A Response to "Believing on the basis of evidence"*

Fahiem Bacchus
Department of Computer Science
University of Waterloo
Waterloo, Ontario
Canada, N2L 3G1

Adam Grove
NEC Research Institute
4 Independence Way
Princeton, NJ 08540

Joseph Y. Halpern
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120–6099

Daphne Koller
Computer Science Department
Stanford University
Stanford, CA 94305

1 Introduction

"Believing on the basis of evidence" outlines a program of research Kyburg has pursued over most of his career. The program addresses the difficult problem of how a rational agent can acquire beliefs about the world given the intrinsic uncertainty of empirical facts. This problem is also central to AI. Early AI systems depended simply on deduction for generating beliefs, which can perhaps be viewed as "believing on the basis of proof." Of course, deduction still requires a set of premises and the question of how these are acquired returns one to the same problem. Furthermore, even given the premises, i.e., a knowledge base, deduction is not sufficient to generate all of the conclusions that need to be drawn. This initiated research in non-monotonic inference.

Kyburg addresses the issue of how we can use our information to reason to a set of conclusions that are plausible but not certain. This set of conclu-

*This paper is essentially identical to one that appears in *Computational Intelligence* **10**:1, 1994, pp. 21–25.

sions goes beyond the set of deductive conclusions. His program is certainly ambitious and the approach he suggests meets with varying degrees of success and failure when it comes to particular issues. In the rest of the paper, we examine a few parts of Kyburg's approach.

2 To HEDGE or to hedge?

Kyburg starts out by focusing on what he suggests are two fundamentally distinct ways of dealing with non-deductive inferences. The conclusions drawn by such inferences do not possess the assurance of logical validity, so we are faced with a choice: do we bother to represent their uncertainty or not?

Schema 1 takes the route that we represent the uncertainty. The conclusion of this schemas is "C, hedged" which is quite different from "C." If we wish to characterize such an inference formally it is important to have a language that can represent both "C" and "C, hedged." If like Kyburg, we give "hedged" a probabilistic meaning, then we require a language that can represent probabilities assigned to assertions. Very expressive versions of such languages have been developed [Bac90, Hal90]. In Halpern's [Hal90] language, e.g., we can write "$\Pr(C) = 0.9$" to represent that the degree of certainty, or degree of belief, in "C" is 0.9. What is important here is that the probabilities, or the hedges, are represented inside of the language, i.e., they are not meta-linguistic constructs. A single unified language is much easier to deal with, both conceptually and formally, than language/meta-language combinations.

Schema 2, on the other hand, requires no such representational power. Here we bury the inferential uncertainty in the inference procedure, and the final conclusion "C" is not tagged by its uncertainty. Such an approach is typical of work on non-monotonic logic (see, e.g., [Gin87]).

Do these schema really represent fundamentally distinct approaches to the problem? We believe that the distinction is not sharp. In some cases it reduces simply to a matter of representation.

For example, given a realization of Schema 1 it is quite possible to realize Schema 2 through a simple syntactic modification of the conclusions. That is, Schema 1 generates conclusions of the form "C, hedged," so to realize Schema 2 we could simply take the conclusions generated via Schema 1 and drop the explicit mention of the hedge. Thus we would conclude "C" via

Schema 2 if we can conclude "C, hedged" via Schema 1. Note that this can work even if "C" is a probabilistic assertion of the form "the probability of α is p," as allowed by Kyburg. If Schema 2 is realized in this manner it is clear that its behavior will be determined by the behavior of the underlying Schema 1, and it is not clear that there is any fundamental difference between the two.

Of course this is not the only way in which Schema 2 can be realized. Other realizations need not require an underlying Schema 1. The point is, however, that *as schema* the division between the two approaches seems rather arbitrary.

A good example is Kyburg's own approach, which follows exactly the steps described above. He has a system of evidential probability that generates inferences according to Schema 1, i.e., it generates conclusions with explicit probabilities attached. He then suggests adding a rule of acceptance, based on high probability, that subsequently drops these probabilities and generates the raw conclusions in accord with Schema 2. Note, however, the high probability rule of acceptance, his version of Schema 2, cannot be effected without first generating the probabilities to determine if they are high enough.

This makes his claim that "the first schema represents a perfectly classical deductive inference" rather puzzling. This means that the probability assigned to a conclusion will be a deductive consequence of $\mathcal{BK}$, his background knowledge, and E, his evidence. But then, whether or not the probability of the conclusion is high enough to be accepted will also be a deductive consequence, and all of the conclusions he generates via Schema 2 are thus deductive!

Just because the conclusion is hedged *does not* mean that it follows with complete certainty from the premises. If we examine the case of probability we can see more clearly what is going on. Kyburg states that *given a prior distribution* the probability of a conclusion C is simply the conditional probability of C given $\mathcal{BK} \wedge E$ which is determined deductively from the axioms of probability. This is correct, but it ignores the main point: the choice of a prior distribution is *not* deductive!

