Vol. III. *Edited by* H. Feigl and G. Maxwell. University of Minnesota Press, Minneapolis, MN, pp. 98–169.
KYBURG, H. E., JR. 1987. Objective probabilities. Proceedings of the International Joint Conference on Artificial Intelligence, Milan, Italy, pp. 902–904.
MCDERMOTT, D. 1987. A critique of pure reason. Computational Intelligence, **3**: 151–160.
MORGAN, C. G. 1971. Hypothesis generation by machine. Artificial Intelligence, **2**: 179–187.
NILSSON, N. J. 1986. Probabilistic logic. Artificial Intelligence, **28**: 71–87.
SCHUBERT, L. K. 1978. On the representation of vague and uncertain knowledge. COLING 78, Bergen, Norway.
——— 1987. Remarks on "A critique of pure reason" by Drew McDermott. Computational Intelligence, **3**: 210–214.

Comments on *An inquiry into computer understanding* by Peter Cheeseman

GLENN SHAFER
School of Business, University of Kansas, Lawrence, KS 66045, U.S.A.

Comput. Intell. 4, 121–124 (1988)

I share Peter Cheeseman's conviction that probability ideas have much to offer workers in artificial intelligence. I fear, however, that the tone of Cheeseman's article will alienate rather than persuade those who are not yet fond of probability. I myself am offended by Cheeseman's dogmatic dismissal of non-Bayesian methods of probability judgment.

In the following comments, I will spell out some of the many points where I disagree with Cheeseman, and, more importantly, I will try to supply a broader perspective on the issues and problems that he raises. Most of the issues have been the subject of long debates, and most of the problems have been the subject of considerable study. Readers who are encountering the issues and problems for the first time deserve some signposts pointing to the existing literature.

The Bayesian controversy

The choice between Bayesian and non-Bayesian uses of the mathematical theory of probability has been a subject of debate for over a century. The Bayesian method, or the method of inverse probability, as it was then called, was championed in the 19th century by Laplace. But it also had many critics, and their criticisms were so effective that by the early 20th century non-Bayesian probability methods were dominant in statistical work. The past three decades have seen a revival of interest in Bayesian ideas, a remarkable flowering of theoretical Bayesian work, an intensification of the debate between Bayesians and non-Bayesians, and an increasing use of the Bayesian method in engineering work and decision analysis. The Bayesian view remains, however, a minority view. In most scientific work, non-Bayesian methods are used to analyze statistical evidence.

Where should the reader turn for a good perspective on the Bayesian controversy? One good starting point would be *Comparative Statistical Inference* by Barnett (1973), a book that tries to lay out the issues without taking sides. Another reasonable, but older, starting point would be *The Foundations of Statistical Inference*, edited by Savage (1962); this small book records a symposium involving prominent Bayesian and non-Bayesian statisticians. For a perspective on the richness and vigor of current work on Bayesian methods, the reader might turn to two volumes edited by Bernardo *et al.* (1980, 1985), *Bayesian Statistics* and *Bayesian Statistics 2*. For information on the philosophical underpinnings of the Bayesian revival, the reader might turn to *Studies in Subjective Probability*, edited by Kyburg and Smokler (1980).

Interpretations of probability

The mathematical theory of probability derives, historically, from the study of games of chance (Hacking 1975; Shafer 1978). Numerical probabilities that obey the rules of this theory clearly do exist in games of chance. But why should such numbers exist in other domains?

There is a long history of argument on this point. Savage (1954), a leader of the 20th century revival of Bayesian ideas, classified those taking part in the argument into three broad groups: objectivists, personalists, and necessarians. The *objectivists* (or frequentists) believe that numerical probabilities can be assigned only in situations that are like games in chance in that repetition under fixed conditions is possible. The *personalists* hold that rationality requires a person to assign a numerical probability to every proposition and every event, but that this number should be interpreted as the person's betting rate rather than as a property of the world. The *necessarians* hold that numerical probabilities always exist, and that they measure "the extent to which one set of propositions, out of logical necessity and apart from human opinion, confirms the truth of another" (Savage 1954).

Both personalists and necessarians are Bayesians. Most of the participants in the 20th century Bayesian revival, including Savage, have been personalists.

