A Note on the Foundations of Bayesianism
Research Note

Stefan Arnborg

Royal Institute of Technology, Nada, SE-100 44 Stockholm, Sweden.
Email:stefan@nada.kth.se

Gunnar Sjödin

Swedish Institute of Computer Science, ARC, Box 1263, SE-164 29 Kista, Sweden. Email: sjodin@sics.se

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. We 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 can be 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.

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.

We will here show assumptions that are strong enough to strictly imply Bayesianism and at the same time convincing in a subjective way (common sense). The background is given in section 2, and our assumptions and their algebraic consequences are discussed in section 3. In section 4, an example of non-rescalability is discussed and 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.

An example shows that some new assumption is required for the infinite case and we 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 is inevitable under our assumptions, and with a partially ordered domain of plausibilities we get sets of extended probability distributions.

2 Arguments for the Inevitability of the Bayesian View

In 1946, R.T. Cox published his findings [6] 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 [9, Ch. 2]. He stated three requirements:

I: Divisibility and comparability- The plausibility of a statement is a real number between 0 (for false) and 1 (for true) 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 [6] and most other authors analyzing this problem assume that F is associative[13] 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.

A more precise derivation of rescalability with significantly weaker assumptions was published by Aczél[1]. 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, associativity and that F distributes over G, to prove rescalability. The use of the auxiliary function G describing disjunction instead of Cox's function S describing logical complement (negation) turns out to simplify the analysis.

3 Common sense assumptions

Halpern[7] gave a non-rescalable example. The example is a finite model, and the function F is not associative. Its plausibility is thus not rescalable to probability. Whether or not this example really is a counterexample to a theorem of Cox is discussed in [13], but is not the topic of this note.

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 a refinability assumption seems more acceptable: 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'[14,8], and it is equally fundamental in a 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. We thus 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 $B|AC = B|C$?

- **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. 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 noticed by Adams[2] and recently elaborated by Wilson[15].

3.1 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 such that $S_a|S_b = a$, $S_b|S_c = b$ 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 $S_a(S_bS_c)|S = F(a, F(b, c))$. Similar arguments can be used to show that F is bound by common sense to be symmetric, that G is associative and symmetric for arguments where it is defined, and that F distributes over G in the sense that if $G(x, y)$ is defined, then $F(z, G(y, x)) = G(F(z, x), F(z, y))$. In other words, refinability and information independence ensures that these laws are inherited by F and G from the corresponding laws of propositional logic. 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 $F(x, u) > F(x, v)$). It is assumed in most related analyses, including [1,6]. We summarize:

Claim 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 do not violate the laws of associativity and symmetry, and moreover that F will distribute over G. 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 The finite case

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 $x_i = x_j$.

If a partially specified function can be completed to a full function over the support points (and some more points) satisfying associativity, symmetry and strict monotonicity, then the system is solvable. A simple case where the partially specified function triples satisfy the laws, but no completion over the support points does so, is the following:

$$F(x_4, x_4) = a \tag{1}$$
$$F(x_3, x_5) = a \tag{2}$$
$$F(x_2, x_4) = b \tag{3}$$
$$F(x_1, x_5) = b \tag{4}$$
$$F(x_4, x_6) = c \tag{5}$$
$$F(x_3, x_7) = c \tag{6}$$
$$F(x_2, x_6) = d \tag{7}$$
$$F(x_1, x_8) = d \tag{8}$$

Here we have assumed that the x_i quantities are ordered increasingly in the open interval $(0, 1)$, but the quantities a, b, c and d can have any values. If the plausibilities were scalable to log probabilities l_i, there should be a solution to the system:

$$l_4 + l_4 = l_a \tag{9}$$
$$l_3 + l_5 = l_a \tag{10}$$
$$l_2 + l_4 = l_b \tag{11}$$
$$l_1 + l_5 = l_b \tag{12}$$
$$l_4 + l_6 = l_c \tag{13}$$
$$l_3 + l_7 = l_c \tag{14}$$
$$l_2 + l_6 = l_d \tag{15}$$
$$l_1 + l_8 = l_d, \tag{16}$$

together with the conditions $l_i < l_{i+1}$.