For example, in [BGHK95], we present an approach for assigning probabilistic degrees of belief to conclusions using a knowledge base of statistical information. The approach is based on Laplace's *principle of indifference* [Lap20]. There are different realizations of the approach each of which generates a different prior distribution on our beliefs. Although we can certainly provide good motivation for each of these choices of prior, none is a logical consequence of the knowledge we have. Since the choice of prior is not deductive in [BGHK95], neither are the conclusions that we draw.

Kyburg assigns probabilities in a different manner, based on selecting an appropriate reference class, but his rules for choosing the appropriate reference class are no more deductive than our rule for choosing a prior so, again, neither are his conclusions deductive consequences of the knowledge base.

Although he is not very clear on this point, it could be that Kyburg has something more general in mind when he talks about deductive inference. In the paper Kyburg says that any useful way of hedging conclusions should be accompanied by some axiomatization; deductions can be, therefore, completely certain *relative to the axioms*. So if we consider a procedure based on reference classes, as Kyburg does, presumably the axioms refer to rules for choosing the reference class. It seems clear that there can be countless different theories, according to which axioms we favor.

So to attain deductive certainty in this sense, some subjective choice must be made: either a prior distribution or an axiomatization of the acceptable inferences. (By "subjective" we simply mean that it not deductively determined by the given knowledge.) But once such choices are allowed, isn't everything deductive?

We would argue that if is this is in fact the notion of deduction that Kyburg is appealing to then it is too general. With the right set of axioms anything computable can become deductive.

3 Evidential Probability and Reference Classes

After his discussion of the inference schema, Kyburg launches into a discussion of acceptance. However, since his rule for acceptance is simply to accept conclusions with sufficiently high probability, it seems more reasonable to first discuss his mechanism for assigning probabilities. So let us skip ahead to Section 4 of his paper, and discuss evidential probabilities.

Kyburg attempts to assign probabilities to conclusions by using the statistics of an appropriate reference class. For example, if we wish to assign a probability to the conclusion that John will die in the next year, we find some class of individuals to which John belongs and look at the frequency of deaths in that class. So if John is a college professor and 0.1% of college professors die every year we would assign a probability of 0.001 to "John will die in the next year." The notion of reference class, due originally to Reichenbach [Rei49], has the advantage of yielding intuitive results in simple cases, but unfortunately it is highly problematic as a foundation for a general theory. The problem lies in what Kyburg briefly refers to as the "indefinitely large set of possible reference classes." He refers to three principles for choosing among alternate reference classes and claims: "So far as I have been able to tell, no other principles are needed for resolving disagreement between reference classes."

To us, the whole approach of locating reference classes seems flawed. First, in some cases no single reference class is the right one. For example, in the Nixon diamond, Nixon is a member of the reference class of Republicans as well as the class of Quakers. Neither is the "right" class. Given that we still need to assign a degree of belief to our conclusion, it seems appropriate in such cases to find a mechanism for combining the statistics of the alternate reference classes. The notion of finding "the right" class simply does not apply here. Second, the rules used to choose between reference classes are *ad hoc*. Their only justification is that they seem to work in various examples; however, there are well-known examples [Lev80] where the answers they give seem quite unreasonable. Third, the approach is subject to *ad hoc* restrictions. For example, disjunctive reference classes (such as the reference class consisting of the union of Quakers and Republicans) are not allowed [Kyb83]. Hence, if we possess statistical information about a disjunctive class we cannot use it. Fourth, note that the original problem—computing probabilities based on one's knowledge—makes no mention of the notion of reference class. While the notion of reference classes is conceptually appealing, it runs into difficulty when we move beyond simple cases. A great deal of machinery has been introduced in the literature in an attempt to extend the reference class approach to more complex cases. However, it is not clear that this machinery is actually required to solve the original problem.

In recent work [BGHK95] we have investigated an alternate approach to assigning probabilities to conclusions. Like Kyburg our approach assigns probabilistic degrees of belief to conclusions using a knowledge base of statistical information. However, our approach avoids all mention of reference classes, and is instead based on what seems to us a more fundamental notion of indifference. It is able to deal well with the traditional problems, like preferring more specific information, and does not suffer from the *ad hoc* restrictions of the reference class approach. Furthermore, it is able to combine statistical information from many different sources. It simply uses whatever information is expressed in the knowledge base without having to separate that information into appropriate and inappropriate classes. Subtle issues still arise, and more work needs to be done to fully understand how these issues should be resolved. Nevertheless, we feel that this work does clearly demonstrate at least two things. First, there are alternate, and we think better, ways of looking at the problem rather than the approach of locating reference classes. And second, the issues that arise in the generation of probabilistically hedged conclusions are far from being solved.

