AD-A250 440

[DOCUMEN]TATION PAGE

Form Approved
OPM No. 0704-0188

(2)

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1989	Unknown

4. TITLE AND SUBTITLE

Evidential and Practical Certainty

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-0032

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

It is argued that there are significant advantages to using a two-level framework for knowledge representation. It allows both what corresponds to default assumptions, based on high probability, and also it allows a way to abandon those defaults under special circumstances. The basic mechanism for determining membership in the lower level representation is a criterion of high probability relative to the evidence embodied in the upper level. Probabilities are also defined, in the same way, relative to the lower level, and these are the probabilities that are used in the computation of decisions.

14. SUBJECT TERMS

Artificial Intelligence, Data Fusion, Evidential and Practical Certainity, Statistical Inference

15. NUMBER OF PAGES

7

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

EVIDENTIAL AND PRACTICAL CERTAINTY*
(area: B7)

by

Henry E. Kyburg, Jr.
University of Rochester

kyburg@cs.rochester.edu

abstract

It is argued that there are significant advantages to using a two-level framework for knowledge representation. It allows both what corresponds to default assumptions, based on high probability, and also it allows a way to abandon those defaults under special circumstances. The basic mechanism for determining membership in the lower level representation is a criterion of high probability relative to the evidence embodied in the upper level. Probabilities are also defined, in the same way, relative to the lower level, and these are the probabilities that are used in the computation of decisions.

1. Introduction.

The search for an unquestionable basis as a foundation for knowledge has been a philosophical grail at least since Descartes' *Discourse on Method*. While philosophers have sought such a foundation, practical men, engineers, and scientists have been quite content with some form of practical certainty. This desire for practicality, for getting on with the job, also motivates some of the concern with non-monotonic logic (McCarthy [1980], [1987], Reiter [1980], McDermott [1980]). We will illustrate these practical concerns in two cases: measuring distance and measuring frequency. We will then offer a proposal for a two-level knowledge representation framework based on an epistemic notion of probability. We will show that this accommodates the two examples; that it provides a natural way of representing default reasoning; and that it provides for the simplest applications of decision theory. Finally, we shall discuss some shortcomings, and some directions for future research.

2. Measuring Distance.

Measurement provides a simple and clear illustration. To obtain the distance between two points, we measure. There are a variety of techniques, appropriate to a variety of contexts, for measuring the distance between two points. We apply some appropriate method M and conclude that the distance is $D \pm \Delta$.

The claim that the distance between the two points in question is $D \pm \Delta$ is not "certain." It is, indeed, the sort of claim that might well be reported as an observation (particularly if it is the result of averaging several distinct measurements); but it is also a claim whose denial has a finite (and calculable!) probability.

We assume that Δ is chosen so that the claim in question meets whatever are the conventional requirements for confidence in the context at hand. (Sometimes measurements are reported in the form "$D \pm sd$ " where sd is the standard deviation of the assumed distribution of error characteristic of the method M. This has the advantage that, assuming the error distribution is roughly normal, the reader can calculate his own interval for whatever degree of confidence he wishes.)

So the assertion that the distance is $D \pm \Delta$ is accompanied by a certain confidence. This confidence, clearly, comes from our knowledge of the distribution of error that is characteristic of the method of measurement M.

In general, we suppose that the distribution of error characteristic of M is approximately normal, and has a mean of close to 0, and a standard deviation of sd . Note that we say "approximately;" it would be unreasonable to claim that we knew the error distribution exactly. But more than this, if the error distribution were really normally distributed, there would be a finite probability that that the distance between the chosen points was in fact negative. (If our reading is 23 cm., and the standard deviation is 1 cm, a negative error of 24 standard deviations would mean that the distance was negative.) "The probability of this is too small to take seriously," you say. Precisely. The normal distribution is too precise to take seriously.

Very well, where does the approximate distribution come from? The full answer to this is rather complicated (a discussion can be found in [1984], but for present purposes, we can say is is just a distribution that we take for granted, in the same sense that we take the results of our individual measurements for granted. I assume that I have made one measurement (or a sequence of measurements). This result is a numeral (or a sequence of numerals). I apply an assumed distribution of error to those results, and infer, with practical certainty, that the distance in question is $D \pm \Delta$.

