ON THE FOUNDATIONS OF BAYESIANISM

STEFAN ARNBORG
Royal Institute of Technology,
Nada, SE-100 44 Stockholm, Sweden. ‡

AND

GUNNAR SJODIN
Swedish Institute of Computer Science,
ARC, Box 1263, SE-164 29 Kista, Sweden. §

Abstract. We discuss precise assumptions entailing Bayesianism in the line of investigations started by Cox, and relate them to a recent critique by Halpern. We show that every finite model which cannot be rescaled to probability violates a natural and simple refinability principle. A new condition, separability, was found sufficient and necessary for rescalability of infinite models. We finally characterize the acceptable ways to handle uncertainty in infinite models based on Cox's assumptions. Certain closure properties must be assumed before all the axioms of ordered fields are satisfied. Once this is done, a proper plausibility model can be embedded in an ordered field containing the reals, namely either standard probability (field of reals) for a real valued plausibility model, or extended probability (field of reals and infinitesimals) for an ordered plausibility model. The end result is that if our assumptions are accepted, all reasonable uncertainty management schemes must be based on sets of extended probability distributions and Bayes conditioning.

1. Introduction

Several ways are possible for dealing with uncertainty and ignorance in AI and other applications. It has not been possible to find a unique correct way to handle it. This is because it is not a purely mathematical question but, since the time of Aristotle, a central problem in philosophy. *Bayesianism*, claiming that all types of uncertainty must be described by probabilities, is one possible way that has been tried in many application areas and with convincing results. One family of arguments consists of observations that even if other ways to deal with uncertainty

‡Email:stefan@nada.kth.se
§Email: sjodin@sics.se

are possible, they either have some easily stated deficiency or are equivalent to Bayesianism.

Indeed, such arguments have been put forward, but they have not been unanimously accepted. This also would follow from Bayesianism itself, since prior prejudices are predicted by the theory to outweigh every informal argumentation, and there is no proof method relating to real-world phenomena with the persuasiveness of pure logic and mathematics. This note was inspired by a recent critique [1] of [2] and [3].

We will here show assumptions, mainly refinability, that are strong enough to strictly imply Bayesianism and at the same time convincing in a subjective way (common sense). A proposed counterexample to Cox's argument is shown in section 3 and discussed sin section 4, where a theorem is shown saying that for a plausibility model with finite domain, natural refinements are possible if and only if the plausibility measure is rescalable to probability.

We discuss the extension to infinite but non-dense domains in section 5. We show with an example that some new assumption is required for the infinite case and give one, separability, that is both necessary and sufficient. We introduce the concept of extended probability models which have infinitesimal probabilities. We define the concept of a closed plausibility model and show that a model that can be closed in the real numbers is rescalable to a probability model. If we are content with a totally ordered domain of plausibilities, extended probability emerges as canonical uncertainty measure. If we weaken the assumptions and accept partially ordered plausibility values, we end up with sets of extended probability distributions.

2. Arguments for the Inevitability of the Bayesian View

In 1946, R.T. Cox published his findings [2] on some properties required by any good calculus of plausibility of statements. A very lucid elaboration of Cox's findings can be found in E.T. Jaynes posthumous manuscript [3, Ch. 2]. He stated three requirements:

- I: Divisibility and comparability- The plausibility of a statement is a real number and is dependent on information we have related to the statement.
- II: Common sense - Plausibilities should vary sensibly with the assessment of plausibilities in the model.
- III: Consistency - If the plausibility of a statement can be derived in two ways, the two results must be equal.

After introducing the notation $A|C$ for the plausibility of statement A given that we know C to be true, he finds the governing functional equation for defining the plausibility of a conjunction: $AB|C = F(A|BC, B|C)$ must hold for some function F. Since $ABC|D \equiv (AB)C|D \equiv A(BC)|D$, F must satisfy the equation of associativity: $F(F(x, y), z) = F(x, F(y, z))$, for $x = A|BCD$, $y = B|CD$ and $z = C|D$. At this point [2] and most other authors analyzing this problem assume that F is associative[4] on a dense domain where it is also differentiable or continuous. The result, under one of a couple of alternative assumption sets, is that no matter what our choice of $A|C$ is, there must be a function w such that $w(AB|C) = w(A|C)w(B|AC)$. The existence of a function that translates the plausibility measure to another measure satisfying the rules of probabilities will be called *rescalability*, and the main topic of investigation in this note is under what reasonable and precise assumptions rescalability obtains. From rescalability all the machinery of Bayesian analysis follows, except the way to assign prior probabilities.

