(2)

REPORT DOCUMENTATION

AD-A255 279

Public reporting burden for this collection of information is estimated to average 1 hour per response, i... maintaining the data needed, and reviewing the collection of information. Send comments regarding this ... gathering and ... for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations ... ing suggestions ... the Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, DC 20... 32-4302, and to

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 1989	Unknown

4. TITLE AND SUBTITLE Convention, Confirmation, and Credibility	5. FUNDING NUMBERS DAAB10-86-C-0567
6. AUTHOR(S) Henry E. Kyburg, Jr.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Rochester Department of Philosophy Rochester, NY 14627	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army CECOM Signals Warfare Directorate Vint Hill Farms Station Warrenton, VA 22186-5100	10. SPONSORING/MONITORING AGENCY REPORT NUMBER 92-TRF-0047

DTIC ELECTE SEP 1 0 1992 S C D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT Statement A; Approved for public release; distribution unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words) There are good arguments for conventionalism in science, but that if they are pursued too far they leave us with empty bodies of knowledge. We must have grounds for preferring one convention to another, and furthermore, they must be epistemic grounds. Such grounds cannot be provided directly by obeservation. As many conventionalists have pointed out, observation is always subject to error. The notion of error, and hence of observationality, is relativized to a theory, but is determined, for a given theory, by our experience - by what happens to us. Thus even though thories are regarded as "conventional" or a priori, even though the notion of error that determines the content of the practical certainties of a theory is internal to the language of the theory, we can still have an objective measure of the degree to which one theory rather than another satisfies our desire to anticipate the future. This picture leaves certain puzzles to be resolved, but seems to offer a picture of scientific knowledge that is plausibly rooted in the empiricist tradition. It may be that the general framework of metalinguistic report at one extreme, and object language prediction at the other.

14. SUBJECT TERMS Artificial Intelligence, Data Fusion, Philosophy	15. NUMBER OF PAGES 18
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540 01-280-5500

Standard Form 298 (Rev. 2-89)

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines to meet optical scanning requirements.**

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	**PR** - Project
G - Grant	**TA** - Task
PE - Program Element	**WU**- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - DOD - Leave blank.
DOE - DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - NASA - Leave blank.
NTIS - NTIS - Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

Convention, Confirmation, and Credibility*

1. Convention

Conventionalism, particularly in the treatment of scientific theories, has a long history, though a rather unclear meaning. There may be important historical roots going arbitrarily far back, but the first widely familar modern conventionalist was Mach[1], who argued that Newton's second law, $f = m\,a$, was nothing but a disguised definition of "force", on the ground that we have no other way to measure force than by directly or indirectly measuring acceleration. If we measure a force in terms of the strain on a spring, for example, we are indirectly measuring force in terms of the stress on the spring, and we know the stress in terms of experiments in which the spring accellerates known masses (for example). How far beyond Newton's second law conventionalism extended for Mach is not clear, and not really to the point of our inquiry. It suffices for us that he took a statement that was clearly regarded as an empirical generalization, and argued that it should better be construed as a convention concerning one of the terms of discipline involved.

A view of convention as somewhat more pervasive within a single discipline is that of Poincare. Poincare took all of the axioms of geometry to be implicit definitions of "space" and thus conventional. He argued that whether you regard space to be

Euclidean, or Riemanian, or Lobachevskian, or some combination thereof, was not determined by "facts" about space, but was a matter of convention.[2] He also made a famous prediction that was falsified in the event (though this does not prevent his argument from having been a good one in its time): "Euclidean geometry is, and will remain, the most convenient..."[3] He argued this on the ground -- of interest to us in the latter part of this essay -- that the increased complexity of geometries of variable curvature would never pay for itself in increased simplicity in the physical theory itself. In any event, for Poincare it was not just the odd physical or mathematical law that was conventional in character, but all the assertions of a certain discipline often construed as empirical in character.

It was not merely geometry to which Poincare ascribed a conventional element. Mach remarks, "Poincare ... is right in calling the fundamental propositions of mechanics *conventions* which might very well have proven otherwise."[4]

It was Pierre Duhem, a physicist, who made the strongest claim regarding the conventionality of what most people think of as empirical theory near the turn of the century.[5] He distinguished between <u>emprical</u> laws and <u>theoretical</u> laws, and argued that while empirical laws (by which he meant such laws as the law of thermal expansion) could be confirmed or disconfirmed by experiment, any really interesting law could not, since testing it involved a whole bundle of theoretical physical laws that could in principle perfectly well be called into question.

It is not germain to our purpose to examine in detail the ideas of Mach, Poincare, and Duhem. That they had the idea that it could be philosophically respectable to regard scientific theory, the crowning glory of human knowledge, the pinnacle of human inquiry, as in some sense conventional is important and sufficient for the time being.

Their modern successor, Quine, is not altogether a successor, since his claim is not that certain statements of physics or mathematics should be regarded as conventional, but rather that the very distinction between what is conventional -- definitional, analytic -- and what is not, what is synthetic, what has empirical content -- is an illegitimate and unenlightening distinction. In what follows later we shall argue that this distinction is indeed useful and enlightening, but only in a rather stronger conventionalist framework.