This situation is illustrated in figure 1. The knowledge of the distribution of error, as well as the result of the individual measurement, appears in what I shall call the *evidential corpus*. Knowledge about the distance appears in the *practical corpus*. What constitutes practical certainty will be dependent on context. What appears in the evidential corpus may in turn be questioned: we can ask what grounds we have for accepting the error distribution we do accept.

Finally, the inference -- we shall take it to be a probabilistic inference -- that leads to the inclusion of the sentence "the distance is $D \pm \Delta$" among our practical certainties is not automatic. We may well have other information in our evidential corpus that may undermine this sentence. (For example, that the distance has already been measured to be D' ; in that case what is practically certain should be determined by both measurements.) We require that the measurement be a random one in the appropriate epistemic sense.

3. Measuring Frequency.

Let us consider measuring the long run frequency of an event in a class of events. (The frequency of survival for five years of patients exhibiting a certain cluster of symptoms, for example.) A "measurement" is just the selection of a sample by a method M, and the observation of the relative frequency of the property in question in that sample.

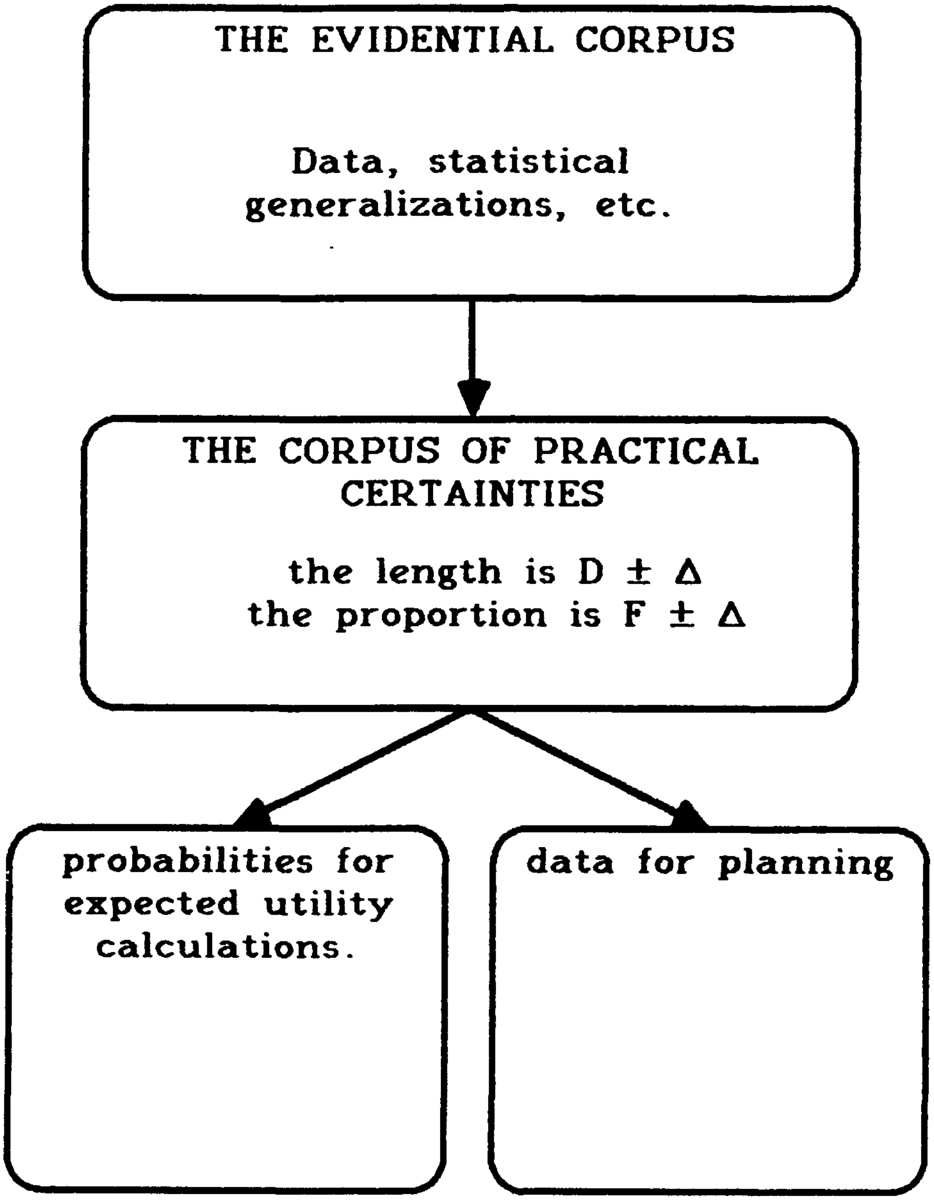
THE EVIDENTIAL CORPUS
Data, statistical generalizations, etc.
THE CORPUS OF PRACTICAL CERTAINTIES
the length is D ± Δ
the proportion is F ± Δ
probabilities for expected utility calculations.
data for planning

