself ψ and method γ then determines an infinite play sequence ***play***(γ, ψ), where γ produces its act and conjecture on the empty data sequence, ψ responds with a truth assignment and the next datum, and so forth, forever. For each position n, play$(\gamma, \psi)_n$ denotes some tuple $\langle a_n, h_n, x_n, f_n \rangle$ where $a_n \in A$ is the method γ's act, $h_n \in H$ is γ's conjecture, $x_n \in E$ is the current datum returned by ψ and f_n is the current truth assignment returned by ψ. Then we define sequences $A(\gamma, \psi)$, $H(\gamma, \psi)$, $E(\gamma, \psi)$, $C(\gamma, \psi)$ so that for each n, $A(\gamma, \psi)_n = a_n$, $H(\gamma, \psi)_n = h_n$, $E(\gamma, \psi)_n = x_n$ and $C(\gamma, \psi)_n = f_n$. So the model reflects the "hermeneutic circle" of relativism, in which the conceptual scheme depends upon the evidence and the evidence depends upon the conceptual scheme. But here the circle is bent into a spiral through time, so we have mutual dependency without circularity.[77]

Relativistic discovery admits of various different senses of success. Let γ be a relativistic hypothesis generator and let K be a set of evidentially local worlds-in-themselves. One sense of success requires only that after some time, the conjecture of γ is correct with respect to the acts of γ.

> γ ***makes reliable discoveries in the limit given*** K $\Leftrightarrow$
> *for each* world-in-itself ψ in K
> *there is* a time n such that
> *for each* later time m
> the conjecture produced by γ in ψ at time m is correct at time m with respect to all the acts and conjectures ever performed or to be performed by γ in response to ψ.
> (i.e. $C(\gamma, \psi)_m(H(\gamma, \psi)_m) = 1$).

This definition is general enough to apply at once to local and to global worlds-in-themselves. It is also quite liberal. Reliable relativistic discovery permits γ to vacillate forever both in its conjectured hypothesis and in its other semantically relevant acts. Thus it is permitted for γ to initiate "scientific

---

[76] I.e. $\gamma: E^* \rightarrow (A \times H)$.

[77] If the world-in-itself is not evidentially local, then the circle is of a more interesting sort that must be handled with more powerful machinery than we will develop here.

revolutions" infinitely often, so long as after some time the hypothesis produced is correct relative to the conceptual scheme operative when the hypothesis is conjectured. Strange? Yes! But the possibility for such success is implicit in the standard relativistic philosophies, whether or not it is advertised. And when these relativisms are taken seriously (as we take them here), it is hard to say what would be wrong with this sort of success from a limiting reliabilist point of view.

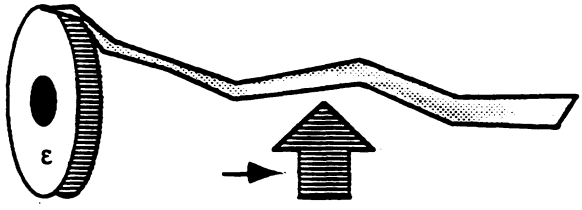
Those who are subject to semantic vertigo at the prospect of an eternity of bounces between worlds or conjectures are free to impose extra, syntactic stability conditions on conjectures and on other semantically relevant acts in various combinations. ***Stable*** discovery requires that eventually both the acts and the conjectures stabilize to some specific choice of <h, a>.

III.E. Anything Goes

P.K. Feyerabend's "anarchist" response to meaning variance and relativism in science has been summarized in the slogan *anything goes*. As we have seen, it is a limiting reliabilist commonplace that different methods are better in different inductive settings, so if "anything goes" means only that we should not force a fixed method on everybody, we don't need *relativism* to prove it. And if "anything goes" means that there is no universal architecture for relativistic discovery, it is mistaken both for relativistic and for naive realist science, as we shall see in the next section.