In order to discuss the elements of convention in scientific theory, we should have a clear notion of what a convention is. As a first approximation, let us take the characterization offered by David Lewis in his book, *Convention*[6]. (Surely the author of a book should have a clear idea as to the meaning of its title!) Lewis writes,

> It is redundant to speak of an arbitrary convention. Any convention is arbitrary because there is an alternative regularity

> that could have been our convention instead. (p.70)

Let us take this notion of conventionality as a starting point, and see how far convention permeates our scientific knowledge. To give us a visual handle on what we are doing, suppose that the vision of Quine and Ullian[7] is roughly correct: Our knowledge of the world is like a web. At the periphery of the web are statements that are direct reports of experience -- occasion sentences. In the center of the web are statements representing the truths of logic and mathematics. In between are the statements representing the theories, laws, and generalizations of empirical science.

Quine, at least in *Mathematical Logic*[8], regarded first order logic as conventionally true. The principle of charity invoked in *Word and Object*[9] is also used to defend conventional two-valued, extensional, first order logic as the basic framework for human thought, or at least for human conversation. Even those who would like to see some other form of logic taken as basic rarely argue that ordinary first order logic is <u>false</u>; it is argued, rather, that some other logic might be "simpler", or less "unnatural" for certain purposes. The tradition that regards logic as being "without content" or analytic is an old one, and respectable. Let us not argue with it, since it surely "could have been otherwise" by making a judiciously different choice of logical primitives and axioms.

To keep our attention on the important part of our scientific knowledge, let us eliminate from our Quinean/Ullianian image of scientific knowledge, these strongly conventional elements. From the center of our web, that is, let us eliminate the propositions of logic and mathematics.

Now a well-known result of Craig[10] shows that if you distinguish, as all these people seem to, between observational and theoretical terms, you can find a constructive procedure by which you can re-axiomatize your scientific theories in such a way that the theoretical terms are completely redundant and unnecessary. More explicitly, the theorem shows the following: if you can distinguish recursively between "theoretical" and "observational" terms, then your theory can be expressed recursively exclusively in observational terms. The theoretical terms are only of heuristic value in deducing consequences from your axioms.

Note, though, that this is not to say that theoretical terms "serve only a heuristic function." It may perfectly well still be the case that these theoretical terms and the axioms that govern them are central to explanation. It may also perfectly well be the case that these theoretical terms denote objects and properties and relations that actually exist -- that is, that the theories with their abstract terms are semantically correct. This is quite consistent with the replaceability of theoretical terms.

Given this replaceability, however, we are quite free to eliminate theoretical terms and the axioms that govern them and that relate them to observational terms as "merely conventional"; they not only "could" be replaced, but we have a recipe for replacing them. We are left with a corpus of scientific knowledge that is written in a language containing only "observation" terms, and yet is just as useful for predicting as the corpus we think of as warranted by the collective institution of science.

But worse! As I have shown elsewhere[11], you can replace a given theory, employing a given theoretical vocabulary, and embodying given theoretical axioms and given "coordinating definitions", with a new theory, employing whatever theoretical, non-observational terms you like, embodying whatever axioms strike you as plausible for those theoretical terms, in such a way that the new theory has <u>exactly</u> the same observational consequences as the old one.

If this is so, then everything inside the periphery of the web could be regarded as "conventional" -- that is, it can equally well be replaced by something else. But the periphery itself is not immune to charges of conventionality. Any realistic view of observation must allow for <u>errors</u> of observation. Thus what happens to us does not uniquely determine what sentences at the periphery of our web of knowledge we should accept.

Note that this is <u>not</u> merely a matter of the fact that what sound you make when yoiu see a crow is arbitrary. That would be what Grunbaum calls "TSC" -- trivial semantic conventionalism.[12] What is importantly conventional is the fact that the boundaries of the categories into which we take our observations to fall are subject to arbitrary modification. It is open whether we should regard this observation rather than that as being erroneous.

All this suggests that the standard, "might equally well have been otherwise" is a poor touchstone for conventionality. And in fact no conventionalist has suggested that other alternatives might really equally have been chosen. What the conventionalist suggests is rather, as is clear from what Poincare says about geometry, that it is factors of simplicity, convenience, even familiarity, that dictate the choice of one alternative over another. There are always reasons for preferring one convention over another.

Since our concern here is epistemology, let us come right out and characterize an arbitrary linguistic convention (from the epistemological point of view) as a convention whose reason for adoption is non-epistemic. (By "non-epistemic" one means to include such considerations as simplicity, familiarity, computational efficiency, and the like.)

Henceforth, then, we want to consider only non-arbitrary conventions. If any!

2. Formal Framework

We will consider formalized theories.[13] We will construe a theory in what is possibly a somewhat old-fashioned way as consisting of four parts:

(a) A recursive specification of the set of terms and predicates of the theory. We include logical and mathematical terms and predicates, as well as "observational" and "theoretical" empirical terms and predicates. We do not suppose that there is a procedure for telling which is which.

(b) A recursive characterization of the sentences and formulas of the theory. Logic and mathematics are to be included.

(c) A recursive characterization of the axioms of the theory. These are to include both sufficient axioms for whatever logic and mathematics you want, and "meaning postulates", and axioms "pertaining" to the empirical subject matter with which the theory is concerned. We do not require a procedure for distinguishing these classes of axioms.