figure 1

We do not conclude that the long run frequency is exactly the same as that in our sample, but we take account of a theory of error about the representativeness of samples to infer, with a certain degree of confidence, that the long run frequency lies within a certain interval about our observed frequency.

As in the measurement of length, when things go right, we can become practically certain that the long run frequency in question lies in the interval $F \pm \Delta$. The error distribution characteristic of the method we have employed, like the error distribution relevant to the measurement of length, is among our evidential certainties. The data providing the observed frequencies is also represented at the evidential level.

Of course things aren't automatic: we may have knowledge of a sample, and knowledge of the error distribution of the method, and not have the claim that the long run frequency is $F \pm \Delta$ among our practical certainties. This would be the case, for example, if we happened to have among our evidential certainties knowledge of another sample relevant to the estimation of the frequency in question.

A new feature of this example is that the statement about a long run frequency that appears in our practical corpus may play a different role than the statement about the distance between the two points. If we wish to build a bridge between the two points, we will use as a constraint on our engineering design: we simply take it for granted -- take it as practically certain -- that the distance we must span is in the interval .

But what may concern us in the second example is making a decision to which whether or not the next item has or lacks the property in question is relevant. Perhaps we are an insurance company being asked to quote a premium for five year life insurance on one of the patients having the collection of symptoms at issue.

For making that quotation, what we need is the probability that a patient -- and quite possible, the probability that a particular patient -- will survive for five years. For this purpose, what interests us is not $F \pm \Delta$ itself, since we will not be insuring the whole class of patients, but the probabilities that can be derived from this knowledge.

In short, just as we can base probabilities on the evidential corpus, so we can base probabilities on the practical corpus. And it is these latter probabilities that we need to employ in determining expected utilities and in making decisions among alternative courses of action.

4. A Two-Level Representation.

The Evidential Corpus: Let us take the evidential corpus to consist of a finite set of axioms. These axioms may include both general and particular statements. For example, we might include the general statement that the distribution of errors generated by measurement method M is distributed nearly normally, with a mean of approximately 0.0 and a standard deviation of approximately sd. We might include a statement to the effect that three measurements of the distance in question have been made, yielding the results 23.4, 23.8, 23.6. We might include the statement that a sample of size n has been observed, and m of items observed have had the property in question. In general, we include statements in the evidential corpus that no further observations, in the context at hand, are going to impugn.

The Practical Corpus: What goes into the practical corpus, in principle, are exactly those statements whose probability relative to the evidential corpus exceeds some number (the appropriate degree of confidence, a.k.a. practical certainty) determined by the context.

Practically, this is an inappropriate standard in a number of respects. First, logical and mathematical truths will have probability 1 relative to any evidential corpus. But we cannot expect our practical corpus to contain them all. Furthermore, we cannot even decide whether an arbitrary statement is a theorem.

Second, even if we disregard mathematical and logical statements, any empirical statement may have an infinite number of distinct logically equivalent forms. (These forms include, for example, S & T, for empirical statement S and logical theorem T !)

Third, even if we look only at "purely" empirical statements (however they may be defined) there will be a great many of them.

We therefore construe the practical corpus as a potential set of statements. Formally, it is the set of all those statements whose probability relative to the evidential corpus exceeds the canonical value p ; practically, we need only be able to tell, of any given statement S, whether or not it belongs to the practical corpus. We can do this if, for any given statement S, we can tell what its probability is, relative to the evidential corpus.