Perhaps "anything goes" means that it is wrong to force a particular conceptual scheme on the scientist, or to require him to stabilize to a particular such scheme. We may think of each shift in correctness and evidence due to the agency of the scientist as a ***scientific*** or ***conceptual revolution***. Stable discovery fits with the intuition that each scientific revolution is a clean break with the past, so that convergent success can succeed only within a fixed conceptual scheme. Thus, after some time science should fix upon a particular scheme and diligently seek the truth within it. Unstable discovery embodies a more anarchistic attitude, in that it countenances an infinite number of conceptual revolutions. This kind of anarchism is in fact vindicated from the point of view of convergent relativism, because there are relativistic discovery problems that cannot be solved in the stable sense, but that can be solved in the unstable (anything goes) sense. For a trivial example, suppose that $K = \{\psi\}$ where ψ is semantically spontaneous so that at even stages of inquiry the only correct hypothesis is h and at odd stages of inquiry the only correct hypothesis is h', no matter what the scientist does. It is impossible to stabilize to a true hypothesis in ψ, even in the limit, but the trivial method that conjectures h', h, h', h, ... for eternity, irrespective of the evidence, succeeds in the unstable sense.

The preceding example made stabilization to a correct hypothesis impossible for purely semantic reasons. A more interesting example would be one in which stabilization is semantically possible, but reliable stabilization to a particular scheme-conjecture pair is not possible for properly epistemic reasons, because stabilization would prevent the scientist from seeing important data. To construct such a case, suppose the scientist's only semantically relevant act is to adopt one of two conceptual schemes, 1 or 0. Let ψ_1 be a world-itself in which h is true exactly when the current scheme is 1 and in which the current datum is 0 no matter what the scientist does. Let ε be a fixed data stream in which only 0's or 1's occur. In world ψ_2^ε, h is correct exactly when the current scheme is 1, and the data is produced as follows: a pointer is initialized to ε_0, and each time conceptual scheme 1 is visited, the next entry in ε is presented as data. Whenever conceptual scheme 0 is visited, however, datum 0 is returned.



Increment pointer and return datum pointed to when conceptual scheme 1 is visited.

Else return datum 0.

In world ψ_3^ε we have the same situation, but the role of schemes 1 and 0 is reversed. Thus, data is drawn by the same pointer mechanism from ε in scheme 0, datum 0 is always returned in scheme 1, and h is correct in scheme 0 but not in scheme1. Finally suppose that in each of the worlds under consideration, $\neg$h is correct whenever h is not correct, and no other hypotheses are correct under any circumstances. Suppose we know in advance that the actual world-in-itself is either ψ_1, ψ_2^ε or ψ_3^ε, where it is certain that infinitely many 1's occur in ε. Thus, $K_0 = \{\psi_1, \psi_2^\varepsilon, \psi_3^\varepsilon$: infinitely many 1's occur in $\varepsilon\}$.

Intuitively, the dilemma posed by the problem is that if the scientist is actually in ψ_1 and stabilizes the truth value of h after some time, say to 1, then he must eventually settle into performing act 1 forever. But then if the scientist had really been in ψ_3, he might never have seen a 1 in the data prior to deciding to settle down into act 1, so he would never have discovered that in fact the truth value of h is 0 under act 1. This dilemma can be turned into a rigorous, relativistic version of Putnam's diagonal argument:

Fact 11.1: Effective, relativistic discovery is possible in the limit over K_0, but stable, relativistic discovery is not possible over K_0 even by an ideal method.

Proof: The effective scientist ywho succeeds in the unstable sense follows this strategy: He flip-flops between schemes 1 and 0 forever, producing hypothesis h in scheme 1 and hypothesis -»h in scheme 0 so long as only O's are seen in the data. As soon as a 1 is seen, the scientist remarks which scheme a it occurred under and stabilizes to that scheme and the corresponding hypothesis (h if a = 1, and -ih if a = 0).

In v i , y alternates forever between schemes 0 and 1 and hypotheses h, —»h, but ys conjecture is always correct. In H4 eventually datum 1 appears under scheme 1 since y visits scheme 1 at odd stages until a 1 is seen. Then y stabilizes correctly to <1, h>, which is correct tor v l - By a similar argument, *y* eventually sees a 1 under scheme 0 in V3[5], and thus stabilizes to act 1 and truth value 0, which is correct for V3[5].