It must be said that neither Cox nor Jaynes are completely rigorous in defining their assumptions, and recent critiques can be found in [5,1]. Halpern notes that all variations of the derivation of rescalability theorems make an assumption on the denseness of the domain of plausibility and some type of regularity assumption on the functions F and S. The critiques can be taken as evidence that Cox's common sense assumptions are not the only ones possible. We do not think anyone would argue against the desirability of the three general conditions I, II and III as given above. However, they must be interpreted to fine detail, and they may conflict with other desirable conditions. Particularly, common sense is a rather open-ended condition and it can certainly be debated what is required by common sense, and what is not.

The first condition I has been characterized as the 'dogma of precision' and is sometimes found unacceptable essentially by arguments saying that we cannot know which exact real numbers to use. Several alternatives based on intervals instead of numbers have been designed and motivated, by [6] and others. This may in turn lead to problems in choosing the exact real values used as end-points of the intervals. Although some interval based schemes can be seen as multiple-context Bayesian inference in analogy with multiple-criteria decision making, they are usually not presented as such. The insistence on consistency is only halfway to 'correctness', and one can argue that a consistent somewhat arbitrary method is no better than an inconsistent method that 'works in practise'. Floating point computation is a good example of such a method, but the analysis of which errors we make with floating point computation is also a major research field in numerical analysis. There are many possible objections, but here we will concentrate on how the density assumption in [2] and its followers can be relaxed. We will not give a full account of the background to the discussions and developments of Cox's ideas.

A more precise derivation of rescalability with significantly weaker assumptions was published by Aczél[7]. He relaxed the differentiability assumption of Cox and introduced the function G with the use: $A \vee B|C = G(A|C, B\overline{A}|C)$. It is then only necessary to assume continuity of G, and associativity and joint distributivity of F and G, to prove rescalability. The use of the auxiliary function G describing disjunctions instead of Cox's function S describing logical complement (negation) turns out to simplify the analysis.

3. Halpern's Example

The consistency assumption only says that associativity holds for values actually occurring as plausibilities of statements $A|BCD$, $B|CD$ and $C|D$. A small world is analyzed in [1] where there are no 4-tuples of statements to which the associativity condition could apply. In the notation of [2], the example consists of four groups of three statements each($\to$ stands for implication):

A, B, C, where $A \to B$ and $B \to C$ hold
D, E, G, where $D \to E$ and $E \to G$ hold
H, I, J, where $H \to I$ and $I \to J$ hold
K, L, M, where $K \to L$ and $L \to M$ hold

C, G, J and M exclude each other, so only one of them can hold. Plausibilities are assigned to these statements as follows:

$D|EG = H|IJ = 3/5$
$E|G = A|BC = 5/11$
$B|C = L|M = 11/19$
$A|C = I|J = 5/19$
$D|G = K|LM = 3/11$
$K|M = 3/19$
$H|J = 3/19 - \delta$, for some small $\delta > 0$.

From the above we find $5/19 = A|C = AB|C = F(A|BC, B|C) = F(5/11, 11/19)$ and $3/11 = D|G = DE|G = F(D|EG, E|G) = F(3/5, 5/11)$. Moreover, $3/19 = K|M = KL|M = F(K|LM, L|M) = F(3/11, 11/19)$, but $3/19 - \delta = H|J = HI|J = F(H|IJ, I|J) = F(3/5, 5/19)$. It is easy to see that these plausibilities are consistent in all ways, as shown by the detailed model in [1]. With the exception of $H|J$, all quantities could have been probabilities. The function F is not associative, because $F(3/11, 11/19) = F(F(3/5, 5/11), 11/19) = 3/19$, but $F(3/5, 5/19) = F(3/5, F(5/11, 11/19) = 3/19 - \delta$.

4. Common sense assumptions

Halpern's example is a finite model, and the function F is not associative. Its plausibility is thus not rescalable to probability. What happens in the example is that the same plausibility values are assigned to seemingly unrelated conditional statements. Therefore, a violation of associativity yields no immediate inconsistency. Whether or not this example really is a counterexample to a theorem of Cox is discussed in [4], but is not of our concern here.