What logical structure can we attribute to these corpora? Since everything that gets into the practical corpus gets there by being probable relative to the evidential corpus, we cannot expect the conjunction of two statements that appear in the practical corpus to appear in the practical corpus. It follows that the practical corpus cannot be deductively closed.

We do have the following theorem, though: If S is in the practical corpus, and S entails T, then T will also be in the practical corpus [1961, 1974]. There is, of course, no reason that conjunctions cannot sometimes be probable enough to get into the practical corpus, and if they do, their consequences do, too. This reveals something important about argument: What we demand of an argument in order to be rationally persuaded of its conclusion is not merely that it be valid, and not merely that each premise be acceptable; we demand also that the conjunction of the premises be acceptable.

Since we may suppose that in another context our evidential corpus may be construed as a practical corpus justified by a yet more demanding evidential corpus, these same properties should be attributed to the evidential corpus: deductive closure under single premises, failure of deductive closure in general.

5. Probability.

Epistemic probability represents a relation between a statement (whose probability we're after) and a set of statements (representing a body of evidence). It is interval valued (probability greater than p is to mean lower probability greater than p). It is objective: two entities with the same evidence will assign the same probability. It is based on knowledge of frequencies: a probability can have the value [p,q] only if the body of evidence contains relevant statistical knowledge mentioning the same interval. All statements known to be equivalent in truth-value have the same probability. Three principles, a subset principle, a supersample principle, and a cross product principle, are required to eliminate conflicting reference classes. A further principle, the strength principle is required to pick out the reference class about which we have the most precise (useful) information. (For details, see [1961, 1974, 1983].)

Subjective probability is different from epistemological probability. For one thing, it can vary from agent to agent independently of differences in evidence. For another, it supposes that the result of observation (or measurement) is a full probability distribution. Thus when I observe the value 23.4 in measuring the distance between the two points in question, the result according to the subjectivistic view is not $D \pm \Delta$, but rather an entire normal distribution with mean D and variance characteristic of the method of measurement. It is this distribution we are to use in designing our bridge.

The subjective view may be viable in simple cases. One may conjecture that it becomes hopelessly complex in any real world situation.

6. <u>Accommodating the examples.</u>

Distance: In our evidential corpus we have the statistical information that method M is subject to errors distributed approximately normally with mean 0.00 and standard deviation 0.05. We make a measurement yielding the value 23.40. Three standard deviations is taken to yield a practical certainty.

Case I. This is the only measurement we have of the distance, and we know of nothing peculiar about it. It is an epistemologically random member of the set of possible measurements, with respect to yielding any given amount of error, relative to what we know. We may be practically certain that the length is between 23.25 and 23.55.

Case II. We also have the results of another measurement by the same method, 23.50. It follows from our knowledge about error that the distribution of error among the averages of <u>pairs</u> of measurements is approximately Normal with mean 0.00 and variance $.05/\sqrt{2}$; if the pair of observations is epistemologically random, then we may be practically certain that the distance is in the interval 23.45 $\pm .15/\sqrt{2}$.

Case III. We happen to be evidentially certain that the distance between the points is 23.00. Then none of the observations is epistemologically random and we should be practically certain than the distance is 23.00, regardless of what we know about errors of measurement.

Frequency. We know that almost all (where "almost all" corresponds to practical certainty) n-membered subsets of the set of A' s reflect, within an amount d, the proportion of B' s among A' s in general. (In fact this is a set theoretic truth, and should be included in all evidential corpora, though it may not be the most relevant statistical knowledge in these cases.) We know that m of our sample were B' s.

Case I. This is the only sample we have, and we have no other knowledge bearing on the frequency of B' s among A' s. We may be practically certain that in the long run $m/n \pm d$ of the A' s are B' s.

Case II. This is merely part of a larger sample. The supersample principle referred to dictates that we base our inference on the larger sample.

Case III. We have theoretical grounds, in our evidential corpus, for supposing that the long run frequency in question is f. In that case the sample is not epistemically random for determining the long run frequency, and we should be practically certain that the long run frequency is *f*.

7. <u>Defaults.</u>

Suppose that our evidential corpus contains the information that almost all (i.e., a fraction corresponding to practical certainty) birds fly, and that Tweety is a bird.