Now the diagonal argument. Suppose for *reductio* that *y* stably succeeds over Ko- Then either (a) y converges to <1, h> in yi or (b) Y converges to <0, —ih> in \yi. Consider case (a). Let n be the time at which Y converges to <1, h> in y i . Let k be the number of times scheme 1 is visited by Y in \|f 1 up to stage n. Define E SO that e has O's up to k and 1 thereafter. Thus V3[5] € K<>. Moreover, the data seen by *y* in *vl* is exactly the same up to stage n as the data seen in yi by y. Since Y converges to scheme 1 at stage n, and since V3[5] produces the same data under scheme 1 that vi does (namely, all O's), *y* converges to <1, h> in V3[5], which is incorrect. The cases for -rh are similar. 1^

III.F. Relativistic Transcendental Deductions

Kant proposed both a moderate, metaphysical relativism and the notion of transcendental deductions. Since he thought that the human conceptual scheme is fixed, his transcendental deductions were not relativistic: they were directed at the character of this fixed scheme. But convergent relativism raises the prospect for rigorous, properly relativistic, transcendental deductions. And just as in the realist case, such results can be used to establish completeness for relativistic discovery architectures. The existence of such architectures tor reliable, relativistic discovery overturns the quick inference from relativism to methodological nihilism.

For a simple example, when truth depends upon the current act of the scientist but evidence does not, we have an easy reduction to the situation in Section I. When evidence is spontaneous, each world-in-itself is characterized by a unique data stream e_v (the one generated in successive times by y, regardless of what y does). Since y is semantically immediate, we may

let fy(<h, a>) denote the correctness assignment produced by y at any time in light of semantic **act** <h, a>. Now the characterization condition may be stated as

Theorem III.F.1: Suppose the worlds-in-themselves in K are
(a) evidentially spontaneous and
(b) semanticaily immediate

Then correct hypotheses are effectively stably discoverable in the limit given K <=>

there is some R^f **c** (E° X H), S **c** (H X A) s.t.

(0) V\|f e K_t <h, a> G H X A, R'(e_v, h) =* $f_{v(<h, a>)}$(h) = 1 and

(1) S is recursively enumerable and

(2) S covers K according to R^f and

(3) R'e3H\K]2

Condition (0) replaces the realist condition R' Q R in Theorem I.F.2. It is a relativistic way of saying that the act <h, a> is "correct[11] if R^1 says it is. Conditions (1), (2) and (3) mimic the conditions with the same numbers in Theorem I.F.2, where we take the pair <h, a> to be "correct" or "incorrect" **for** a world-in-itself y rather than for a particular data stream e. The complete architecture **for** discovery will be a minor variant of the bumping pointer architecture that takes these minor alterations into account.

We have also characterized a special class of problems in which data also depends upon the current act of the scientist [Kelly and Glymour 92]. When the scientist's history is relevant to the data presented (i.e. the world-in-itself can "remember[11]) the situation becomes more complex. Such historical dependencies can make success harder to achieve, because some desired datum may be forever unobservable for a scientist who has elected to perform some sequence a of acts in his past.[78] These new difficulties are hardly unique to philosophical relativism, however, since they are familiar features of experimental science in general. In classical mechanics, it is usually assumed that we can perform essential experimental acts independently and get the same information, but in relativity theory[79] and in quantum mechanics[80], this assumption is challenged. More concretely, experimental compatibility assumptions clearly fail in historical studies like

[78][Kelly 92a], chapter 11.

[79]The idea is that irregular features of the global topology of space-time can become forever causally disconnected from observers on some world lines [Malament 77].

[80]!n some developments of quantum logic, *experimental compatibility* \s a central concept. [Cohen 89], p. 25.

archaeology, in which a pillaged or mis-managed site can shut the book on an ancient culture for eternity. Global relativism poses further technical challenges to the transcendental logician, issues we leave for future study.

