Kyburgian Acceptance: A Rejection, Hedged

Michael P. Wellman
University of Michigan
Department of EECS
Ann Arbor MI 48109
wellman@engin.umich.edu

Appeared as a *Taking Issue* response to Henry E. Kyburg, Jr., Believing on the basis of the evidence, *Computational Intelligence* **10**(1), 1994.

There seems to be something about probability that offends some people's (not necessarily Kyburg's) sense of taste. Not being of this aesthetic persuasion, I can only speculate that assigning degrees of belief to propositions is seen as a cop-out; a true believer should either adopt a proposition or not, eschewing compromise. Those who would reject probability need not deny uncertainty, however. Instead, they argue, one can accommodate the inevitable incompleteness in immediate knowledge by an appropriate belief revision policy.

Philosophers have assigned the term *acceptance* to this notion of categorically adopting an uncertain belief and acting as though it were the case. That acceptance is semantically elusive and technically complicated is evidenced by the volume of philosophical debate on the subject as well as the mass of AI research on nonmonotonic logic aimed at working it out computationally. That this AI work is essentially attacking the problem of acceptance is not often explicitly recognized, and Kyburg's clear exposition of this point is one of his essay's important contributions. The other, of course, is his own proposed account of acceptance.

In taking issue with Kyburg, I attempt first to explain why we might want a theory of acceptance. My claim is that our quest can and should be justified on pragmatic computational grounds, rather than those purely philosophical (I'm not sure exactly what this means, but I don't identify it with "frivolous") or aesthetic. Next, I argue that we cannot have acceptance, at least not in terms Kyburg seems to want it. Finally, I suggest some ideas that would have to be addressed in order to obtain an account of acceptance

worth accepting.

Unfortunately, Kyburg's proposed acceptance approach comes bundled with all sorts of other baggage, including his own scheme for probability based on interval measures derived from statistical *reference classes*, as well as the general idea of *evidential probability*. These are ideas that Kyburg has certainly given a fair run for their money, and it is only natural that his acceptance theory would incorporate them. But since the issues they address are somewhat orthogonal to acceptance per se, I am afraid they confound the main points. In my response I will attempt to focus on the essential stance on acceptance, and defer taking issue with other elements of the Kyburg program to some other forum.

1 The Virtues of Acceptance

Kyburg alludes to some potential computational disadvantages of the pure probabilistic approach, but is careful not to make specific claims. As he clearly understands, the comparison is quite difficult to draw in the general case. But without a precise understanding of the nature of computational benefits, we have only our taste to discriminate among technically valid approaches.

If there is an inherent advantage to categorical acceptance over expressly maintaining degrees of belief, it must be that acting for all purposes as though something is the case narrows the set of contingencies or factors that need be considered for relevant decision problems. This seems obviously true, but it appears equally obvious that neglecting contingencies or factors, if they are real, can degrade the quality of the decisions produced. Assessing the magnitudes of loss on either side of this equation is a problem of cost-benefit analysis, and there is every reason to expect the result to depend entirely on the specific context.

What is not obvious is whether in relaxing categorical acceptance by qualifying the "acting for all purposes" to include hedges for consistency (in other words, to make acceptance nonmonotonic) we retain any of the computational advantages. For example, one of the main representational and computational advantages of deterministic over probabilistic relationships are that the former are more modular. We can make this assertion precise in a specific context like dependency-graph formalisms (e.g., probabilistic

networks), where deterministic relationships induce greater separability than probabilistic ones [2]. Probabilistic relationships are inherently non-modular in that the overall degree of belief in a proposition potentially depends on the total set of evidence, whereas in a deterministic system one line of proof is sufficient to render all other evidence irrelevant. Nonmonotonic systems do *not* possess this property—it is only with respect to the entire set of evidence that we can be sure whether a proposition is to be believed. More generally, it appears that much of the additional complexity of uncertain reasoning is inherent in the uncertain nature of relationships themselves; we have no evidence that their expression in probabilistic versus other terms affects the complexity of reasoning.

So we see that categorical acceptance, where we act for all purposes and immutably as though something is the case, can truly simplify reasoning. The case for acceptance based on a nonmonotonic logic (like Kyburg's) is far less clear.

2 Limitations of Acceptance as Practical Certainty

By *practical certainty* I refer to an approach based on accepting propositions believed with probability exceeding some threshold. This I take to be the central feature of Kyburg's acceptance proposal, ignoring the details about his interpretation of probability and his use of intervals. There are two fundamental reasons why any such approach will be inadequate.