A person interested in finite models would not find an assumption that models are infinite very compelling. But models are in practice crafted incrementally by refinement of simpler models, and it is completely plausible that such a person would insist that refinements should not be arbitrarily and unnecessarily restricted. Indeed, when and where to stop the refinement process cannot be known in advance. There is a standard method for developing probability models by splitting cases into subcases and assigning probabilities conditional on such subcases, to the degree required for an application. This method is an informative refinement method, sometimes known as 'extending the discussion'[8,9], and it is equally fundamental in any plausibility model. A weaker form of model refinement is a non-informative refinement where we do split cases into subcases, but do not assign new plausibilities of existing events contingent on the new subcases. Such refinements should never change the information obtainable from the model, nor should they render a consistent model inconsistent. We also want to have at our disposal the possibility of claiming that two statements are independent in a given context, so that knowledge of one does not change the plausibility of the other. This condition we call information independence. We argue that a well-informed method choice can be obtained by considering questions like these:

- **Refinability:** If we have already made a particular splitting of a statement into sub-cases, by adding new statements implying it, should it then always be possible to refine another statement in the same way, and with the same plausibilities in the new refinement? As an example, if we defined A' with $A' \to A$ and $A'|A = a$, should we for any existing statement B be allowed to define B' as a new symbol with $B' \to B$ and $B'|B = a$?
- **Information Independence:** If a statement is refined by several new symbols, should it then be possible to state that they are information independent, so that knowledge of one does not affect the plausibility of the other? As an example, if A and B are introduced as refinements of C, should we be permitted to claim that $A|BC = A|C$ and $B|AC = B|S$?
- **Strict Monotonicity:** Will it always be the case that the plausibility of a conjunction is less than those of the conjuncts, if these are independent and their plausibilities are not 0 or 1?

We mean that 'yes' answers to all are minimal precise conditions that entail Bayesianism for finite models. We also mean that they reflect common sense desiderata on a calculus for uncertainty better than the alternatively used associativity, density and regularity assumptions. For infinite models there turns out to be another possibility that Cox apparently did not realize, namely plausibilities taking values in an ordered field of reals and infinitesimals. This possibility was eventually (around 1965) noticed by Adams[10] and recently elaborated by Wilson[11].

4.1. ASSOCIATIVITY AND STRICT MONOTONICITY

If we accept refinability and information independence as reasonable assumptions, associativity and other algebraic laws for F and G follow: if a model has a violation of associativity for F, then there exists a simple and finite refinement (in three steps) that is arbitrarily blocked. If we have worked out a model where $F(a, F(b, c)) \neq F(F(a, b), c)$ for some plausibilities a, b and c, then we take an arbitrary statement S (not false) and refine with S_a, S_b and S_c: $S_a|S_b = a$, $S_b|S_c = b$ and $S_c|S = c$. Now the value $S_aS_bS_c|S$ can be computed in two ways giving different results, as $(S_aS_b)S_c|S = F(F(a, b), c)$ and as $S_a(S_bS_c)|S = F(a, F(b, c))$. In Halpern's example, it would be perfectly reasonable that one wants to add a new statement A' to the model, and such that $A' \to A$. Moreover, it would be reasonable to allow any value to be assessed for the plausibility $A'|A$, because there is no link to the rest of the model. In particular, we have already refined E by defining a sub-statement D with $D \to E$ and $D|E = 3/5$, so it should be safe to do the same thing with A and say $A'|A = 3/5$ for a new statement A' with $A' \to A$. Thus far we have not introduced any informative change in the model, and we would expect that nothing has happened. But we actually have got a violation of the associativity law for the statement $A'AB|C$, which can now be proved to have plausibility both 3/19 (the value of $(A'A)B|C$) and $3/19 - \delta$ (the value of $A'(AB)|C$). One can claim that this effect is a violation of common sense. It involves nothing that is infinite. Similar arguments can be used to show that F is bound by common sense to be symmetric, that G is associative and symmetric, and that they are jointly distributive:

Observation 1 *In order to satisfy natural requirements on consistency being preserved by non-informative refinements of models, we must work with models where F and G are partially specified in such a way that they satisfy the laws of associativity and symmetry, as well as joint distributivity.*