Laplace could be classified as a necessarian. The 19th-century frequentists criticized Laplace's necessarian view for its lack of empirical content. How can you say that something exists if you have no way of measuring it? The personalists have dominated the 20th century Bayesian revival because necessarians have never been able to give a good answer to this question. The personalists do have an answer. They say that their conception of probability is just as empirical as the frequentist conception; probability is a property of a person, and it can be measured by observing the person's behavior as he chooses among gambles.

Printed in Canada / Imprimé au Canada

In practice, necessarians try to determine probabilities by using linguistic or mathematical symmetries. If there are five possibilities, and we know nothing else, each possibility should have probability one-fifth. Unfortunately, there are often competing symmetries (Shafer 1976).

Cheeseman, I gather, is a necessarian. It is of some interest to note that he was trained in physics. The necessarian view seems, in general, to be more attractive to physicists than to statisticians. Harold Jeffreys, E. T. Jaynes, R. T. Cox, Erwin Schrodinger, and Judea Pearl were all trained in physics. Why are physicists less influenced by the empiricist criticism of the necessarian view than statisticians? The obvious explanation is that probabilities are calculated from symmetries in physics textbooks. Physicists should remember, however, that probabilities in physics have a frequentist interpretation. Experiments are available to arbitrate between conflicting symmetries. Photons obey Bose–Einstein statistics, and electrons obey Fermi–Dirac statistics (Feller 1968).

My own stance is neither frequentist, personalist, nor necessarian. I take what I call a *constructive* view. I believe that when we make a numerical probability judgment, we are subjectively measuring the strength of our evidence by comparing it to the evidence we have in a game of chance where the chances are known. When we combine probabilities to calculate new probabilities, we are making an argument by analogy, where the analogy is between the problem at hand and a certain game of chance. When we use Bayesian methods, we are drawing a very direct analogy; when we use belief functions (Shafer 1976) or non-Bayesian statistical methods, we are drawing more subtle analogies. Because of the directness of the analogy used by the Bayesian method, Bayesian numerical probabilities (degrees of belief) always follow the rules for chances in games of chance. Because of the indirectness of the analogies used by belief-function and non-Bayesian statistical methods, these methods can produce degrees of belief that do not follow those rules.

The constructive view differs from the frequentist view, because it allows for the possibility that the analogy to games of chance will be convincing even though we are not working with frequencies. It also differs from the personalist and necessarian views, because it allows for the possibility that the analogy to games of chance will not be convincing, and that consequently we will be unable to assign probabilities in our problem.

For more on the personalist view, see Savage (1954) and Lindley (1972). For more on the necessarian view, see Jeffreys (1961). For more on the constructive view see Shafer (1981, 1986*b*) and Shafer and Tversky (1985).

Cox's argument for Bayes

Modern Bayesians, both personalists and necessarians, are fond of formal arguments supporting their view that numerical degrees of belief should follow the same rules chances follow in games of chance. On the personalist side, such arguments have been given by Ramsey (1931), de Finetti (1937), Savage (1954), and Lindley (1982). On the necessarian side, we can cite Jeffreys (1961), Schrodinger (1947), and Cox (1961).

I have made my case against the personalist arguments elsewhere; see Shafer (1986*b*) and Shafer and Srivastava (1987). Here I will comment on Cox's necessarian argument, which Cheeseman cites.

Cox gave a list of qualitative rules for numerical degrees of belief, and he proved that these qualitative rules are equivalent to the usual numerical axioms for Bayesian probability. I find this valuable, because it makes clearer the assumptions we are making when we use the Bayesian method. But does it provide much of an argument for using Bayesian methods? Is there any particular reason for adopting Cox's qualitative rules, aside from the fact that they do lead to the Bayesian numerical rules?

I think not. Indeed, I believe some of Cox's rules would be incomprehensible if we did not have the Bayesian numerical rules in the back of our minds to help us understand what is going on. The fifth rule Cheeseman lists, hypothetical conditioning, is a good example. What reason would we have for adopting this rule, and why would we even think that there exists, for every pair of propositions, a well-defined number called "the belief in the first proposition given the second," if we did not have in the back of our minds either the standard numerical rule $P(A \cap B) = P(A)\,P(B \mid A)$ or else the picture of games of chance from which it derives?