At the beginning of the essay, Kyburg makes a neat distinction between unhedged inferences producing hedged conclusions and hedged inferences producing unhedged conclusions. He associates the latter approach with an action-oriented perspective and labels it *pragmatic* because the conclusions correspond to beliefs that would rationalize relevant behaviors. But it is important to note that the so-called pragmatist's stance is not at all sufficient for action. Every day we take actions that are not rational with respect to *any* categorically accepted set of beliefs. For example, if we buckle up when getting in a car does that mean that we believe that are going to crash? Obviously not—if we really believed that we were to crash we would not get in the car at all. But if we accepted the proposition that we we

would not crash, then we would not wear the safety belts (assuming that they entail some mild bother or discomfort). Buckling up makes sense *only* in some intermediate circumstance where we believe it is possible but unlikely that we will crash, and that the value of added safety even in this unlikely case exceeds the certain cost of wearing the belts. Moreover, we can easily find other commonplace examples where the usual probability is in any given intermediate range.

So the first limitation of practical certainty is that it cannot possibly account for the large class of decisions that themselves hedge on contingencies. Who is buying all those lottery tickets, anyway? Sure, buying lottery tickets may be irrational. But one cannot make this judgment based solely on the low probability of winning; one must also consider the payoff.

The second limitation applies as well to situations where the action may be rationalized by categorically accepted beliefs. The problem is that it is simply unrealistic to define a meaningful probability threshold over a usefully broad set of decision problems. Kyburg acknowledges that acceptance thresholds may be situation-dependent, and proposes that we define "contexts" over which particular thresholds are applied. But he also suggests that two or three will be sufficient, a claim about which I am dubious. If decision problems arose in neat categories like "nuclear power plants" and "groceries", each with meaningful global criticality thresholds, then the scheme might be workable. But AI aspires to build agents capable of adapting to a range of decision situations, including those for which the criticality level is not assigned in advance. Homer-the-nuclear-plant-automaton might make a decision about which donut to eat (or whatever the robotic analog of mundane choice) on the same day he decides whether to release radiation into Springfield to avoid a possible meltdown. And even within a "nuclear emergency" context, the acceptance threshold should be variable depending on the alternatives available. Perhaps Homer should accept the proposition "meltdown on the way" at one probability when the option is to shut down the plant, and at a more stringent threshold when the option is to release radiation. There is no reason to think that a small number of levels of practical certainty will suffice for even a moderately broad range of decisions, and even less that we could identify the significant ones in advance. It is not that we will definitely require the potentially infinite precision of probability measures, just that we have no way of knowing at design time which part of the spectrum will turn out to be important.

Kyburg anticipates three Bayesian arguments against acceptance. The points I raise here fit best in the second category, those denying the feasibility of formally characterizing the notion of sufficiently high probability. But this is not quite analogous to denying that heavier than air machines can fly. Rather, we are merely criticizing one particular design as unpromising. Probability alone does not appear to offer enough lift to get an ample theory of acceptance off the ground.[1]

3 An Acceptable Characterization of Acceptance

A satisfactory theory of acceptance must admit the possibility of error and justify its risk by a comparable benefit in computational utility. An example of this kind of an account is the cost-benefit analysis performed by Ginsberg [3], which identifies several of the critical factors. Unfortunately, I cannot cite (or produce) an operational general acceptance theory at this point, and I suspect that the enterprise will prove quite difficult. But I can point out that current research on decision-theoretic control of inference (e.g., the work of Doyle, Russell and Wefald, Horvitz and Breese, Dean and Boddy, Greiner, Smith) is addressing related questions, and successful models of computational choice will also be applicable to the question of acceptance. The application will not be straightforward, as even computational utility is context-dependent—the value of saving time in decision making depends on the time stress (opportunity cost of inference) faced at execution time.

References

[1] Jon Doyle and Michael P. Wellman. Impediments to universal preference-based default theories. *Artificial Intelligence*, 49:97–128, 1991.

[1] This argument applies to any attempt to provide a purely probabilistic semantics for default reasoning (e.g., [4, 5]). There are numerous reasons one might jump to a default conclusion; likelihood of correctness is only one of them [1].

[2] Dan Geiger. Graphoids: A qualitative framework for probabilistic inference. Technical Report R-142, UCLA Computer Science Department, January 1990.

[3] Matthew L. Ginsberg. The computational value of nonmonotonic reasoning. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 262–268, 1991.

[4] Eric Neufeld. Defaults and probabilities; extensions and coherence. In *First International Conference on Principles of Knowledge Representation and Reasoning*, pages 312–323, 1989.

[5] Judea Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In *First International Conference on Principles of Knowledge Representation and Reasoning*, pages 505–516, 1989.