It is also reasonable to argue that F must be strictly monotone when none of its arguments represents falsity (*i.e.*, if x is not falsity and $u > v$, then $F(u, x) > F(v, x)$ and $F(x, u) > F(x, v)$). The requirement of strict monotonicity is stated in [3]: "If $A|C$ becomes more plausible, and $B|AC$ is not falsity, then $AB|C$ also becomes more plausible, if nothing else (namely $B|AC$) changes". This statement can certainly not be verified mathematically, but is something you have to believe to accept. It is assumed in most related analyses, including [7,2]. We summarize:

Observation 2 *It is reasonable to assume that the functions F and G of a plausibility model are strictly monotone for non-zero arguments, and that $F(x, y) < \min(x, y)$ and $G(x, y) > \max(x, y)$ for non-trivial plausibility values of x and y.*

4.2. THE FINITE CASE

It remains to consider whether any partially specified function can be extended to an associative function if it is associative on its range of definition. This is not generally the case, even if it also satisfies the other properties that will be required from the completed function: strict monotonicity and symmetry. If an appropriate rescaling to probabilities exists, we can find it by solving a finite linear system of equations and inequalities for the log probabilities $l_i = \log w(x_i)$ excluding the value for falsity. The system has an equation $l_i + l_j = l_k$ for each triple $x_k = F(x_i, x_j)$ and an inequality $l_i < l_j$ for every pair with $x_i < x_j$, and an equality $l_i = l_j$ when $x_i = x_j$.

We are now ready to state that rescalability of the F function is equivalent to finite refinability. The argument goes as follows: If rescalability obtains, it is trivial to extend F to an associative, symmetric and strictly monotone function over the dense interval $(0, 1)$ which covers any refinement. If rescalability does not hold, then this is equivalent to non-solvability of a linear program. But this means that a dual program has a solution and it so happens that this solution defines a refinement that is a proof of non-compliance of F with strict monotonicity. It is also possible to modify Aczéls analysis of Cauchy's and Euler's equations to prove simultaneous rescalability of F to $*$ and G to $+$. The following is proved in [12]:

Definition 3 *An* **extension base** *B of a sequence X of length L is a sequence (n_i) of length L of non-negative integers, multiplicities. A set of partial functions* *can be* **extended to extension base** B *if the partial functions can be extended to a domain such that every nested expression in the function symbols with arguments in* X *has a defined value if, for all* i*, the number of occurrences of* x_i *in the expression is not larger than the corresponding multiplicity* n_i *in* B*.*

Theorem 4 *Let* $X = (x_i)_{i=1}^L$ *be an increasing sequence of distinct values in the open interval* $(0, 1)$*, and* $S = \{1, \ldots, L\}$*. Given two sets of triples* $T_F, T_G \subset S^3$ *interpreted as specifications of two partial functions* F *and* G *satisfying also* $F(1, x_i) = x_i$ *,* $F(0, x_i) = 0$ *and* $G(0, x_i) = x_i$*.*

The following are equivalent:

(i) There is a finite extension base B *of* X *to which* F *and* G *cannot be jointly extended as symmetric, associative and strictly increasing functions satisfying joint distributivity.*

(ii) There is no increasing sequence of real numbers $(p_i)_{i=1}^L$ *such that if* $(i, j, k) \in T_F$*, then* $p_i * p_j = p_k$*, and if* $(i, j, k) \in T_G$*, then* $p_i + p_j = p_k$*.*

5. Infinite models

Infinite models, without regularity assumptions on F and G, are more complex. We first introduce the assumption of separability, under which any consistently refinable model must be rescalable to a probability model, and then we find a richer probability model family into which all models that can be closed are rescalable.

5.1. SEPARABILITY

Finite refinability is insufficient for infinite domains, as shown by the following consideration: in a probability model, if $x < y$ then the union of the intervals $[x^i, y^i]$ is a finite set of disjoint intervals, since the intervals will overlap for large i. But the number of intervals is invariant under strictly monotone rescaling. So a model where the union of such intervals (exponent now denoting iteration of F, so that $x^1 = x$ and $x^{n+1} = F(x, x^n)$) is an infinite set of disjoint intervals cannot be rescalable.