(d) Rules of inference. Of course we can get by with just *modus ponens*, but we would like to be able to accommodate those who prefer to think of "material rules of inference" rather than axioms.[14]

We suppose that among the theorems of this formal theory we have all the theorems of first order logic, and as much set theory and mathematics as we need. Controversy about these items can be generated, but they are not typically what epistemologists worry about. We suppose that any "meaning postulates" or "logical" or "analytic" relations among terms of the theory (if there are any) are captured by our axioms and rules of inference.

We thus suppose that we have a standard deductive logic built into our theory, that we can characterize in a standard proof-theoretic way. In a similar vein, we suppose that we also have an inductive logic available. One can hardly call it "standard", of course, since the very idea of there being an inductive logic is controversial. Furthermore, our inductive logic is parasitic on our notion of probability, the general idea of the inductive logic being that you can believe stuff that is probable enough. All this will be unpacked as part of the generic formal framework within which we shall attempt to account for the epistemic status of scientific theories.

Let us look at probability first. Probability is defined for all of our theories in the same way. It is a syntactical notion, like that of proof, that we can spell out explicitly in the metalanguage.[15] We list here some of its properties.

(a) Probability is defined for a given language or theory.

(b) Probability is relativized to a corpus **K** of statements, representing the body of *evidence* relative to which the probability of a sentence is to be evaluated. **K** need not in general be deductively closed, but it will contain, at least potentially, all the theorems of the theory. In addition, **K** may contain <u>observation</u> statements, and statements warranted by <u>inductive inference</u>.

(c) Probability is an objective relation, a syntactically definable function from sentences and sets of sentences of a given language to subintervals of $[0,1]$. It is objective in the same sense that provability is objective.

(d) Probability is also *objective in another, indirect,* sense. All probabilities are based on the statistical syllogism, in which the statistical premise represents <u>known</u> (in **K**) frequencies or propensities. (In general, these statistical statements will be empirical, but they <u>can</u> be set-theoretical truths such as: almost all subsets of a given set contain approximately the same relative frequency of objects with a given property as the given set exhibits.)

(e) Probability is interval valued. The form of a probability statement (a metalinguistic statement, be it noted) is:

$$\mathbf{Prob_T(S, K)} = [p, q]$$

(f) Probability can equally well be defined for the metalanguage; thus we can talk about the probability that a certain <u>statement</u>, for example, has a certain metalinguistic property, such as that of being in error.

The general idea behind the epistemological view being presented here is that when something is probable enough, you can simply believe it -- that is, <u>accept</u> it. Given an epistemic notion of probabilty, one is immediately led to ask: "Probable enough, relative to what?" One answer, the one we will endorse, is that the probability in question is to be computed relative to the total evidence you have. Thus we will say:

(1) A sentence **S** will belong to your corpus of practical certainties **K'** if and only if there is a p and a q such that $\mathbf{Prob_T(S,K)} = [p, q]$, and p exceeds whatever we have taken as a level of "practical certainty", where **K** is the corpus of evidence.

Of course we may now inquire into the source of the statements in the evidential corpus **K**.

One possibility would be to suppose that they were phenomenological reports, incorrigible deliverances of the senses. But this is not of much help to us; if we are talking about "evidence" in any ordinary sense, we must include as evidence the results of measurement ("the table is 3.6 $\pm$ .05 meters long"), the content-laden results of observation ("there are thirty-eight black crows in the cage"), and even -- we shall see how shortly -- the results of technically sophisticated observation ("microscopic examination shows the presence of gram-negative bacteria").

So we can raise the same old question again. One answer, this time, is to say that an evidential statement gets into the evidential corpus **K** by being probable enough relative to a corpus **K*** of incorrigibilia. The contents of **K*** might be taken to be the propositional content of phenomenological events, expressed in the language of the theory **T**.

There are difficulties with this view, and we shall ultimately abandon it, but let us look at the picture it yields of the structure of our knowledge. We have, at base, a set of incorrigible statements, $\mathbf{K}^*$. The set of evidential statements, $\mathbf{K}$, consists of those statements $\mathbf{S}$ whose probability relative to $\mathbf{K}^*$ is greater than $\boldsymbol{p}$: $\mathbf{S} \in \mathbf{K}$ if and only if $(\exists p, q)(\mathbf{Prob_T}(\mathbf{S}, \mathbf{K}^*) = [p, q] \;\&\; \vdash p > \boldsymbol{p})$.

The set of practical certainties, $\mathbf{K}'$, is correspondingly defined as the set of statements $\mathbf{S}$ whose probability relative to $\mathbf{K}$ is greater than $\boldsymbol{p}'$, where $\boldsymbol{p}$ is greater than $\boldsymbol{p}'$: $\mathbf{S} \in \mathbf{K}'$ if and only if $(\exists p, q)(\mathbf{Prob_T}(\mathbf{S}, \mathbf{K}) = [p, q] \;\&\; \vdash p > \boldsymbol{p}')$. It is the set of practical certainties that we use for making practical judgments. It is relative to the practical corpus that we compute the probabilities that we multiply by utilities to get the expectations we need for decision theory.

