

AD-A250 600



IDENTIFICATION PAGE

Form Approved
OPM No. 0704-0188

To average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and on. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Washington, DC 20503.

REPORT DATE

1986

3. REPORT TYPE AND DATES COVERED

Unknown

4. TITLE AND SUBTITLE

Knowledge and Uncertainty

5. FUNDING NUMBERS

DAAB10-86-C-0567

6. AUTHOR(S)

Henry E. Kyburg, Jr.

DTIC
SELECTED
MAY 27 1992
S E D

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-0036

1. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

One purpose -- quite a few thinkers would say the main purpose -- of the seeking knowledge about the world is to enhance our ability to make good decisions. An item of knowledge that can make no conceivable difference with regard to anything we might do would strike many as frivolous. Whether or not we want to be philosophical pragmatists in this strong sense with regard to everything we might want to enquire about, it seems a perfectly appropriate attitude to adopt toward artificial knowledge systems. If it is granted that we are ultimately concerned with decisions, then some constraints are imposed on our measures of uncertainty at the level of decision making. If our measure of uncertainty is real valued, then it isn't hard to show that it must satisfy the classical probability axioms.

14. SUBJECT TERMS

artificial intelligence, data fusion, knowledge and uncertainty
Rational Belief, Probability, Theory and Measurement

15. NUMBER OF PAGES

7

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT

UNCLASSIFIED

20. LIMITATION OF ABSTRACT

UL

KNOWLEDGE AND UNCERTAINTY

Henry E. Kyburg, Jr.

Accession For
NTIS GRAD
DTIC TAB
Unannounced
Justification

By
Distribution/

Availability C
Special
A-1

One purpose -- quite a few thinkers would say the main purpose -- of seeking knowledge about the world is to enhance our ability to make good decisions. An item of knowledge that can make no conceivable difference with regard to anything we might do would strike many as frivolous. Whether or not we want to be philosophical pragmatists in this strong sense with regard to everything we might want to enquire about, it seems a perfectly appropriate attitude to adopt toward artificial knowledge systems.

If it is granted that we are ultimately concerned with decisions, then some constraints are imposed on our measures of uncertainty at the level of decision making. If our measure of uncertainty is real valued, then it isn't hard to show that it must satisfy the classical probability axioms. For example, if an act has a real-valued utility $U(E)$ if event E obtains, and the same real-valued utility if the denial of E obtains, ($U(E) = U(-E)$) then the expected utility of that act must be $U(E)$, and that must be the same as $p*U(E) + q*U(-E)$, where p and q represent the uncertainty of E and $-E$ respectively. But then we must have $p + q = 1$.



1. There are reasons for rejecting real-valued -- i.e., strictly probabilistic -- measures of uncertainty, though not all the reasons that have been adduced for doing so are cogent. One is that these probabilities seem to embody more knowledge than they should: for example, if your beliefs are probabilistic, and you assign a probability of .01 to a drawn ball's being purple (on no evidence), and a probability of .02 to a second ball's being purple on the evidence that the first one is, and regard pairs of balls as "exchangeable", then you should be 99% sure that in the infinitely long run, no more than 11% of the balls will be purple. You know beyond a shadow of a doubt (with probability .99996) that no more than half will be purple. (In fact, we need much less than full exchangeability for this: all we need is that both individual events and pairs of events are treated the same way -- Kyburg, 1968.)

Peter Cheeseman (1985) has given a defense of classical probability, and perhaps would not find even such results as the foregoing distasteful. But it is hard to see how to defend the real-valued point of view from charges of subjectivity. Cheeseman refers to an "ideal" observer, but offers us no guidance in how to approach ideality, nor any characterization of how the ideal observer differs from the rest of us. It is therefore quite unclear what the ideal observer offers us, other than moral support: each of us is no doubt convinced that the ideal observer assigns probabilities just like himself. One man's subjective bias is another man's rational insight.

One defense against charges of subjectivity is to be found in the "convergence" theorems, of which the most famous is de Finetti's (de Finetti, 1937). Roughly: If S is a sequence of trials, each resulting in success or failure, and you and I agree that the sequence is exchangeable, then no matter how divergent our initial views of the probability of

92-13681

92 5 22 001



success -- so long as they are not given by probabilities of 0 or 1 -- and no matter what degree of agreement we seek, there is a number n such that after observing n trials you and I will agree to that degree on the probability of success on the next trial.