III.G. Relativistic Theory-Building

So far, we have viewed the scientist as producing a particular theory and other semantically relevant acts that may affect the correctness of this theory. Perhaps a more realistic situation portrays the scientist as building up a theory in response to data, sometimes removing hypotheses and sometimes adding them, with the intention of choosing new hypotheses so that the old ones remain true. In such a situation it is natural to suppose that a scientist is committed to the logical consequences of his conjectures, so we need to introduce a relativistic version of semantic entailment. Intuitively, h entails h' relative to method γ and time n just in case the correctness of h relative to γ at n implies the correctness of h' relative to γ at n.

$$K, h \models_{\gamma, n} h' \Leftrightarrow \forall \psi \in K,\ C(\gamma, \psi)_n(h) = 1 \Rightarrow C(\gamma, \psi)_n(h') = 1.$$

The logical structure of relative entailment can vary radically from one time to the next. Nonetheless, we might hope that for each h, there is a time after which h is entailed (with respect to the sense of entailment operative at the time) if and only if h is correct (with respect to the sense of correctness operative at the time). Thus:

γ ***non-uniformly discovers the complete truth given*** $K \Leftrightarrow$
$$\forall \psi \in K\ \forall h \in H\ \exists n\ \forall m \geq n\ C(\gamma, \psi)_m(h) = 1 \Leftrightarrow K, H(\gamma, \psi)_m \models_{\gamma, m} h.$$

Non-uniform theory construction is closer to the diachronic image of inquiry operative in many philosophy of science discussions.[81] In realist settings, truth and well-formedness are naively fixed, and truth stays put while science tries its best to home in on it in light of increasing data.[82] In the picture, T is the set of all true hypotheses, wf is the set of all well-formed strings, and S is the set of hypotheses entailed by γ's current conjecture:

---

[81] For example, consider the diachronic model in [Lakatos 70].

[82] For a non-relativistic development of the logic of non-uniform or "incremental" theory construction, c.f. [Kelly and Glymour 89].

naive realist convergence

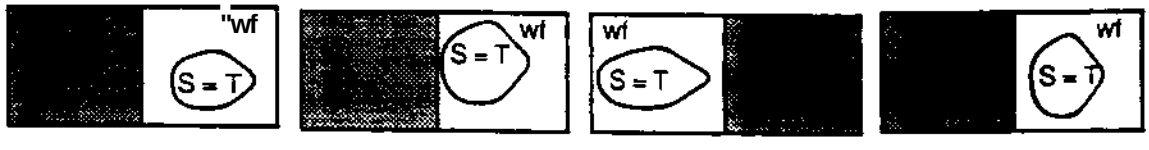
This picture is very little changed in the case of positivism. We must add to the picture the set C of all conventionally selected meaning postulates that tie down the language. Assuming with the positivists that inquiry is conjecture-independent (inductive methods doni fiddle with analytic truth) then the picture is the same, except that it is now guaranteed that the meaning postulates are well-formed and true.

positivisticconvergence

If the meaning postulates C are changed, however, then there is a sattative break with the past. Positivists didn't entertain the possibility of convergent success *through* changing conventions, and this is what left them vulnerable to the possibility of conjecture dependency raised by Kuhn and Quine, as we have seen.

Proponents of conjecture dependency often assume that when truth depends on the scientist's conjectures, it meets them half-way to make science easier. Kuhn and Hanson speak of theory-ladenness, as though the theories we entertain *color* truth and the data so that truth and evidence always meet us half way. Radical coherentists go further, and assume that truth unshakably chases us around, so long as certain unspecified standards of coherence are observed.