C. I. Lewis said that nothing can be probable unless something is certain.[16] In a very special sense, we will find that to be true; but a stronger claim is embodied in the structure suggested: Something can be judged probable only in relation to evidence that is of yet greater dependability. This is embodied in our requirement that $\boldsymbol{p} > \boldsymbol{p}'$.

The next stage in the development of our theory is to consider <u>meta</u>corpora. This is inspired by our natural interest in error. We must be able to attribute error to the sentences embodying our observations. In the metalanguage we will have predicates, "veridical", "erroneous", that we can use to separate the sheep among our observations from the goats. The set of incorrigibilia here is quite straight-forward: it is the set of sentences inscribed (say) by responsible scientists in their notebooks. No responsible scientist would ever withdraw (erase) such an inscription. But no one would ever take it as evidence without a consideration of the possibility of its being in error.

So let us emulate the object-language structure in the metalanguage. **MK*** will correspond to **K***: but now there is no problem of interpreting it. It consists of just those sentences (of the metalanguage of **T**) that are written down in our notebook, or in the notebooks of our community of scientists.
these sentences mention sentences of the object language. **MK**, the metacorpus of evidential certainties, will consist of those metalinguistic statements whose probability, relative to **MK***, is at least p. We'll explain this in detail in just a moment. Finally, **MK'** corresponds to the set of practical certainties: it is the set of metalinguistic statements whose probability, relative to **MK**, is at least p'.

We have so far mentioned no metalinguistic predicates other than the classical ones of 'is true' and 'is false'. More important than those are the predicates corresponding to 'is inscribed in **X**'s laboratory notebook, where **X** may denote either an individual or a group of individuals. (Chemists, for example.) We may introduce such predicates, expressing relations between sentences of our theory **T**, individuals, and times, as:

(a) $\mathbf{O(X,S,t)}$, to mean that the individual, or group, **X** wrote in his (their) notebook the sentence **S**, at some time before the time **t**; **O** will be a primitive relation of the metalanguage of **T**.

We may use this primitive to define such interesting subsets of the set of sentences of **T** as:

(b) $$\mathbf{VO(X,S)} = \{S|(\exists t)(\mathbf{O(X,S,t)} \;\&\; S)\}$$

This is the set of veridical observations of **X**. Correspondingly, we have, for **X**'s errors,

(c) $$\mathbf{EO(X,S)} = \{S|(\exists t)(\mathbf{O(X,S,t)} \;\&\; \sim S)\}$$

Clearly, we may define the sentences of a given form, the sentences involving a given predicate, etc., and may divide those sets of sentences into the veridical and the erroneous. In the next section we will consider how to make use of this machinery.

3. Representing Bodies of Knowledge

You recall that $\mathbf{K^*}$ contained incorrigible statements of the object language. Are there any? It is hard to believe that there are. What observation statement about the world is proof against correction? Surely any statement that purports to be about the world can be explained away, if only by the strained device of appealing to illusion or hallucination. Incorrigibility can be bought only at the price of vacuity.

Within the framework under discussion, however, we can eliminate **K*** with no loss. This is how it comes about. In place of incorrigible object language sentences in **K***, let us look at the metalinguistic sentences in **MK***. There is no problem in regarding *these* sentences as incorrigible, since they merely report **X**'s judgments about the world. That **X** made such and such a judgment about the world does of course represent a fact about the world: **X** is in the world. But it is a fact of psychology (or sociology, or cognitive science), and not, in **MK**, an empirical claim. If we restrict our theories to physics, biology, and the like, it should not be difficult to keep object and metalanguage clear. We may have difficulty doing this if **X** is an individual introspecting, or if **X** collection of sociologists observing the behavior of sociologists. Even in these cases, the benefits may well outweigh the difficulty.

We are now in a position to make a simplification of our framework. The corpus **K*** will be empty if we suppose that we can make errors in any kind of object language statement judgment. The corpus **MK'** seems to serve no very interesting purpose. So we take **MK*** to contain our observation reports; **MK** to contain metalinguistic statements that are highly probable (p) relative to **MK***, and object-language statements that are highly probable (p) relative to statistical information in **MK** itself. We are thus left with three sets of statements:

(1) **MK***, containing observation reports. These reports

may or may not be veridical, but we take the report itself to be incorrigible. It also contains logical and mathematical statements.

(2) **MK**: the set of <u>evidential</u> certainties. A statement is in **MK** if and only if

> (a) its probability, relative to **MK*** exceeds **p**, or
>
> (b) it is a statement in the object language, and its probability relative to the part of **MK** that does <u>not</u> include the object language is greater than **p**.

(3) **MK'**: the set of <u>practical</u> certainties. A sentence is in **MK'** if and only if its probability, realtive to **MK** exceeds **p'**.

Let us suppose we have had a lot of experiences that we might record as the observation of an alligator and of its blueness, of a non-alligator and its blueness, of an alligator and its non-blueness, and of a non-alligator and its non-blueness. Suppose that among these judgments, practically all the alligator judgments are accompanied by blue judgments.