If we now add together the equations (9-16) multiplied with the coefficient sequence $(1, -1, -1, 1, -1, 1, 1, -1)$, we find after cancelling that they imply $l_7 = l_8$, contrary to the condition $l_7 < l_8$.

But if it were possible to complete the partially specified F so that it satisfies symmetry and associativity, we can reach the same conclusion by composing, with the function F, equations (1-8), after first swapping the equations with negative coefficient. The resulting equation is $F(F(x_4, x_4), F(a, F(F(x_1, \ldots))))$ $= F(F(x_3, x_5), F(a, F(F(x_2, \ldots))))$, and thus by symmetry and associativity we can rearrange it to $F(x_7, F(a, F(b, \ldots))) = F(x_8, F(a, F(b, \ldots)))$, where the omitted (dotted) parts of the left and right sides are equal. This entails, because of strict monotonicity and because no variable is zero, that $x_7 = x_8$, contrary to the assumption that $x_7 < x_8$. This also means that it is possible to add a finite set of statements by refinement with plausibilities that leads to inconsistency in the plausibility assignment. In this example we can add statements $\{A_i\}_{i=1}^7$, B_4 and C, with $A_i|C = x_i$ and $B_4|C = x_4$. If the A_i and B_4 are independent given C, the statement $A_1A_2A_3A_4B_4A_5A_6A_7|C$ can be shown to have two different plausibilities, $F(q, x_7)$ and $F(q, x_8)$ for $q = F(a, F(b, F(c, F(d, x_1, F(x_2, F(x_3, F(x_4, F(x_4, F(x_5, x_6) \cdots)$.

We can now 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[1] of Cauchy's equations to prove simultaneous rescalability of F to $*$ and G to $+$:

Definition 2 *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** *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 3 *Let $X = (x_i)_{i=1}^L$ be an increasing sequence 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 $G(0, x_i) = x_i$. Then 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 increasing functions satisfying joint distributivity.

(ii) There is no increasing sequence of real numbers in $(0, 1)$ $(p_i)_{i=1}^L$ 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.

Suppose instead 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 4 *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 is proved in the appendix:

Theorem 5 *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}.

5.2 Extended probability models

Definition 6 *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.*

For algebraic terminology used from now, see [11]. Extended probability was studied by Wilson[15] as a way to handle conditioning on rare events. It is closely related to Adams's proposal for the logic of conditionals[2]. 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 and closed model.

Definition 7 *A plausibility model satisfying strict monotonicity, refinability and information independence assumptions can be* **closed** *if its functions* F *and* G *can be extended to an ordered 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**.

Theorem 8 *Every plausibility model that can be closed can be rescaled to an extended probability model.*

Conway derives the structure of transfinite numbers using a real ordered field **No** that he shows[5, 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, every element of the model is generated by its infinitesimals and reals. Thus it suffices to show that every closed plausibility model can be embedded in an ordered field. An proof sketch is in the appendix.

Corollary 9 *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 5, except possibly the separability condition. We know by Theorem 8 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: the embedding process described in the proof of Theorem 8 does not introduce infinitesimals into a separable and closed model.

Finally, if we accept an ordered instead of a real 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). Moreover, if we allow a set of plausibility values that are only partially ordered, our assumptions would lead to uncertainty management schemes where uncertainty is modelled by a set (via the set of total orderings compatible with the partial order) of extended probability distributions.

6 Conclusions

We proposed to weaken the common sense assumptions used previously to prove rescalability, from domain denseness and continuity of auxiliary functions to refinability and allowing information independence. We showed such assumptions sufficient for finite models. For the infinite case, 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 [4]. Thus, for any scheme that cannot be described as based on sets of extended probability distributions, it would be interesting to see what is lost by violating our assumptions, and what is gained in terms of alternative good properties. As an example, there is a current example in epidemiology, where there seems to be an incompatibility in a non-parametric inference problem between coherence and frequentist coverage, even asymptotically[12].

