Reply to Cheeseman's "An Inquiry into Computer Understanding"

This paper covers a fairly wide range of issues, from a basic review of probability theory to the suggestion that probabilistic ideas can be used to solve the learning problem. Much of the material is therefore quite old hat, and I will limit my response to the portion of the paper that strikes me as original. In this regard, I believe Cheeseman's main points to be the following:

- The repeated claim that, "McDermott has shown that a direct translation of commonsense reasoning into logical form leads to unsurmountable difficulties" [Abstract, and following].
- A series of suggestions [Section 2] that probabilities are a good idea, because:
  - It is important to be able to label statements as other than simply unquestionably true or unquestionably false,
  - These labels should indicate something about the contexts in which they are valid, and
  - These labels should include numeric information.
- That the assignment of context-dependent labels can help solve the problem of referential opacity [Section 2].
- That the priors problem can be solved using the existence of "objective priors available for most of the commonly encountered situations" [Section 2] such as Jeffrey's priors [Section 3].
- That the use of probabilities solves the learning problem [Section 4].
- That "logic is just a special case within probability theory, but the converse is not true." [Section 5].
- That probabilities can be used to solve the ravens paradox and related difficulties [Section 5].
- That a randomly selected passage from [3] deals with probabilistic information [Section 6]; Cheeseman attempts to infer the ubiquity of probabilistic reasoning from this.

I guess I can manage to find something to respond to in this.

McDermott's paper Cheeseman's remark that, "McDermott has shown that a direct translation of commonsense reasoning into logical form leads to unsurmountable difficulties," is simply false. I am reluctant to enter, at this late date, into the debate on McDermott's article, but it seems to me that what McDermott was arguing (and I agree with him) was that none of the existing methods suggested for nonmonotonic reasoning is satisfactory. McDermott appears to find this disheartening, but nowhere does he make the claim that the logical approach is doomed. He argues that its lack of success makes it suspect, perhaps, and that alternatives (such as proceduralism, and probabilistic methods) should be *tried*, but nowhere does he seem to say that the logicists' approach is fundamentally without hope. And how could he? Simply because no one has managed to make it work *yet* is slim evidence from which to conclude that no one will in the future.

Additional truth values Here, I am at least partially in agreement with Cheeseman's views. I have been suggesting for some time [2, and subsequent papers] that it is important to be able to label sentences more descriptively than simply as "true" or "false." In many cases, it is useful for these labels to include information regarding the context or contexts in which a particular conclusion is valid; deKleer's ATMS's [1] are a typical example. But there is a substantial computational cost associated to maintaining these labels, and I see no justification for the belief that it is *always* right to include contextual information in the truth values.

I also see very little evidence to support Cheeseman's claim that there should be a *numerical* component to the truth value. Certainly this will be appropriate in many situations. But the objections to the probabilistic approach really do have merit: what, for example, is the probability that an astronaut walking on the surface of Mars will encounter a lion? McCarthy has argued that default logics (if we can ever get them to work) will treat the probability of a statement such as this one simply as ϵ, which seems a more reasonable approach than the assignment of a specific numeric value. Once again, it seems that Cheeseman is jumping from the observation that an approach is useful in some instances to the conclusion that it should be used all the time. He does point us to a reference (his paper, "In defense of probability") that addresses this issue, but this seems too crucial a point to be skipped over so lightly in the current work.

Referential opacity Cheeseman argues here that, "if all real world propositions are statements of belief, as required in probability theory, then ... the problem of referential opacity never arises."

This argument is made without support. There is probably a germ of truth to what Cheeseman is saying, in that it is possible to overcome some of the difficulties of referential opacity by keeping track of the believer of any particular sentence. But this is simply an argument for maintaining contextual information in the truth values – and only an argument for doing so in special cases. Once again, I see no justification for Cheeseman's suggestion that the problem of referential opacity is in some way an argument for the use of probability theory.

Priors Here, too, I find Cheeseman's claims unconvincing. The sharpest one I can find is the claim from Section 3 that a Martian, faced with the problem of determining a prior probability distribution for the heights of Earthlings, could use the "truely [sic] noninformative prior" of $(1/x)dx$, "based only on the information that height is always positive and that we want the final answer to be independent of scale." What is x here? Height? Volume? Is the prior probability supposed to be noninformative with respect to the height of Earthlings, or with respect to their volume? (After all, the size of land creatures is limited much more by volume than by height.) The "priors problem" is not that one cannot write down a probability distribution that is noninformative with respect to some particular measurement, but that it is *impossible* to produce one that is noninformative with respect to *all* possible measurements – and that deciding which measurement should be unbiased is a difficult and open problem.