Consider two theories, $\mathbf{T_1}$ and $\mathbf{T_2}$. $\mathbf{T_1}$ contains an axiom to the effect that all alligators are blue: $(x)(Ax \rightarrow Bx)$; $\mathbf{T_2}$ does not. **MK***, on both theories, will contain a lot of statements of the form "$\mathbf{O(X,}$"Aa_1"$\mathbf{,t)}$", $\mathbf{O(X,}$"Ba_1"$\mathbf{,t)}$, and so on. It will also generally contain some "negative" statments "**O(X,** "$\sim Ba_i$"**, t)**".

MK*, since it is populated by incorrigibilia, will also contain statistical information -- e.g., to the effect that 98% of the observation reports that have the form "Aa_i" are accompanied by observation reports having the form "Ba_i".

Now let us see what happens according to the theory we employ. If we employ theory $\mathbf{T_2}$, we have no reason to think that any of the observation reports are in error, while if we employ theory $\mathbf{T_1}$, we do. But someone might say that we have no reason, on either theory, to think that our observations are <u>not</u> in error! We need some principled way to deal with the possibility (or certainty, in the case of $\mathbf{T_1}$) of observational error.

Various possible answers are possible, but one answer I have given elsewhere[17] is particularly simple: Adopt two principles:

<u>Minimization Principle:</u> Minimize the attribution of error to your observation reports.

<u>Distribution Principle:</u> Subject to that condition, distribute the error you <u>must</u> attribute to your observational reports in as even a way as possible.

It is hard to see why anyone would want to impugn the first principle. It is true that a lot of our alligator observations might turn out to be wrong. Maybe they all are, since alligators are material objects and maybe there are no material objects. But this seems like a silly thing to worry about, though it may not be a silly project to analyze what it is to be a material object. On the other hand, the principle seems a natural one: it is our knowledge about the world that provides us, in ordinary life, with a touchstone for reality. We may judge an object in a tree to be a black cat, but when it flies away, we know that our judgement was in error.

The second principle is less natural. In particular, it would seem that negative judgments ("$\sim Aa_3$") are less prone to error than positive ones. But this in fact follows from plausible assumptions concerning our judgments: We first minimize error: to do this, it seems plausible to suppose that far fewer negative judgments will be need to be labelled false than positive judgments. Thus when we choose the distribution that is most even, it may well still be the case that the relative frequency of false positive judgments must be presumed to be much higher than the relative frequency of false negative judgments.

From the application of these two principles, we do not, of course, discover which observation reports are misleading. (If we did, we would cleverly excise them so that the rest could be taken at face value!) What we get are the "observed" relative frequencies of error for the various sorts of observation report. This sample data is in **MK***, which we may suppose to be deductively closed in virtue of the fact that the items in it are incorrigible.

In theory $\mathbf{T_2}$ we don't have to suppose there are any errors of observation.

In theory $\mathbf{T_1}$ we must suppose that a certain number of alligator judgments and a certain number of non-blue judgments are in error.

Now turn to **MK**. We have sample information in **MK*** about the relative frequencies of various sorts of errors. We can use this as the basis for a straightforward statistical inference concerning the long-run or general frequencies of errors of these various sorts. (We assume here that there are no problems concerning statistical inference, and that the result is the acceptance, in the evidential corpus **MK** of level **p**, of a statistical hypothesis such as: "In general, alligator judgments turn out to have to be rejected with a relative frequency between .04 and .08.")

In theory $\mathbf{T_2}$, since we not only have no evidence of any errors in our sample of judgments, but know that we can have no such evidence, we accept in the evidential corpus the metalinguistic generalization "Between 0% and 0% of alligator judgments turn out to have to be rejected."

In theory $\mathbf{T_1}$, we do have some evidence of error, and so we might accept the metalinguistic statistical statement, "between 4% and 8% of the alligator judgments have to be rejected.

We also have in **MK** all the observation reports from **MK***. This accounts for the interesting metalinguistic contents of **MK**. We also have in the evidential corpus some probable statements of the object language. In particular, if "Aa_i" is a random member of the set of (syntactically characterized) observation reports, and the proportion of such statments that belong to **VO** (the veridical ones) is known in **MK** to exceed **p**, then the probability that "Aa_i" belongs to **VO** is more than **p**. If this is so, then since "Aa_i iff "Aa_i" ϵ **VO**" is in **MK**, "Aa_i" itself will be in **MK**.

Of course "Aa_i" is not in general a random member of the general set of observation reports, but rather a random member of the subset that have the form "Ax". Whether this makes any difference or not depends on whether we are using $\mathbf{T_1}$ or $\mathbf{T_2}$. There is nothing in $\mathbf{T_2}$ to entail any errors of observation, so that if we have an observation report, we can accept it at face value, and include the corresponding statement in **MK**. In the case of $\mathbf{T_1}$, if the relative frequency with which observation statements must be rejected is high, the fact that "Aa_i" is an observation report may <u>not</u> suffice to justify its inclusion among the evidential statements. In particular this is so for those statements "Aa_i" that are paired with statements "$\sim Ba_i$" in **MK***. We are bound to lose some statements, on this ground.

On the other hand, since it is a theorem that if **S** entails **S'** then the probability of **S'** exceeds that of **S**, and in $\mathbf{T_1}$, "Aa_i" entails "Ba_i", and "$\sim Ba_i$" entails "$\sim Aa_i$", the contents of **MK** should be accordingly expanded. In $\mathbf{T_2}$ there are no interesting entailments. ("Aa_i" entails "$Aa_i \vee Ba_j$", but who cares?)