References

[1] J. Aczél. *Lectures on Functional Equations and their Applications.* Academic Press, 1966.

[2] E. Adams. Probability and the logic of conditionals. In J. Hintikka and P. Suppes, editors, *Aspects of Inductive Logic*, pages 265–316. North Holland, Amsterdam, 1966.

[3] S. Arnborg and G. Sjödin. Bayes rules in finite models. In *European Conference on Artificial Intelligence*, Berlin, 2000.

[4] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92:259–276, 1997.

[5] J.H. Conway. *On Numbers and Games.* Academic Press, 1976.

[6] R.T. Cox. Probability, frequency, and reasonable expectation. *Am. Jour. Phys.*, 14:1–13, 1946.

[7] J. Halpern. A counterexample to theorems of Cox and Fine. *Journal of AI research*, 10:67–85, 1999.

[8] D. Heckerman and H. Jimison. A Bayesian perspective on confidence. In L. N. Kanal, T.S. Levitt, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 149–160. Elsevier Science Publishers, 1989.

[9] E. T. Jaynes. *Probability Theory: The Logic of Science.* Preprint: Washington University, 1996.

[10] H. W. Kuhn. Dual systems of homogeneous linear relations. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 3–18. Princeton University Press, 1956.

[11] S. MacLane and G. Birkhoff. *Algebra.* The MacMillan Company, New York, 1967.

[12] J. Robins and Y. Ritov. Towards a curse of dimensionality appropriate(CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine*, 16:285–319, 1997.

[13] P. Snow. On the correctness and reasonableness of Cox's theorem for finite domains. *Computational Intelligence*, 14:452–459, 1998.

[14] M. Tribus. Bayes's equation and rational inference. In *Rational Descriptions, Decisions and Designs*, pages 73–106. Pergamon Press, 1969.

[15] N. Wilson. Extended probability. In *Proceedings of the 12th European Conference on Artificial Intelligence*, pages 667–671. John Wiley and Sons, 1996.

A Proofs of Theorems

Lemma 10 (Kuhn[10]) *The system of equations $Ax = 0$ has a positive solution $x > 0$ if there is no u such that $A^T u \geq 0$ and $u \neq 0$.*

Lemma 11 *Let F be a linear subspace of R^n. The following conditions are equivalent:*

(i) There is no element in F with all components positive.
(ii) There is a non-negative non-zero vector d orthogonal to F.

PROOF.

(ii)→(i): This direction is obvious, since a vector orthogonal to a non-zero and non-negative one cannot have all components positive.

(i)→(ii): Assume (i): There is no element in F of R^n with all components positive. Let F be the space spanned by the rows of matrix B, $F = \{B^T y : y \in R^k\}$. Let the rows of A be a base for the orthogonal co-space of F, $AB^T = 0$. Thus, $F = \{x : Ax = 0\}$ and $Ax = 0$ has no positive solution x by our assumption that (i) is the case. Since $Ax = 0$ has no positive solution, by Lemma 10 there is a u such that $A^T u \geq 0$ and $u \neq 0$. Now $u^T A$ is a non-negative vector, and it is orthogonal to every vector in F because $AB^T = 0$ and thus $(u^T A)(B^T x) = 0$ for all $x \in R^k$. So (ii) applies, *i.e.,* (i)→(ii).

Lemma 12 *Let $X = (x_i)_{i=1}^L$ be an increasing sequence of real values. Let $S = \{1, \ldots, L\}$ and $T_\circ \subset S^3$ be a finite set of triples. The partial function $\circ$ satisfies $x_i \circ x_j = x_k$, for all $(i, j, k) \in T_\circ$. Then the following conditions (i) and (ii) are equivalent:*

(i) There is a finite extension base B of X to which $\circ$ cannot be extended as a symmetric, associative and increasing function.
(ii) There is no increasing sequence of numbers $(f_i)_{i=1}^L$ such that if $(i, j, k) \in T_\circ$, then $f_i + f_j = f_k$.