As an example with an infinite number of intervals thus not being rescalable, consider a domain generated from two statements with plausibilities $b = 1/4$ and $a = 1/5$. Let exponents of plausibilities denote iteration of the F function. The model is defined by: $F(b^j, a^k) = 1/(3(j+k)+(j+2k)/(j+k))$. Now $a^p = 1/(3*p+2)$, $b^p = 1/(3 * p + 1)$, and separation is not obtained, because no b^{p+1} is larger than a^p for any positive integer p, and therefore all intervals are disjoint. There appears to be no finite argumentation for the inadequacy of this model, at least not using reasonable refinability arguments.

Instead of generalizing Theorem 4 to infinite dimension, we solve a slightly easier problem: Suppose that a model is defined, and its F function is completed to a minimal function that already covers all refinements. Which are the properties required for rescalability of such a function? If the domain and range of F is D and R, respectively, and $R \subset D$, then we need only one new condition before we can prove rescalability, at least for the function F, and this is that the set of intervals defined above is finite! We call this property separability, for the following reason: if the condition obtains, then for any non-trivial plausibility c in the model, and for every non-trivial plausibilities x and y with $x < y$, there are integers p and q such that $y^p < c^q \leq x^p$, *i.e.*, some power of c separates some (equal) powers of x and y.

Definition 5 *Two non-trivial elements a, b of a plausibility model are called* **separable** *if $a < b$ and $a^p < b^{p+1}$ or $b < a$ and $b^p < a^{p+1}$ for some natural number p where the powers in the condition exist. A value a is separable from 0 or 1 if a and $F(a, a)$ are separable. Otherwise the elements are* **non-separable**. *A plausibility model is separable if all distinct plausibility values are separable.*

Non-separability is easily seen to be an equivalence relation: It is obviously reflexive and symmetric. It is also transitive: If a, b and c are plausibility values, $a < b < c$, and $b^{p+1} < a^p$ and $c^{p+1} < b^p$ for all $p > 0$, then $c^{p+2} < a^p$, *i.e.*, $F(c^{q+1}, c^{q+1}) < F(a^q, a^q)$ for $2q = p$ and by strict monotonicity of F we have $c^{q+1} < a^q$ for all integers $q > 0$. The following can be proved by modifying the analysis of the equation of associativity found in [7]:

Theorem 6 *Let the function $\circ : D^2 \to R$ have the following properties: $R \subset D$, $\{0, 1\} \subset D$ and $D \subset [0, 1]$; Associativity; Strict monotonicity on $D - \{0\}$; Symmetry; $0 \circ x = 0$ and $1 \circ x = x$; Model is separable.*

Then for $x, y \in (D - \{0\})^2$, $x \circ y = f(f^{-1}(x) + f^{-1}(y))$, for a partial strictly monotone function f whose inverse is a strictly monotone function f^{-1}.

Once we have accepted assumptions strong enough to ensure rescalability of F (or G), the arguments for joint rescalability of F to $*$ and G to $+$ are the same as before, based on the analysis Cauchy's or Euler's equation. If we insist that separability is not an appropriate common sense assumption, we cannot claim rescalability to the standard probability model. Previous work like [7] shows that continuity is an adequate assumption. We will now see what we can achieve without continuity or separability, in the next section. We introduce closure assumptions which makes the analysis amenable to standard algebraic methods.

5.2. EXTENDED PROBABILITY MODELS

We define an extended probability into which we can always rescale a well-constructed plausibility model:

Definition 7 *An* **extended probability model** *is a model based on probabilities taking values in an ordered field generated by the reals and an ordered set of infinitesimals. An* **infinitesimal** *is a non-zero element smaller in magnitude than any positive real.*

Extended probability was studied by Wilson[11] as a way to handle conditioning on rare events. It is closely related to Adams's proposal for the logic of conditionals[10]. The intuition behind extended probability models is that the probability value given by a real number is 'accompanied' by a set of probabilities that are ranked different but do not have measurably different values. The non-separable example in section 5.1 can be mapped into an extended probability model by mapping a to $1/2$ and b to $1/2+\epsilon$, where ϵ is a non-negative infinitesimal. We will show that such a mapping can always be found for a consistently refinable model.