In theory $\mathbf{T_1}$, containing the generalization, we lose some *observation statements in* **MK** *due to error; but we gain some* observation statements from entailment.

In theory $\mathbf{T_2}$, lacking the generalization, we retain all of our observations. Since the chance of error is 0, the probability of a conjunction of observation statements is the same as the probability of each of them, and we may also have in **MK** a statement to the effect that in a sample of n alligators, 96% have been found to be blue.

It is, however, the corpus of practical certainties **K'** that we require in order to plan our lives and make our decisions. The part of the corpus of practical certainties that interests us is the part that contains sentences in the object language. These sentences fall into a variety of categories. First there are the observational sentences that are inherited (indirectly) from **MK***. In general $\mathbf{T_2}$ will provide more of these than $\mathbf{T_1}$. Second, there are statistical statements that are rendered practically certain by the data in **MK**. Thus the statistical statement, "about 95% of alligators are blue" may be rendered so probable by the evidence in **MK** that it becomes included in **K'**. Then relative to the corpus **K'** the probability of "Ba_j" may be (assuming randomness) $0.95 \pm .03$. Finally, singular statements, such as "Ba_j" can be rendered probable enough, relative to **MK**, to be included in **K'**, even when they are not probable enough to be included in **MK** itself. The most important case of this sort is when there are entailments in a theory that lead from observation statements to their implications.

In the theory $\mathbf{T_2}$ we may have, in the set of practical certainties **K'**, all of the observation statements (e.g., "Aa_8") that correspond to observation reports in **MK*** (e.g., "$\mathbf{O(X,}$"Aa_8"$\mathbf{,t)}$". (Of course we also have everything entailed by these statements, e.g. "$Ba_6 \vee \sim Ba_6$".) Since we have sample data represented in **MK**, there will be statistical hypotheses that are probable enough to be included in **K'**. For example, we may be practically certain that almost all alligators are blue. Relative to such a corpus we should predict that an alligator of unspecified color will be blue, though that statement will not appear in **K'**.[18]

In the theory $\mathbf{T_1}$ the corpus of practical certainties **K'** may not contain all the observation statements corresponding to the observation reports in **MK*** (in fact, if our observations are, according to our theory, highly prone to error, we could end up with *no* observation statements at all in **MK**). But we have the additional observation statements entailed, according to the theory, by those we do retain. In addition, as in theory T_2, we have statistical statements rendered practically certain by the evidence in **MK**.

4. Choosing Between Theories

By considering what happens to our three-part bodies of knowledge when hypothetical incomplete observation reports are added to the metacorpus **MK***, we can formulate criteria for the preferability of one theory over another in terms of the <u>predictive observational content</u> of the corpus of practical certainties **K'**. The predictive observational content of a corpus of practical certainties **K'** consists of observation sentences in **K'** that do *not* correspond to observation reports in the corresponding **MK***, together with sentences whose probability relative to **K'** exceeds the level of **K'**.[18]

$$(4) \qquad \mathbf{POC(MK^*, p, p') = \{S | \; S \in K' \; \& \sim S \in MK^* \; v} \\ \mathbf{Prob_T(S, K') > p'\}}$$

In the example we have just been considering, we could add such statements as "$\mathbf{O(X,}$"Ax_1"$\mathbf{,t)}$, "$\mathbf{O(X,}$"$\sim Ax_2$"$\mathbf{,t)}$", "$\mathbf{O(X,}$"$\sim Bx_3$"$\mathbf{,t)}$", to our corpus **MK***, where x_1, x_2, x_3, ... are distinct terms new to the language. Which of the two theories is preferable, depends on our past experience. If our past experience has involved a lot of non-blue alligators, then we may obtain no predictive observational consequences under either theory, though we could still get some statistical advice from $\mathbf{T_2}$. If our past experience has involved a few non-blue alligators, then $\mathbf{T_2}$ may provide more in the way of predictive power, since it requires less attribution of error to our observations. If our past experience has involved very few non-blue alligators, $\mathbf{T_1}$, in which "Ax_i" entails "Bx_i" and "$\sim Bx_i$" entails "$\sim Ax_i$", may well provide more predictive observational content.

What happens depends on what partial observations we add to **MK***. We want to add enough, of various kinds, so that we can see what is happening; but we don't want what we are hypothetically adding to affect the error frequencies. (Since the new terms are all distinct, they will dilute the error frequencies of the original corpus, if we count them as ordinary observation reports.) Since we have a formal object to play with, we can accomplish all this. Let us add partial observations to **MK*** in the same proportion as our past observations of the corresponding sorts; let us call the result "aug-**MK***"; let us not use this information to update the statistical components of **MK** and **K'**; and let us then compare the **POC** of aug-**MK*** on the two theories.

We still have the parameters **p** and **p'** to take account of. For present purposes, let us just take them to be fixed by context.[20]

In "Theories as Mere Conventions"[21], I have argued that this standard of preference for one theory over another (I perhaps perversely called them "languages" there, to emphasize the a priori character of the theoretical axioms and generalizations), accounts for much of what Kuhn and Feyerabend have drawn our attention to. I claimed that we could not only account relatively neatly for the replacement of one theory by another, but for the expansion of a theory by the addition of new generalizations, the (rare) contraction of a theory by the "refutation" of generalizations, and even for the replacement of "observational" terms as theories change. I will not repeat those arguments here, but merely claim that the epistemological framework adumbrated above supports a plausible empiricist view of scientific inference.