Learning Cheeseman tells us that, "What makes the untutored learning problem difficult (at least for current ad hoc AI methods) is that the number of classes (if any) and their definition is unknown." Nothing could be further from the truth. What makes learning difficult is that the search space is simply too large for any sort of exhaustive search to be practical, and that the nature of the search space is such that heuristic techniques such as hill climbing are unlikely to be effective.

The approach Cheeseman suggests to the learning problem does not address these difficulties at all. In fact, the procedure he suggests for traversing the search space is simple hill climbing; Cheeseman defends this by arguing that, "a search for the global maximum is usually computationally too expensive." He's right, of course. But progress in learning will be made when people find ways to effectively reduce this computational cost, and the application of probabilities to the problem makes no contribution in this regard.

The learning argument also lacks the support of examples of the method at work, although Cheeseman once again points us elsewhere to find them (this time, to his paper "Automatic discovery of optimal classes"). As with the statement that numeric truth values are to be preferred over symbolic ones, this is simply too important a point for us to be expected to accept it on faith.

Logic and probability Cheeseman would also have us believe that, "Probability can be regarded as more fundamental than logic, because logic is just a special case within probability (where all probabilities are 0 or 1), but the converse is not true." Remarkably, the sentence that precedes this one begins, "Similarly, although probability can be embedded in logic ..." Cheeseman argues that probability is not a special case of logic because the embedding to which he refers adds too much baggage to the logical approach for the result to be accurately called "logic". The same thing can be said of probability where all of the probabilities are 0 or 1, however: here, probability theory adds a tremendous amount of baggage (the dependence on contexts, for example) not needed by the logicists.

Both probability theory and first-order logic are universal, and each is capable of describing the other. The suggestion that either approach has some fundamental advantage over the other is simply misguided.

The ravens paradox Cheeseman goes on to suggest that a probabilistic approach avoids the ravens paradox. In reality, however, he does not discuss the ravens paradox at all, but instead investigates another difficulty of his own invention.

The ravens paradox is the following: Suppose that we denote the ravens sentence, "All ravens are black," by r. Now after accumulating some amount of evidence A, we conclude that r has some probability x:

$$p(r|A) = x.$$

In other words, the probability that all ravens are black, given the observed evidence, is x.

Now consider the companion statement, "All non-black things are not ravens," which we denote by $\overline{r}$. Since r and $\overline{r}$ are equivalent, we must also have

$$p(\overline{r}|A) = x,$$

by virtue of the "consistency" condition adopted by Cheeseman himself in Section 2 of the paper.

This is the ravens paradox: evidence for the statement that all ravens are black is also evidence for the statement that all non-black things are not ravens. Cheeseman attempts to sidestep this difficulty by arguing that we are more interested in the conditional probability

$$p(\texttt{black}(x)|\texttt{raven}(x))$$

than we are interested in the probability of the conditional r. But if, as in the original ravens paradox, we really are concerned with the probability of the conditional, Cheeseman's arguments do not apply. Once again, he is using probabilities to answer a question other than the one of interest.

Probabilistic information in English Finally, Cheeseman examines a passage from Nilsson's introductory text on AI. The sentence that he cites as probabilistic is the following:

> To inform a control system completely about the problem domains of interest in AI typically involves a high-cost strategy, in terms of the storage and computations required.

Cheeseman says of this sentence, "This is an example of probabilistic information where the reader is informed of a *likely* consequence of the model under particular circumstances ..."

I see no probabilistic information here. The reader is being informed of a *default* rule. If anything, the form of Nilsson's sentence is an argument *against* the probabilistic approach, since Nilsson seems to feel it sufficient to tell his reader of the presence of the default without providing any way to assign a specific numeric measure of certainty to the conclusion.

Summary In sum, I can find very little with which to agree in Cheeseman's paper. I do agree that probabilities can be used to solve his version of the ravens paradox, and that they can be used to evaluate points in the search space encountered in learning. But these issues are not terribly interesting.

Most of Cheeseman's other points strike me as either unsupported or simply wrong. The single exception is the claim that truth values should be chosen from a set containing more than two elements. But I see little intrinsic connection between this claim and probabilistic issues.

References

[1] J. de Kleer. An assumption-based truth maintenance system. *Artificial Intelligence*, 28:127–162, 1986.

[2] M. L. Ginsberg. Multi-valued logics. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 243–247, 1986.

[3] N. J. Nilsson. *Principles of Artificial Intelligence*. Tioga, Palo Alto, 1980.