Definition 8 *A plausibility model satisfying strict monotonicity, refinability and information independence assumptions can be* **closed** *if its functions F and G can be extended to a domain D, still satisfying refinability, information independence and strict monotonicity in the following way: The domain D of F contains its range. Likewise, on the domain D there is a function S with the property $G(x, S(x)) = 1$, and $G(x, y)$ is defined when $x \leq S(y)$. The range of G is contained in D. Closing a plausibility model results in a* **closed plausibility model**. *The domain D of the closed model must be ordered but need not be contained in the reals. Our arguments for associativity, symmetry and distributivity follow from refinability also for closed models.*

Theorem 9 *Every plausibility model that can be closed can be rescaled to an extended probability model.*

The proof of this theorem becomes embarrassingly complex. It is based on the fact that there is exactly one way to minimally complete the model to a field. An intuitive argument is that every computation in the extension can be mirrored in the plausibility model, by the rule of distributivity. As an example, the expression $a+b$ in the field, where a and b are plausibility values, corresponds to the expression $G(F(e,a), F(e,b))$ in the plausibility model for some e. This expression must be defined for a sufficiently small non-zero e, for example $\min(f, S(f))$ for any non-trivial plausibility value f. Technically, one has to extend the domain by a sequence of constructions similar to the quotient construction taking a commutative ring without zero divisors to a field[13]. In each step one has to verify in detail that an embedding has been produced, and that all relevant algebraic laws (symmetry, associativity, *etc.*) and strict monotonicity have been preserved. Somewhat surprisingly, the resulting ordered field does not necessarily have a unique minimal extension to a field containing all real numbers. So we need a somewhat complex argumentation to show that every ordered field contains an extended probability model:

Conway derives the structure of transfinite numbers using a real ordered field **No** that he shows[14, Th. 28, 29] universal, *i.e.,* every other ordered field is (isomorphic to) a subfield of **No**. This field contains all real numbers and is an extended probability model: Assume **No** contains some non-real element e between 0 and 1. This element is associated with a real number r_e, the least upper bound on reals smaller than e. The solution to $x \oplus r_e = e$ is an infinitesimal, a non-zero element smaller in magnitude than any positive real. Thus, since $e = r_e \oplus x$, every element of the model is generated by its infinitesimals and reals. Thus Theorem 9 follows, since we already explained why every closed plausibility model can be embedded in an ordered field.

If the closed model can be embedded in the real numbers, then it must be a subfield of the field of real numbers, since only these subfields of **No** have least upper and greatest lower bounds on all bounded sets:

Corollary 10 *Every plausibility model which can be closed in the domain of the reals, can be rescaled to a standard probability model.*

Indeed, the closed model has a function F satisfying the premises of Theorem 6, except possibly the separability condition. We know by Theorem 9 that our model can be rescaled into an extended probability model. If F is not separable the model cannot be embedded in the field of reals, otherwise it can, because the embedding process described in the proof of Theorem 9 does not introduce infinitesimals.

Finally, if we accept an ordered domain instead of a real valued domain in Jayne's desideratum I, we arrive rather painlessly at extended probability as canonical uncertainty measure, with the added insight that extended probability is required only in infinite models (although it can be motivated pragmatically also for finite models, as is done in default and other non-monotonic reasoning frameworks).

6. Conclusions

We proposed to weaken the common sense assumptions used previously from domain denseness and continuity of auxiliary functions to refinability and allowing information independence, and showed such assumptions sufficient for finite models. That our proposal uses truly weaker assumptions is shown by its inadequacy for the infinite case, where we proposed an assumption weaker than denseness, namely separability, which entails rescalability to standard probability. Without separability we can only show rescalability to extended probability. Several contemporary reasoning schemes are related (shown more or less equivalent) to infinitesimal or extended probability in [15], so our result seems to re-concile Bayesianism and non-monotonic reasoning. We finally observe that our techniques seem to apply also to weakening the assumptions of the justifications of Bayesianism published by Savage[16] and Lindley[17].