5. More Epistemology

There are some specifically epistemological questions that it is important to consider in this framework. Perhaps the most pressing is the question of observation. I have been freely referring to "observation sentences", though I have also said that this framework did not require a sharp observational/theoretical distinction. Once we have admitted errors of observation, and eliminated **K***, we no longer need to worry about finding sentences of the object language that can be known with certainty on the basis of observation. We can take our observations at face value, in the form of observation reports.

The general idea is that observation reports represent judgments based directly on experience. By "directly", I mean non-inferentially, and by "experience" I mean what happens to me in response to what I do. By merely looking, I can judge that a certain object is an alligator, for example. By hefting a certain piece of iron, I can judge that it weighs about three pounds. Lots of learning can be involved. It is this possibility that I was depending on earlier when I referred to a sentence about gram-negative bcteria as "observational." In fact, we are at last in a position to remove observation from the hands and eyes of the physiologically normal amateur, and allow the experienced expert -- the histologist -- full scope. What is more important, however, is that the theory of error to which we are led in **MK** provides control.

But in comparing theories, at least at an elementary level, we do have to count observation sentences. We cannot avoid the question: What sentences? Clearly we want to count only sentences that are going to be dependable guides to experience.

Dependability in prediction may be construed in the same way as dependability in observation. An observation report represents a dependable observation just in case it is unlikely that future observations will impugn the contents of that report. More explicitly: a kind of observation report is a dependable kind just in case the general frequency of error among reports of that kind is sufficiently low. (Less than 1 − **p**, where **p** is the evidential level) We obtain this frequency from considering the sample proportion of such reports we are forced to construe as false in an initial segment of our experience as an indication of the long run relative frequency with which such reports will have to be construed as false. It is our theory -- that is, our general knowledge about the world -- combined with the minimization principle and the distribution principle, that tells us what we must take the *relative frequency of error* of various kinds of observation judgments to be.

In making an observation judgment, there is only one thing that can go wrong: it may be the case that not-S obtains. Or in epistemic rather than semantic terms, it may be the case that further experience may induce us to regard that kind of judgment as generally unreliable.

In making an observation prediction there is still only one thing that can go wrong: it may be that not-S obtains. Or, in epistemic rather than semantic terms, that further experience may induce us to regard that kind of judgment as generally unreliable.

There is a difference, of course, between verifying a prediction, and simply recording an observation. In order to verify the prediction, I must often put myself in the way of a certain kind of experience: I must open my eyes and look, or focus my attention, or try to see whether or not **S** obtains, or look into an instrument. But this difference is not related to a difference in the reliability of the observation. If I try to see **S**, and fail, that results in an observation report that enters into the data for determining the reliability of non-**S** observation, as well as the reliability of any other observation sentences that went into the prediction.

The point is that reliability is a matter internal to what I am calling a theory, and that observationality is a property that is a matter of degree, and related to the internal measure of reliability.

There is much to be said about special cases. For example, if we have a physical theory, and the physical theory predicts that (say) a certain temperature should be 55° C, and we measure the temperature with a reliable instrument (a reliable instrument is one that gives rise to reliable readings -- that is, to reliable observation reports), and find 75° C, we may well reject the physical theory, or replace it with a more modest theory that does not lead to the prediction in question.

The argument goes as follows: We have a larger and more pervasive theory, of which the law in question is only a small part, which entails that if we get one genuine anomalous result, we can generate any number of them. So it is within the large theory itself that the deep negative import of finding a temperature of 75°C makes itself felt. This is simply the negative form of a procedure that in a positive form allows us to get a lot of information from very few experiments. (For example, from a single careful experiment, we can get a very narrow distribution for the melting point of a new chemical compound; why? Because our general theory requires that all samples of a pure chemical compound melt at the same temperature under standard conditions.)

How about the confirmation and disconfirmation of theories? If we identify a theory with the conjunction of the non-logical, non-mathematical axioms of the theory, then we have no difficulty in assigning a degree of probability to the theory relative to a body of knowledge. There are two cases: the theory is providing the background for the bodies of knowledge. Then the probability of each of its sentences (axioms and theorems) is 1.0. Alternatively some other theory is providing the background. The probability of each of its sentences (axioms and theorems) is 0.0. The probabilities are perfectly well defined, but perfectly useless.

When we talk about the probability of a theory, I think we are not talking of the probability that the theory is true, in the sense in which wehen we talk of the probability of heads on the next toss, we are talking of the probability that the proposition that the coin lands heads on the next toss is true. I think we have in mind something quite different: for example, the probability that we will accept the theory in question ten years from now. This is a quite different question -- a question in the sociology of scientific change, if you will -- and one to which sensible answers might well be provided by a study of the history and sociology of science. That is, it is a question that is internal to another branch of science, and not a philosophical question at all.