Coherentism (positive conjecture dependency)



But once conjecture dependency is out of the bag, we see we see that it is just as conceivable for truth to run *away* from us, in the sense that our conjectured theories are false *because* we conjecture them. Thus, a hypothesis could remain true as the scientist homes in on it, only to melt

spitefully into falsehood (or even worse, into *nonsense)* the minute he comes to believe it, only to become true again after it is rejected. We refer such wortds-in-themselves as ***Sisyphusian.***[33]

Sisyphusian relativism (negative conjecture dependency)

Sisyphusian relativism is not entirely a product of our malicious imaginations. If social constructivists are correct and truth is nothing more than the relevant community's assent to one's beliefs, then Sisyphusian relativism is a distinct possibility, for an offended community could out of spite reject, and therefore falsify, every one of the offender's announced beliefs.

Coherentism, the philosopher's ultimate refuge from skepticism, is not immune from these difficulties. Putnam has correctly remarked that coherentism is a dangerous position for a relativist, for it is vacuous if coherence is nothing at alt and if coherence is the same for everybody then it is naively objectivistic.[84] But if coherentism steers a moderate relativist course between these two extremes, so that coherence, itself, depends to some extent on the acts and history of the scientist, then Sisyphusian skepticism looms once again, for it may happen that what we believe is incoherent *because* we believe it.[85]

Sisyphusian relativism is much worse than the original inductive skepticism that relativism (in its Kantian and positivistic incarnations) was summoned to defeat. Inductive skepticism tells us that belief might happen to be wrong for ail we know. Relativistic skepticism tells us that belief might be self-defeating for all we know: a sorrier situation. All that stands in the way of such possibilities are the philosophers' claims to transcendental knowledge about how truth and evidence can possibly depend upon our acts. All that has happened is that ordinary scientific uncertainty has been replaced with a far more virulent transcendental uncertainty, and the material dogmas of science have been replaced with the metaphysical dogmas of the philosophers.

---

[83]**Sisyphusian relativism is covered by Proposition lll.D.1 in the case of semantic immediacy, for in that case it will be impossible to satisfy condition (2).**

[84]**[Putnam 90], p. 123.**

[85]**This may be seen in a precise way if we define relative satisfiability along the lines of our definition of relativistic entailment. E.g., say that h is satisfiable relative to *y* at time n just in case h is correct relative to *y* at n in some possible world-in-itself.**

But Pandora's box is hard to shut once it has been opened. Perhaps it is vain to hope for a convincing, *a priori* argument against undesirable relativistic possibilities. But this doesn't leave us entirely helpless. At least we can reason backwards, by means of transcendental deductions, to determine what we would have to know about the world-in-itself for reliable, relativistic theory-building to be possible. If we are doomed to transcendental dogmatism, we can at least choose our dogmas to be as weak as they can be conditional on preserving some desired sense of reliability in scientific inquiry. In this manner, limiting reliabilist transcendental deductions can be used to constrain metaphysical theorizing. For example, since Theorem III.F.1 characterizes convergent success over conjecture-dependent worlds, it must exclude the possibilitly of Sisyphusian worlds. Inspection of the theorem will reveal that condition (0) is violated for every choice of R' when ψ is Sisyphusian.

III.H. Relativism, Logic, and Historicism

Our aim in this section of the paper has been to extend Putnam's early, limiting reliabilism to settings in which theory laden data and incommensurability run rampant. We have seen that there are transcendental deductions for convergent relativism that are quite parallel to those we obtained in the naive realist settings in the first section of this paper. Despite this, it is interesting to observe that this basic strategy has has already been anticipated--- and *refuted*--- some years ago by the *cognoscenti* of relativistic nihilism.

> The notion that it would be all right to relativize sameness of meaning, objectivity, and truth to a conceptual scheme, as long as there were some criteria for knowing when and why it was rational to adopt a new conceptual scheme, was briefly tempting. For now the philosopher, the guardian of rationality, became the man who told you when you could start meaning something different, rather than the man who told you what you meant.