The end result of this analysis is that, under our assumptions of refinability, information independence, strict monotonicity and closability, every well founded uncertainty management methodology must be equivalent to a system where uncertainty is completely described by a set of extended probability distributions, a system we might call Extended Robust Bayes, and where conditioning is made by Bayes's rule. This approach has indeed been proposed, by Wilson[18]. Our assumptions are much weaker than thoses based on coherence or consistent betting behaviour, since there we end up with a single standard probability distribution and we can also apply the result to infinite-dimensional and non-parametric inference situations which we cannot reach with Cox's approach. This might be a strength of our analysis, since it is known that standard Bayes analysis, being a method satisfying the strict likelihood principle, gives unwanted results in certain such problems[19]. Needless to say, those authors advocating standard Bayesianism have not been strengthened or weakened by our analysis, since their approach can still be defended by pragmatic considerations. Extended probability, for example, was introduced for pragmatic reasons where standard probability gives unwanted effects in applications where one wants to condition on rare events (like in systems reliability studies), but if we take probability as a betting rate we can hardly expect to measure differences between probabilities an infinitesimal apart. The Robust approach was introduced where one does not want to weight together different subjective probability assessments[20], but on the other hand in most cases where interval-based plausibilities are proposed the documented evidence only calls for probability volatility estimates, and these can be used, *e.g.*, to decide whether or not to wait for more information when a decision is called for[21,22].

References

1. J. Halpern, "A counterexample to theorems of Cox and Fine," *Journal of AI research*, **10**, pp. 67–85, 1999.
2. R. Cox, "Probability, frequency, and reasonable expectation," *Am. Jour. Phys.*, **14**, pp. 1–13, 1946.
3. E. T. Jaynes, *Probability Theory: The Logic of Science*, Preprint: Washington University, 1996.
4. P. Snow, "On the correctness and reasonableness of Cox's theorem for finite domains," *Computational Intelligence*, **14**, pp. 452–459, 1998.
5. J. Paris, *The Uncertain Reasoner's Companion*, Cambridge University Press, 1994.
6. P. Walley, "Measures of uncertainty in expert systems," *Artificial Intelligence*, **83**, pp. 1–58, 1996.
7. J. Aczél, *Lectures on Functional Equations and their Applications*, Academic Press, 1966.
8. M. Tribus, "Bayes's equation and rational inference," in *Rational Descriptions, Decisions and Designs*, pp. 73–106, Pergamon Press, 1969.
9. D. Heckerman and H. Jimison, "A Bayesian perspective on confidence," in *Uncertainty in Artificial Intelligence*, L. N. Kanal, T. Levitt, and J. Lemmer, eds., pp. 149–160, Elsevier Science Publishers, 1989.
10. E. Adams, "Probability and the logic of conditionals," in *Aspects of Inductive Logic*, J. Hintikka and P. Suppes, eds., pp. 265–316, North Holland, Amsterdam, 1966.
11. N. Wilson, "Extended probability," in *Proceedings of the 12th European Conference on Artificial Intelligence*, pp. 667–671, John Wiley and Sons, 1996.
12. S. Arnborg and G. Sjödin, "Bayes rules in finite models," in *European Conference on Artificial Intelligence*, (Berlin), 2000.
13. S. MacLane and G. Birkhoff, *Algebra*, The MacMillan Company, New York, 1967.
14. J. Conway, *On Numbers and Games*, Academic Press, 1976.
15. S. Benferhat, D. Dubois, and H. Prade, "Nonmonotonic reasoning, conditional objects and possibility theory," *Artificial Intelligence*, **92**, pp. 259–276, 1997.
16. L. Savage, *Foundations of Statistics*, John Wiley & Sons, New York, 1954.
17. D. V. Lindley, "Scoring rules and the inevitability of probability (with discussion)," *Internat. Stat. Rev*, **50**, pp. 1–26, 1982.
18. N. Wilson, "A logic of extended probability," in *Proceedings of the First International Symposium on Imprecise Probabilities and their Applications*, G. de Cooman, F. G. Cozman, S. Moral, and P. Walley, eds., pp. 397–404, Gent University, 1999.
19. J. Robins and Y. Ritov, "Towards a curse of dimensionality appropriate(CODA) asymptotic theory for semi-parametric models," *Statistics in Medicine*, **16**, pp. 285–319, 1997.
20. J. O. Berger, "An overview of robust Bayesian analysis (with discussion)," *Test*, **3**, pp. 5–124, 1994.
21. L. Seligman, P. Lehner, K. Smith, C. Elsaesser, and D. Mattox, "Decision-centric information monitoring," *Journal of Intelligent Information System*, **14**, p. to appear, 2000.
22. S. Arnborg, H. Artman, J. Brynielsson, and K. Wallenius, "Information awareness in command and control: Precision, quality, utility," in *FUSION 2000*, to appear, 2000.