6. Conclusion

What is the upshot of all this? That there are good arguments for conventionalism in science, but that if they are pursued too far they leave us with *empty bodies of knowledge*. We must have grounds for preferring one convention to another, and furthermore, they must be epistemic grounds. Such grounds cannot be provided directly by observation. As many conventionalists have pointed out, observation is always subject to error.

So suppose we take theory in a broad and general sense as providing a framework for our body of knowledge. We take our body of knowledge to consist of three sets of sentences: a set of sentences in the metalanguage of the theory, **MK***, containing observation reports. From this set of sentences, together with the background theory, we can derive the statistics of observational error that are required by that background theory. Given these statistics, we can determine the probability of statements of the object language, relative to the metalinguistic and theoretical background.

The statistics of error and the statements of the object language that are highly probable form the evidential corpus **MK**, Statements in the object language that are so probable relative to the corpus **MK** as to be regarded as practically certain form another corpus: **K**, the corpus of practical certainties. **K** contains both statistical knowledge and predictive observational knowledge.

The contents of all of these corpora are determined by the background theory and by the contents of **MK***, and by the selection of two parameters to distinguish evidential certainty and practical certainty.

Finally, the criterion by which we determine that one theory is to be preferred to another is the frankly pragmatic one of the number (or content) of the predictive observational statements in the practical corpus.

The notion of error, and hence of observationality, is relativized to a theory, but is determined, for a given theory, by our experience -- by what happens to us. Thus even though theories are regarded as "conventional" or a priori, even though the notion of error that determines the content of the practical certainties of a theory is internal to the language of the theory, we can still have an objective measure of the degree to which one theory rather than another satisfies our desire to anticipate the future.

This picture leaves certain puzzles to be resolved, but seems to offer a picture of scientific knowledge that is plausibly rooted in the empiricist tradition. It may well be that the general framework of metalinguistic report at one extreme, and object language prediction at the other, tied together by an analysis of error that involves both languages, has something to offer other areas in epistemology as well.

Henry E. Kyburg, Jr. University of Rochester

Notes

* Research bearing on the topic of this paper has been supported by the United States Army Signals Warfare Center.

1. Ernst Mach, *The Science of Mechanics*, Open Court, LaSalle, 1960. The first German edition appeared in 1883. See especially Chapter II on the development of the principles of dynamics.

2. Henri Poincare, *Science and Hypothesis*, Dover Publications, New York, 1952. This book appeared in French in 1903.

3. *Ibid.* p. 50.

4. Mach, *op. cit.* p.306.

5. Pierre Duhem, *The Aim and Structure of Physical Theory*, Princeton University Press, Princeton, 1954. In French, it appeared as *La Theorie Physique: Son Objet, Sa Structure*, in 1906.

6. David Lewis, *Convention*, Harvard University Press, Cambridge, 1969. It is interesting to note the similarity of the final clause quoted, and that occurring in Mach's attribution of conventionalism to Poincare just quoted.

7. W. V. O. Quine, and J. S. Ullian, *The Web of Belief*, Random House, New York, 1970.

8. W. V. O. Quine, *Mathematical Logic*, Harvard University Press, Cambridge, 1940.

9. W. V. O. Quine, *Word and Object*, MIT Press, 1960.

10. William Craig, "Replacement of Auxiliary Expressions," *Philosophical Review* **65**, 1956, 38-55.

11. Henry E. Kyburg, Jr., "How to Make Up a Theory," *Philosphical Review* **87**, 1978, pp 84-87.

12. Adolf Grunbaum, *Geometry and Chronometry in Philosophical Perspective,* University of Minnesota Press, Minneapolis, 1968, p. 20.

13. In previous publications, for example "Theories as Mere Conventions," in *Minnesota Studies in the Philosophy of Science,* I have called the objects with which we are concerned "languages".

14. See, for example, Stephen Toulmin, *The Uses of Argument,* Cambridge University Press, Cambridge, 1958.

15. We won't do that here. Various sources for the notion of probability we will employ are, Henry E. Kyburg, Jr, *The Logical Foundations of Statistical Inference,* Reidel, Dordrecht, 1974; "Epistemological Probability," *Synthese* **23**, 1971, 309-326, reprinted in *Epistemology and Inference,* University of Minnesota Press, Minneapolis, 1983; and "The Reference Class," *Philosophy of Science* **50**, 1983, 374-397.

16. C. I. Lewis, *An Analysis of Knowledge and Valuation,"* Open Court, LaSalle, 1949.

17. In *Theory and Measurement,* Cambridge University Press, 1984.

18. In the more complex version of this scheme, in which incorrigible observation <u>statements</u> appeared in $\mathbf{K}^*$, we could get sample data in $\mathbf{K}^*$ through deductive closure, and thus statistical statements in $\mathbf{K}$ (now collapsed into $\mathbf{MK}$), and thus, in the case of extreme statistical statements, predictive singular statements

could appear in **K'**. We now lose this feature, but since it was a feature of a very unrealistic case, it is a small loss.

19. The second clause is new. Since, however, there is no way in which you can rationally bet against a proposition whose probability relative to your corpus of practical certainties exceeds the level of that corpus, it seems appropriate to include such statements as part of the predictive observational content of the corpus.

20. In "Knowledge and Acceptance", forthcoming, I attempt to show how context determines the levels of practical and evidential certainty.

21 Forthcoming