The sixth rule Cheeseman lists, complementarity, is clearly violated by the theory of belief functions (the Dempster–Shafer theory). In that theory, we are allowed to assign degree of belief zero to both a proposition and its negation, so as to represent the fact that we have no evidence on either side. One of these degrees of belief can later go up without the other, which is already zero, going down. What is wrong with this?

Following Horvitz *et al.* (1986), Cheeseman calls Cox's qualitative rules "intuitive." But today's intuition is simply yesterday's theory. Complementarity is intuitive only because we have learned the Bayesian theory. Its violation becomes equally intuitive once we have learned the belief-function theory.

The third rule that Cheeseman lists, completeness, goes against my whole constructive philosophy, but I do not see that it differentiates between Bayes and belief functions. If we do not care whether the numbers we give are convincing, we can use either formalism to give a degree of belief for anything we want, regardless of how little evidence we have to support it.

The fourth rule that Cheeseman lists, context dependency, is very weak. I fail to understand why Cheeseman thinks it is violated by the belief-function approach, or, for that matter, by any other approach.

Bayesian and Non-Bayesian statistics

Consider a possibility biased coin. If we tossed it many, many times, we would find out whether it is biased or not. In fact, we would find out its true probability for coming up heads. If we toss it only a few times, we will learn less, but we may be able to make a guess or judgment about the true probability of heads. If, for example, we toss it 100 times and get 45 heads, we might be fairly certain that the true probability of heads is between 35 and 55%.

Mathematical statistics has traditionally been concerned with problems of this type—problems where probability is used to model repeatable experiments. A "model," in this context, is a class of probability distributions. We assume that if we were to repeat the experiment many times, the outcomes would follow one of the distributions in the model. But we do not know which one. We actually repeat the experiment only a few times, and we use the resulting outcomes to judge or guess which probability distribution is the correct one.

The controversy between Bayesian and non-Bayesian methods developed in this statistical context. The difference between Bayesians and non-Bayesians in this context is that Bayesians supply prior probabilities for which probability distribution in the model is correct (a Bayesian might, for example, have a prior probability of 20% for the proposition that the true probability of heads falls between 35 and 55%), while non-Bayesians do not. The model is taken for granted by both sides.

As Cheeseman notes, Bayes' theorem uses two ingredients, the likelihood and the prior probability distribution. The likelihood comes from the model, so non-Bayesians object only to the second ingredient, the prior probability distribution.

It is important, however, to realize that moving outside of the traditional domain of mathematical statistics usually means moving outside the domain where models can be taken for granted. In a Bayesian analysis outside the traditional domain of mathematical statistics, the "likelihoods" are usually just as suspect—and just as subjective—as the "prior probabilities." Cheeseman overlooks this point when he asserts, without a shred of justification, that it is generally easier to find $P(E \mid H, c)$ than $P(H \mid E, c)$.

Recognizing the subjectivity of likelihoods outside the statistical domain is a first step towards a broader, constructive understanding of Bayes (Shafer and Tversky 1985). According to this constructive understanding, a Bayesian analysis need not involve Bayes' theorem. But it must involve a deliberate design—a plan specifying what numerical and nonnumerical judgments are to be combined to arrive at the probabilities of primary interest.

Noninformative priors

Cheeseman's untroubled faith in the availability of objective or noninformative priors should be balanced by some reference to the considerable literature on the difficulties and contradictions to which supposedly noninformative priors can lead. See Chap. 10 of Cox and Hinkley (1974). The fact that most Bayesian statisticians are personalists rather than necessarians is partly due to these difficulties and contradictions.

The principle of stable estimation

In the statistical problem, it is often pointed out that the prior distribution makes less and less difference as more and more independent outcomes of the experiment are observed. Savage called this the principle of stable estimation (1963). It is important to remember, however, that this principle, along with the other principles we learn from Bayesian statistics, must be applied gingerly outside the statistical domain. It may apply, as Cheeseman suggests, to the case of a Martian observing the heights of humans. But it may fail in cases where there are few straightforward repetitions.

Model choice and induction

In Sect. 4 of his article, Cheeseman asserts that Bayesian inference is the solution to the problem of induction and model choice. He asserts, moreover, that statisticians, engineers, and physicists have been using it for this purpose for years.