Recall that a relativistic hypothesis generator is exactly a method that tells you what to conjecture and when to mean something different. Rorty claims that any such proposal must somehow presuppose meaning invariance:

> But this attempt to retain the philosopher's traditional role was doomed. All the Quinean reasons why he could not do the one were also reasons why he could not do the other. * * * ...[As] soon as it was admitted that "empirical considerations" ... incited but did not require "conceptual change"... the division of labor between the philosopher and the historian no longer made sense. Once one said that it was rational to to abandon the Aristotelian conceptual scheme as a result of this or that

> discovery, then "change of meaning" or "shift in conceptual scheme" meant nothing more than "shift in especially central beliefs". The historian can make the shift from old scheme to the new intelligible.... There is nothing the philosopher can add to what the historian has already done to show that this intelligible and plausible course is a "rational" one. Without what Feyerabend called "meaning invariance," there is no special method (meaning-analysis) which the philosopher can apply.[86]

But as a matter of fact, nothing in our presentation presupposes invariance in meaning or truth of the scientist's conjectures; much less that such conjectures are translatable across conceptual revolutions. Stabilizing to one's own version of the truth does not require that past theories make sense or can be translated into one's current point of view. It does require knowledge of what past scientists accepted as evidence and what kinds of theories they conjectured when that evidence was accepted; but these are just the sorts of things that Rorty claims the historian can teach us.

Rorty focuses on the question whether a *particular* shift in conceptual scheme is plausible or intelligible. But when we shift our attention to the reliability of general strategies for generating theories and changing meaning through time, the prospect arises for a genuinely logical analysis along the lines indicated above. Such analyses are something that a philosopher (or computer scientist or historian or *anyone who pleases*) can do that amassing loads of historical case studies and explaining them informally would not do. We aren't saying that one of these sources of insight into the workings of science is better than the other. Rorty is.

Perhaps the relativistic nihilist will respond with his ultimate weapon, namely, that even his relativism is relative. Since our logical investigation of relativistic induction assumes fixed, transcendental knowledge K, it assumes an Archimedean point which the nihilist is happy to pull out from under himself. We have two responses.

First, this ***meta*** relativism is a bold, transcendental thesis that is hardly entailed by the sorts of historical anecdotes that historicists like Kuhn have proposed as evidence for their of semantically and evidentially local versions of relativism. All that is indicated by such cases is (a) some dependence of truth and evidence upon one's history and (b) a putative lack of translatability between paradigms. The scientific historian aspires, in fact, to *tell us* when revolutions occurred, which theories are incommensurable, and why certain shifts of meaning occurred. As long as the

[86][Rorty 79], pp. 272-273.

relativist can tell us about how the relativistic dependencies work, our apparatus for convergent relativism will apply.

Second, we may view the pulling out of the rug as the first step up a "hierarchy" of relativisms analogous to Tarski's hierarchy in the conventional theory of truth. As we move up this hierarchy, the world-in-itself may depend upon acts, that dependency may depend upon acts, etc., to any ordinal level. We expect that our convergent relativist gambit will apply, with a corresponding increase in subtlety and complexity, to fixed background knowledge concerning dependencies at any such level. If the nihilist insists that his *ultimate* relativism diagonalizes across each of *these* levels, then we no longer have a response, but neither do we know what is being asserted. Ultimate relativism, the relativism beyond all relativisms, belongs to the battling realm of the greatest ordinal, the set of all sets, the liar paradox, and the neoplatonic One. If it takes *this much* to shut down Putnam's limiting reliabilist methodology, then it stands in excellent company.

Acknowledgments

We are grateful to Teddy Seidenfeld and to Wilfried Sieg for a useful discussion concerning issues in Section II. We are also indebted to Jeff Paris, at the University of Manchester, for some very useful suggestions concerning an earlier draft.

Bibliography

Barzdin, J. M. and R. V. Freivalds (1972), "On the Prediction of General Recursive Functions[11], *Soviet Math. DokL,* 12, 1224-1228.

Blum, M. and L Blum (1975), "Toward a Mathematical Theory of **Inductive Inference", *Information and Control,* 28.**

Faye, J. (1991), *Niels Bohr His Heritage and Legacy,* **Dordrecht: Kluwer.**

Gold, E. M. (1965), "Limiting Recursion", *Journal of Symbolic Logic,* 30:1. **pp.** 27-48.