As we have already pointed out, finding a mechanism for realizing Schema 1, the generation of hedged conclusions, is not a straightforward "deductive" exercise, and probabilistically hedged conclusions cannot be easily obtained by a simple collection of rules for locating appropriate reference classes. In both cases we find Kyburg's discussion of these matters misleading.

4 Acceptance

Now let us return to Kyburg's system of acceptance. In Section 8 he defends the need for a system of acceptance. Acceptance is an issue that has been argued about for a very long time, so it is probably pointless for us to enter into this debate here. Instead let us focus on Kyburg's specific proposal.

Having dismissed Schema 1 as being just deduction, incorrectly as we have argued above, acceptance is the means by which Kyburg seeks to realize Schema 2. Kyburg wants to accept a conclusion if its probability is over some threshold value. This corresponds to throwing away the additional information contained in the precise probability of the conclusion, remembering only that it is "high enough." Hence, the conclusion is no longer represented in its hedged form.

However, because acceptance is based on high probability it is not easy to characterize the set of accepted conclusions. Kyburg's lottery paradox (mentioned in Section 2 of his paper) shows that the accepted conclusions may not be deductively closed, nor necessarily even consistent. Kyburg's solution to this problem is to examine all of maximally consistent subsets of the knowledge base to see if a particular conclusion can be accepted. This seems to defeat the computational advantages of acceptance! And computational advantage is Kyburg's main argument for a system of acceptance.

If one's aim is simply to avoid having to keep track of probabilities, however, there is an alternate mechanism. Instead of removing the probability assigned to the conclusions one can represent the statistical information used to generate these probabilities in an alternate, approximate, form. Instead of using statistical information like "90% of all Quakers are pacifists" one can use approximate statistical information of the form "almost all Quakers are pacifists." In particular, one would represent this as "100-ϵ% of all Quakers are pacifists," and inference would proceed by examining what happens as ϵ approaches zero. This approach is related to, but not coincident with, Geffner and Pearl's ϵ-semantics [GP90]. Here the focus has shifted, instead of ignoring the "hedge" in our conclusions we ignore the "hedge" in our premises.

Now, given that Nixon is a Quaker (and nothing else), instead of concluding that Nixon is a pacifist with probability 0.9 and then dropping that probability, we would conclude with probability 1 that Nixon is a pacifist. In this approach we simply look for conclusions that have probability 1. The end result appears to be similar in both cases: we simply conclude that Nixon is a pacifist. However, the formal characterization is quite different. For example, our approach deals well with certain variants of the lottery paradox, and the conclusions it generates appear to have a far cleaner characterization than those generated by Kyburg's suggested acceptance rule. We have explored this approach in detail in [BGHK93].

This approach also demonstrates once again that Schema 1 and 2 are not necessarily distinct approaches. If we combine ordinary quantitative statistical information with the approximate qualitative information described above, we obtain a single system that can produce both explicitly hedged conclusions, as in Schema 1, as well as hedged conclusions with no explicit hedge, as in Schema 2. The most important thing is that the inferential mechanism is the same; it is just the set of premises that change.

5 Conclusions

As the above comments indicate, we are not convinced that Kyburg's approach is the right one. We have tried to point out some alternate ways of realizing parts of his program. Nevertheless, the importance of the general enterprise is clear. The problem of finding principled ways of drawing plausible conclusions from a knowledge base is a critical one; one simply cannot get very far with deductive conclusions only. Kyburg deserves great credit for keeping us focused on it, and for illuminating some of the subtle issues involved.

References

[Bac90] F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge.* MIT Press, Cambridge, Mass., 1990.

[BGHK93] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. Statistical foundations for default reasoning. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pages 563–569, 1993. Available by anonymous ftp from logos.uwaterloo.ca/pub/bacchus or via WWW at http://logos.uwaterloo.ca.

[BGHK95] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. Reasoning with noisy sensors in the situation calculus. In *Proc. Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, pages 1933–1940, 1995.

[Gin87] M. L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann, San Francisco, 1987.

[GP90] H. Geffner and J. Pearl. A framework for reasoning with defaults. In H. E. Kyburg, Jr., R. Loui, and G. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 245–26. Kluwer Academic Press, Dordrecht, Netherlands, 1990.

[Hal90] J. Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.

[Kyb83] H. E. Kyburg, Jr. The reference class. *Philosophy of Science*, 50(3):374–397, 1983.

[Lap20] P. S. de Laplace. *Essai Philosophique sur les Probabilités.* 1820. English translation is *Philosophical Essay on Probabilities*, Dover Publications, New York, 1951.

[Lev80] I. Levi. *The Enterprise of Knowledge.* MIT Press, Cambridge, MA, 1980.

[Rei49] H. Reichenbach. *The Theory of Probability.* University of California Press, Berkeley, 1949. This is a translation and revision of the German edition, published as *Wahrscheinlichkeitslehre*, in 1935.