Of course each such theorem has a less gratifying counterpart: given any degree of disagreement that is intolerable, and given any n , there exists a degree of initial disagreement such that even after n trials our degree of disagreement about the probability of success on the n plus first trial will be intolerable. And nothing precludes our disagreeing that much to start with.

It could be argued -- and has been -- that subjective probabilities don't vary so dreadfully much, and so in fact subjectivity is a mere hobgoblin. It may be philosophically troubling to those of that turn of mind, but it is of little practical importance.

But I think it can be argued that a small difference in some subjective probabilities can lead to a very large difference in others. Furthermore, it is well known that we all fail to conform to the probability calculus in our degrees of belief. That just means that we have to make some adjustments. Since small differences can lead to large ones, the particular adjustments we make can have large consequences.

There are other ways of representing uncertainty than by real numbers between 0 and 1. If these uncertainties are to be used in making decisions, however, they must be compatible with classical point-valued probabilities. My preference is for intervals, because they can be based on objective knowledge of distributions, and because this compatibility is demonstrable. (Kyburg, 1974)

In what follows, I will sketch the properties of interval-valued epistemic probability, and exhibit a structure for knowledge representation that allows for both uncertain inference from evidence and uncertain knowledge as a basis for decision. Along the way I make some comparisons to other approaches.

2. Probability.

Probability is a function from statements and sets of statements to closed subintervals of $0,1$. The sets of statements represent hypothetical bodies of knowledge. The idea behind $\text{Prob}(S,K) = p,q$ is that someone whose body of knowledge is K should, ought to, have a 'degree' of belief in S characterized by the interval p,q . The cash value of having such a 'degree' of belief is that he should not sell a ticket that returns to the purchaser \$1.00 for less than $100p$ cents, and he should not buy such a ticket for more than $100q$ cents. The relation in question is construed as a purely objective, logical relation.

Every probability can be based on knowledge of statistical distributions or relative frequencies, since statements known to have the same truth value receive the same probability, and every such equivalence class of statements (we can show) contains some statements of the appropriate form. This statistical knowledge may be both uncertain and approximate (we may be practically sure that between 30% and 40% of the

balls are black), but it is objective in the sense that any two people having the same evidence should have the same knowledge.

Classical point-valued probabilities constitute a special case, corresponding to the extreme hypothetical (and unrealistic) case in which K embodies exact statistical knowledge.

The connection between statements and frequencies is given by a set of formal procedures for finding the right reference class for a given statement. The reference set may be multi-dimensional -- the set of urns, each paired with the set of draws made from it. It may be only "accidentally" related to sentence -- as when we predict the act of someone who makes a choice on the basis of a coin toss. What is the right reference class for a given statement S depends (formally and objectively) on what is in K, our body of knowledge. In some cases we can implement a procedure for finding the right reference class.

It is natural to suppose that statistical knowledge in K is represented by the attribution to each reference set of a convex set of distributions -- for example we have every reason in the world to suppose that heads among coin-tosses in general is nearly binomial, with a parameter close to a half. (We have no reason to suppose that the parameter has the real value .49999...). Or we may have good reason to believe that two quantities are uncorrelated in their joint distribution. Or that we can rule out certain classes of extreme distributions. We can know of a certain bent coin that heads will be binomially distributed in sequences of its tosses, with a parameter p at least equal to a half. In a wide range of cases of practical importance, what we can know of the set of distributions is that conditional independence obtains between certain variables. (Judea Pearl has made conditional independence the cornerstone of his constraint propagation approach (Pearl, 1985); conditional independence is what is required to warrant the use of Dempster's rule of combination.