It is true that Bayesian methods can sometimes be used to discriminate among models, but the wholesale application of Bayes to the problem of model choice leads to difficulties and paradoxes (Shafer 1982). Some of the most thoughtful Bayesian statisticians believe that the appropriate division of labor between Bayesian and non-Bayesian methods is to use Bayesian methods for estimation within models and non-Bayesian methods for choice between models. This view has been forcefully advanced, for example, by Box (1980), probably the most widely cited living statistician.

The problem of small worlds

Cheeseman raises the question whether utility, like probability, should depend on context. There is a considerable history of thought about this problem, going back at least to the discussion of small worlds by Savage (1954). For further discussion by a number of authors, see Shafer (1986*b*). For yet further references, see Fishburn (1981).

Paradoxes of conditional probability

Cheeseman gives an example involving cards where conditional probability can lead one astray. His conclusion is that Bayesian inference should be applied with care. I would draw a stronger conclusion. I believe that examples of this type show that there are limits to how far Bayesian inference can be extended outside the domain of experiments whose possible outcomes are known in advance. For further discussion and references to a wider literature, see Shafer (1985).

Black and brown ravens

Cheeseman's application of Bayesian inference to the question whether ravens are black is precisely the kind of application that was the subject of such ridicule in the 19th century. It is analogous to Richard Price's calculation, in 1763, of the probability that the sun will rise again (Pearson and Kendall 1970). For a good discussion of formulas of the type used by Cheeseman and Price, see Good (1965).

In reality, the question whether ravens are black will come down to whether or not brown ravens should really be classified as ravens. Probabilities will be involved, but not simple formulas.

Conclusion

Probability is a broader and more flexible tool than Cheeseman's article might suggest. It has much to offer AI.

One of the most important potential contributions of probability is the guidance it can give us in design. Any complicated probability argument, whether it is constructed by a statistician, a lawyer, or an expert system, must be based on a design (Shafer 1986*a*, 1987; Shafer and Tversky 1985). Readers who want to see how probability theory can help in design in AI should consult the work of Judea Pearl, especially Pearl (1986).

BARNETT, V. 1973. Comparative statistical inference. John Wiley & Sons, London, England.

BERNARDO, J. M., DE GROOT, M. H., LINDLEY, D. V., and SMITH, A. F. M. *Editors*. 1980. Bayesian statistics. Proceedings of the 1st International Meeting, University Press, Valencia.

——— 1985. Bayesian statistics 2. Proceedings of the 2nd International Meeting. University Press, Valencia.

BOX, G. E. P. 1980. Sampling and Bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society, Series A, **143**: 383–430.

COX, R. T. 1961. The algebra of probable inference. The John Hopkins Press, Baltimore, MD.

COX, D. R., and HINKLEY, D. V. 1974. Theoretical statistics. Chapman and Hall, London, England.

DE FINETTI, B. 1937. La prevision: ses lois logiques, ses sources subjectives. *Translated* as Foresight: its logical laws, its subjective sources. *In* Studies in subjective probability. *Edited by* H. E. Kyburg, and H. E. Smokler. Robert E. Krieger Publishing Company Inc., New York, NY. pp. 59–118.