It follows that if Tweety is an epistemologically random member of the set of birds, we can be practically certain that Tweety flies. That Tweety flies is among our practical certainties.

Add to the evidence that Tweety is an emu, and suppose that we know almost no emus fly. The corresponding set of practical certainties will (ceteris paribus) contain the statement that Tweety does not fly. Add the fact that flemus fly, and that Tweety is a flemu: the set of practical certainties will include the statement that Tweety flies after all.

To be sure, these defaults are based on frequencies (or hypothetical frequencies) rather than "typicality." It seems likely that hypothetical frequencies can take care of "typicalities," if any, that do not correspond to actual frequencies.

Sometimes we get cancellation; this is a consequence of our rules of randomness. Represent the generalities of the Nixon Diamond by statistical statements in the evidential corpus. Add to the evidential corpus the statement that Nixon is a Republican; we become practically certain that he is not a pacifist. Add instead that he is a quaker. We become practically certain that he is a pacifist. Add both statements to the evidential corpus. We conclude that we are practically certain of nothing about Nixon's attitude twoard war.

8. <u>Decisions.</u>

In general, particularly when the relevant probabilities are not extreme, we are (or should be) less interested in knowing what default we should adopt than in getting even a vague idea of the probabilities involved. If there is something serious hinging on whether an arbitrary bird named Tweety can fly, I am likely to be more interested in the proportion of birds that can fly -- even a vague proportion -- than I am in the question of whether "most" birds fly or whether birds "typically" fly. In extreme cases (like the two illustrative examples), it seems plausible to suppose that we can achieve practical certainty. I do not think that in most contexts I would be "practically certain" that a random republican would be a non-pacifist, or that a random quaker would be a pacifist, or that a random bird would fly. Before betting on any of these propositions, I would have to know what odds I was being offered.

But this is just to say that what concerns me are probabilities evaluated relative to my corpus of practical certainties. These probabilities, however, require uncertain knowledge for their evaluation. That is provided for by the two-level system outlined.

In view of the fact that probabilities, relative to the corpus of practical certainties are interval valued, decision theory becomes complicated. It no longer suffices (as it does on the subjective view) simply to "maximize expected utility." Like probabilities themselves, expected utilities will be interval valued. It is easy to rule out alternatives that are dominated; it is not clear what the next step should be.

9. Future Concerns.

We need an algorithm for computing probabilities relative to a body of knowledge. (See Loui [1985].)

It would be nice to know that the practical corpus can be finitely axiomatized so long as the evidential corpus can be.

Computing probabilities is expensive in time and space, so it would be nice to be able to have criteria for determining what parts of a body of knowledge are potentially relevant to the computation of a probability.

A decision theory that is designed to deal with interval expectations would be useful.

Principles for determining the level of practical certainty (the corresponding confidence) are desirable. (See [1988].)

It would be useful to unpack in more detail the implications for default inference of this two-tier system of evidential and practical certainties.

* Research on which this work was based was partially supported by the U. S. Army Signals Warfare Center.

References:

Kyburg, Henry E. Jr.[1984]: *Theory and Measurement* Cambridge University Press, Cambridge.

Kyburg, Henry E., Jr. [1988]: "Full Belief." *Theory and Decision* 25, 137-162.

Kyburg, Henry E., Jr.[1974]: *The Logical Foundations of Statistical Inference*, Reidel.

Kyburg, Henry E., Jr.[1983]: "The Reference Class," *Philosophy of Science* 50, pp 374-397.

Kyburg, Henry E., Jr.[1961]: *Probability and the Logic of Rational Belief*, Wesleyan University Press.

Loui, Ronald P.[1986]: "Computing Reference Classes," Proceedings of the 1986 Workshop on Uncertainty in Artificial Intelligence, 183-188.

McCarthy, John "Circumscription -- a Form of Non-Monotonic Reasoning," Artificial Intelligence 13, 1980, 27-39

McCarthy, John: "Applications of Circumscription to Formalizing Common-Sense Knowledge," Artificial Intelligence 28, (1986) 89-116.

McDermott, D., and Doyle, J. (1980): "Non-Monotonic Logic I," Artificial Intelligence 13, 41-72.

Reiter, R. (1980): "A Logic for Default Reasoning," Artificial Intelligence 13, 81-132.