Henceforth, we assume convexity. Here are some immediate results (Kyburg 1961, 1974):

- (1) If $\text{Prob}(S, K) = p, q$ then $\text{Prob}(\neg S, K) = 1 - p, 1 - q$.
- (2) If $(S \text{ \& } T)$ is in K, and $\text{Prob}(S, K) = p_1, q_1$ and $\text{Prob}(T, K) = p_2, q_2$ and $\text{Prob}(T \mid S) = p, q$, then there are numbers in p_1, q_1 and p_2, q_2 whose sum is in p, q . To see that p, q can be a proper subset of $p_1 + p_2, q_1 + q_2$, consider a die that you know to be biased toward the one at the expense of the two, or toward the two at the expense of the one. Reasonable probability for the disjunction, "one or two" would be very close to 1/3, even though the reasonable probabilities for "one" and "two" would be significantly spread above and below 1/6.
- (3) We can show that: given any finite set of sentences, S_i, and a body of knowledge K, there exists a Bayesian function B, satisfying the classical probability axioms, such that for every sentence S in S_i, $\text{B}(S) = \text{Prob}(S, K)$.
- (4) Let KE be the body of knowledge obtained from K when evidence E is added to K. If E is among the finite set of sentences in question, then

there may be no Bayesian function B satisfying both $B(S) = \text{Prob}(S,K)$ and $B(S/E) = \text{Prob}(S,KE)$: classical conditionalization is not the only way of updating probabilities.

(5) The randomness relation is definable, and in fact for one kind of database rules for picking the right reference class have been implemented.

3. Uncertain Knowledge.

One problem that Bayesian and other approaches to uncertainty have is that there is no formal way of representing the acquisition of knowledge. We can represent the having of knowledge (by the assignment of probability 1 to the item), but since there is no way in which $P(S/E)$ can be 1 unless $P(S)$ is already one, conditionalization doesn't get us knowledge. This has been noticed, of course; Cheeseman (1985, p. 1008) simply says, "A reasonable compromise is to treat propositions whose probability is close to 0 or 1 as if they are known with certainty..." But of course it is well known that this cannot be done generally: the conjunction of a number of certainties is a certainty, but the conjunction of a large enough number of certainties in Cheeseman's sense is what he would have to consider an impossibility^o

McCarthy and Hayes (1969) are seduced into following this primrose path, when they suggest (p. 489) "If O_1, O_2, \dots, O_n is a possible deduction, then probably(O_1), ..., probably(O_n) probably (O) is also a possible deduction." This is clearly ruled out, on our scheme -- and even acceptable(O_1), ..., acceptable(O_2) acceptable(O) is ruled out as a consequent of the logical conditional. If we are to formalize uncertain inference at all and not merely the deductive propagation of probabilities, we must somehow accommodate sets of conflicting statements. Purely probabilistic rules of inference do this easily.

We can accommodate Cheeseman's intuition that we should accept what is "practically certain" by considering two sets of sentences in the representation of knowledge. One of them we will call the evidential corpus, and denote by K_e ; the other we will call the practical corpus, and denote by K_p .

We will accept a sentence into K_p if and only if its lower probability relative to K_e is greater than p . The conjunction of two statements that appear in K_p will also appear in K_p only if the conjunction itself is probable enough relative to K_e . Thus K_p will not be deductively closed, though we can prove that if a statement S appears in K_p , and S entails T , T may also appear there because it will have a lower probability greater than that of S . This reflects a natural feature of human inference: we must have reason, not only to accept each premise in a complex argument, but to accept the conjunction of the premises, in order to be confident of the conclusion.

In fact, the uncertain inference that generates K_p from K_e has a number of the desirable features of non-monotonic inference. Add "Tweety is a bird" to K_e , and "Tweety is capable of flight" will appear in K_p exactly because practically all birds fly. In addition, add "Tweety is an ostrich" to K_e , and "Tweety is not capable of flight" will appear in K_p .

In the former case, you should base your decisions on the assumption that Tweety can fly; in the latter, you no longer need worry about that possibility.

But to warrant the detachment that yields the addition of a sentence to our stock of practical certainties, we need more than a mere preponderance of evidence. We don't want to infer that two tosses of a coin will yield one head and one tail just because this is the most likely outcome. Similarly, we don't want to infer that a die will not yield a six: we want to say that the probability of an outcome other than a six is about five sixths.

This is just to say that the level of practical certainty p is exactly what distinguishes (in a given context) sentences that we are willing to bet against from sentences that we take for granted.

We have a picture that looks like this:

```

*****
*      *
* Ke *
*****

```

Uncertain inference: \underline{S} \underline{Kp} iff
 $\underline{\text{Prob}(S,ke)}$ p .

```

*****
*      *
* Kp *
*****

```

It is relative to \underline{Kp} , the practical corpus, that we make our (practical) decisions. It is thus the (convex sets of) distributions -- including conditional distributions -- embodied in the practical corpus that we use in our decision theory.