FELLER, W. 1968. An introduction to probability theory and its applications. John Wiley & Sons, Inc., New York, NY.
FISHBURN, P. C. 1981. Subjective expected utility: a review of normative theories. Theory Decision, **13**: 139–199.
GOOD, I. J. 1965. The estimation of probabilities. The MIT Press, Cambridge, MA.
HACKING, I. 1975. The emergence of probability. Cambridge University Press, London, England.
HORVITZ, E. J., HECKERMAN, D. E., and LANGLOTZ, C. P. 1986. A framework for comparing alternative formalisms for plausible reasoning. Proceedings of the 5th National Conference on Artificial Intelligence, Philadelphia, PA, pp. 210–214.
JEFFREYS, H. 1961. Theory of probability. Oxford University Press, London, England.
KYBURG, H. E., and SMOKLER, H. E. *Editors.* 1980. Studies in subjective probability. Robert E. Krieger Publishing Company Inc., New York, NY.
LINDLEY, D. V. 1972. Bayesian statistics, a review. Society for Industrial and Applied Mathematics, Philadelphia, PA.
——— 1982. Scoring rules and the inevitability of probability. International Statistical Review, **50**: 1–26.
PEARL, J. 1986. Fusion, propagation and structuring in Bayesian networks. Artificial Intelligence, **29**: 241–288.
PEARSON, E. S., and KENDALL, M. G. 1970. Studies in the history of statistics and probability. Hafner Publishing Company, Darien,
RAMSEY, F. P. 1931. Truth and probability. *In* Studies in subjective probability. *Edited by* H. E. Kyburg and H. E. Smokler. Robert E. Krieger Publishing Company Inc., New York, NY, 1980, pp. 227–230.
SAVAGE, L. J. 1954. The foundations of statistics. John Wiley & Sons, Inc., New York, NY.
——— *Editor.* 1962. The foundations of statistical inference. John Wiley & Sons, New York, NY.
SAVAGE, J., EDWARDS, W., and LINDMAN, H. 1963. Bayesian statistical inference for psychological research. Psychological Review, **70**: 193–242.
SCHRODINGER, E. 1947. The foundation of the theory of probability-I and II. Proceedings of the Royal Irish Academy, **51**: 57–66 and 141–150.
SHAFER, G. 1976. A mathematical theory of evidence. Princeton University Press, Princeton, NJ.
——— 1978. Non-additive probabilities in the work of Bernoulli and Lambert. Archive for History of Exact Sciences, **19**: 309–370.
——— 1981. Constructive probability. Synthese, **48**: 1–60.
——— 1982. Lindley's paradox. Journal of the American Statistical Association, **77**: 325–351.
——— 1985. Conditional probability. International Statistical Review, **53**: 261–277.
——— 1986*a*. The construction of probability arguments. Boston University Law Review, **66**: 799–816.
——— 1986*b*. Savage revisited. Statistical Science, **1**: 463–501.
——— 1987. Probability judgment in artificial intelligence and expert systems. Statistical Science, **2**: 3–44.
SHAFER, G., and SRIVASTAVA, R. 1987. The Bayesian and belief function formalisms. Working Paper No. 189, School of Business, University of Kansas, Lawrence, KS. To appear in Auditing: Journal of Theory and Practice, 1988.
SHAFER, G., and TVERSKY, A. 1985. Languages and designs for probability judgment. Cognitive Science, **9**: 309–339.

Comments on *An inquiry into computer understanding* by Peter Cheeseman

DAVID J. SPIEGELHALTER
Medical Research Council Biostatistics Unit, 5 Shaftesbury Road, Cambridge, England

Comput. Intell. 4, 124–125 (1988)

I am in almost entire agreement with this paper. Cheeseman is performing a great service in trying to nail the idea that probabilities exist outside of context, and it has been recognized for many years within the established Bayesian school of statistics that the conditioning events should always be, at least implicitly, acknowledged: "Our fundamental idea will not be simply the probability of a proposition p, but the probability of p on data q" (Jeffreys 1961, p. 15) and "all probabilities are necessarily conditional" (Lindley 1971, p. 30). Furthermore, Cheeseman emphasizes the essential subjectivity of probability assignments—they are not properties of the event, but Your opinion concerning that event, where "You" is used in the sense of de Finetti (1974, p. 27) as the individual whose state of uncertainty is of interest. De Finetti states in his Preface that "PROBABILITY DOES NOT EXIST" (his capitals) to push home the essential lack of objectivity. Indeed, it is best to avoid the phrase "probability *of* X" altogether, and use "probability *for* X" to emphasise the dependence on context and observer.

Furthermore, in contrast to Cheeseman's claim, it is quite reasonable in decision theory to have values or utilities of consequences depend on context, such as other feasible options and outcomes, regret from previous incorrect decisions, and so on. In a medical setting, such considerations may be very important (Llewellyn-Thomas *et al.* 1982), although they should be distinguished from the problem in *assessment* of values when wording of the questions can radically effect the responses (McNeil *et al.* 1982).

The AI community may warm neither to a precise numerical basis for reasoning, nor to the essential subjectivity of the numbers. On the first point, there are increasing developments allowing the use of imprecise probabilities yet remaining within strict probabilistic calculus (Spiegelhalter 1986). The second point seems incongruous—why is it reasonable to use subjective, heuristic structure as a basis for a program, and yet balk at subjectivity in the numerical assignments made on that structure? By acknowledging essential subjectivity, the arguments about *whose* judgments are being represented—the iden-

Printed in Canada / Imprimé au Canada