Hanson, N. R. (1958), *Patterns of Discovery,* Cambridge: Cambridge University Press.

Hempel, C. G. (1965), *Aspects cf Scientific Explanation,* New York: Macmillan.

Horwich, P. (1991), "On the Nature and Norms of Theoretical Commitment", *Philosophy of Science,* 58: 1.

Kant, I. (1950), *Prolegomena to any Future Metaphysics*, L. Beck, Trans., Indianapolis: Bobbs-Merrill.

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

Kelly, K. and C. Glymour (1992), "Inductive Inference from Theory-Laden Data", *Journal of Philosophical Logic*, forthcoming.

Kelly, K. (1992) "Learning Theory and Descriptive Set Theory", forthcoming, *Journal of Logic and Computation.*

Kelly, K. (1992a), *The Logic of Reliable Inquiry*, in preparation.

Kemeny, J. (1953) "The use of simplicity in induction", *Philosophical Review*, LXII, 391-408.

Kugel, P. (1977), "Induction, Pure and Simple", *Information and Control*, 33, pp. 276-336.

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

Lakatos, I. (1970), "Falsification and the Methodology of Scientific Research Programmes", in *Criticism and the Growth of Knowledge*, Lakatos and Musgrave, eds., Cambridge: Cambridge University Press.

Laudan, L. (1980), "Why Was the Logic of Discovery Abandoned?", in *Scientific Discovery, Logic, and Rationality*, T. Nickles, ed., Boston: D. Reidel.

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

Lucas, J. R. (1961), "Minds, Machines, and Goedel", *Philosophy.*

Lycan, W. (1988), *Judgement and Justification*, New York: Cambridge University Press.

Malament, D. (1977), "Observationally Indistinguishable Spacetimes", in *Foundations of Space-Time Theories*, vol. VIII, *Minnesota Studies in the Philosophy of Science*, J. Earman, C. Glymour, and J. Stachel, eds., Minneapolis: University of Minnesota Press.

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

Osherson, D., and S. Weinstein (1991), "A Universal Inductive Inference Machine", *Journal of Symbolic Logic*, 56:2.

Peirce, C. S. (1958), "Some Consequences of Four Incapacities", in *Charles S. Peirce: Selected Writings*, Philip P. Wiener, ed., New York: Dover.

Plato (1949), *Meno*, B. Jowett, trans., Indianapolis: Bobbs-Merrill.

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

Putnam, H. (1963), "'Degree of Confirmation' and Inductive Logic", in *The Philosophy of Rudolph Carnap*, A. Schilpp (ed), Lasalle, Illinois: Open Court. Reprinted in *Mathematics, Matter, and Method. Philosophical Papers*, Vol. I., Cambridge: Cambridge University Press (1979). All page citations refer to the more recent source.

Putnam, H. (1963a), "Probability and Confirmation", *The Voice of America, Forum Philosophy of Science,* 10, U.S. Information agency. Reprinted in *Mathematics, Matter, and Method, Philosophical Papers,* Vol. I., Cambridge: Cambridge University Press.

Putnam, H, (1974), "The Corroboration of Theories", in *The Philosophy of Kari Popper,* vol II., La Salle: Open Court.

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

Putnam, H. (1989), *Realism and Reason: Philosophical Papers,* Vol. III, Cambridge: Cambridge University Press.

Putnam, H. (1990), *Reason, Truth and History,* Cambridge: Cambridge University Press.

Quine (1951), "Two Dogmas of Empiricism", *Philosophical Review,* 60.

Reichenbach, H. (1938), *Experience and Prediction,* Chicago: University of Chicago Press.

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

Rogers, H. (1987), *The Theory of Recursive Functions and Effective Computability,* Cambridge: MIT Press.

Rorty, R. (1979), *Philosophy and the Mirror of Nature,* Cambridge: Princeton.

Savage, L. J. (1972), *The Foundations of Statistics,* New York: Dover.

Toretti, R. (1990), *Creative Understanding: Philosophical Reflections on Physics,* Chicago: University of Chicago Press.