But there are questions. What is the value of p that we are taking as practical certainty? How do statements get in \underline{Ke} ? What is the decision theory that goes with this kind of structure?

Let us first consider the value of p . Suppose the widest range of stakes we can come up with is 99:1. For example, Sam and Sally are going to bet on some event, each has \$100, and neither has any change. Then a probability value falling outside the range of .01,.99 would be useless as a betting guide. A probability less than .01 would (in this context) amount to a practical impossibility; one greater than .99 would amount to a practical certainty.

The range of stakes can determine the level of "practical certainty" p . What counts as practical certainty depends on context, but in an explicit way: it depends on what's at stake.

How do statements qualify as evidence in \underline{Ke} ? Not by being "certain." It can be argued that anything that was really incorrigible would have to be devoid of empirical content. (The worry about uncertain evidence is not misplaced; it's just misconstrued.) One typical form of evidence statement is this: "The length of x is $\underline{d} + \underline{r}$ meters". Whatever our

readings, these statements are not "certain" -- they admit of error. The same is true of all ordinary observation statements.

So a statement gets into K_e by having a low probability of being in error; equally, by having a high probability (at least e) of being veridical. How high? In virtue of the fact that conjunctions of pairs of statements in K_e appear in K_p , it seems plausible to take $e = (p)^{1/2}$. For a number of technical reasons (Kyburg, 1984) it turns out to be best to construe the corpus containing the theory of error as metalinguistic. This is as one might think: after all, the theory of error concerns the relation between readings -- e.g. numerals written in laboratory books -- and values: the real quantities characterizing things in the real world. For present purposes we need note only that this is not the beginning of an infinite regress. We can maintain objectivity; we can avoid "presuppositions" and other unjustified assumptions.

4. Decision.

It has been objected (Seidenfeld, 1979) that there is no decision theory that is tailored to Shafer's theory of evidential support. Indeed, it is pretty clear that support functions alone would conflict with expected utility. On the other hand, since Shafer's system of support functions is a special case of the representation by convex sets of distributions, we can have very nearly a normal decision theory using Shafer's system. In computing the value of an act, we need to consider not only the support assigned to various states of affairs (corresponding to lower probabilities), but also the plausibilities -- corresponding to the upper probabilities.)

This is true for the more general convex set representation: We can construct an interval of expected utility for each act. A natural reinterpretation of the principle of dominance would take an alternative a1 to dominate an alternative a2 whenever, for every possible frequency distribution, the expectation of a1 is greater than the expectation of a2.

This eliminates some alternatives, but in general there will be a number of courses of action that are not eliminated. What we do here is another matter, one which is certainly worthy of further study. But it seems natural that minimax and minimax regret strategies are appropriate candidates for consideration under some conditions. There may well be others, such as satisficing. And it may even be that the guidance provided by the motto: eliminate dominated alternatives, is as far as rationality alone takes us. Further pruning may depend on constraints that are local to the individual decision problems.

BIBLIOGRAPHY

- Cheeseman, Peter (1985): "In Defense of Probability," IJCAI 1985, II, pp. 1002-1009.
- Finetti, Bruno de (1937): "La Prevision: Ses Lois Logiques, Ses Sources Subjectives" Annales de L'Institute Henry Poincare 7, 1937, pp. 1-68.
- Kyburg, Henry E., Jr. (1968): "Bets and Beliefs," American Philosophical Quarterly 5, pp. 54-63.
- _____ (1961): Probability and the Logic of Rational Belief, Wesleyan University Press, Middletown.
- _____ (1974): The Logical Foundations of Statistical Inference, Reidel, Dordrecht.
- _____ (1984): Theory and Measurement, Cambridge University Press, Cambridge.
- Loui, Ronald P. (forthcoming): "Computing Reference Classes."
- McCarthy, John and Hayes, Patrick (1969): "Some Philosophical Problems from the Standpoint of Artificial Intelligence," Machine Intelligence 4, pp. 463-502.
- Pearl, Judea (1985): "Fusion, Propagation, and Structuring in Bayesian Networks," TR CSD-850022, UCLA, Los Angeles.
- Seidenfeld, Teddy (1979): "Statistical Evidence and Belief Functions", PSA 1978, Asquith and Hacking (eds.).