PROOF.

(i)→(ii) : If (ii) is not the case, there exist appropriate f_i. Define $l(x)$ by interpolation to an increasing function between the constraints $l(x_i) = f_i$. The function $x \circ y = l^{-1}(l(x) + l(y))$ is associative, symmetric and increasing. So also (i) is not the case, which shows (i)→(ii).

(ii)→(i): Assume (ii) is the case. Define the $|T_\circ|$ by L matrix M to have one row for each tuple in $T_\circ$. For such a tuple (i, j, k), the row has the value 1 in columns i and j, the value -1 in column k, and zero otherwise. Matrix D is $L-1$ by L and has value $D = I' - I''$ where I' and I'' is the L by L unit matrix with the first and last row, respectively, deleted. From now on we regard f as a sequence of variables f_i. Since (ii) is the case, there is no L-vector solution f to $Mf = 0$ that also satisfies $Df > 0$, since such a solution would contradict non-existence of the f_i.

The solution space F of $Mf = 0$ is such that the linear subspace DF is orthogonal to some non-zero vector d with non-negative components, by Lemma 11. In other words, a linear equation $d^T Df = 0$ for f can be derived from $Mf = 0$ only, *i.e.*, the null space $\{f : Mf = 0\}$ of M is included in the null space $\{f : d^T Df = 0\}$ of $d^T D$, and $d^T D = c^T M$ for some vector c. Since M and D have integer elements, and the condition is homogeneous in d, we can assume that d consists of natural numbers and c of integers. Thus, a linear equality $d^T Df = 0$ for f can be obtained as a linear combination with integer coefficients of the linear equalities given by the rows of the system $Mf = 0$. But each row r of M is derived from a constraint $x_k = x_i \circ x_j$ for the function $\circ$. By composing these constraints with the associative and commutative operator $\circ$ in the pattern indicated by c we can derive a functional constraint on $x \circ y$, and at last obtain a functional constraint corresponding to the linear constraint $d^T Df = 0$. We compose the constraints coded by a triple of $T_\circ$ a number of times given by the magnitude of the corresponding coefficient c_i of the linear combination, reversing the equation if the coefficient is negative. In this way we derive a functional constraint:

$$a_1 \circ a_2 \circ \cdots \circ a_m = b_1 \circ b_2 \circ \cdots \circ b_n. \tag{A.1}$$

The corresponding linear constraint $d^T Df = 0$ can be written as

$$d_1 f_1 + d_2 f_2 + \cdots d_{L-1} f_{L-1} = d_1 f_2 + d_2 f_3 + \cdots d_{L-1} f_L, \tag{A.2}$$

where no d_i is negative and at least one is positive. But (A.2) results from the linear form of (A.1) by cancelling certain elements in both sides. Thus, $n = m$ and either $a_i = b_i$ (for quantities cancelling in the linear combination) or $a_i < b_i$ (for quantities remaining in (A.2), with at least one strict inequality since at least one d_i is non-zero.

But then, from strict monotonicity, we must also have: $a_1 \circ a_2 \circ \cdots \circ a_m < b_1 \circ b_2 \circ \cdots \circ b_n$.

There can thus not be an increasing extension of $\circ$ to an extension base defined by the union of the (a_i) and (b_i) sequences, in other words (i) is the case.

Lemma 13 (Aczél[1]) *The solutions to Cauchy's equation $f(x+y) = f(x) + f(y)$, when constrained to be bounded and monotone, are $f(x) = kx$.*

Proof of Theorem 3 Lemma 12 applies both to the function F and the function G of a consistently refinable plausibility model, since these functions are both, by Claim 1, associative, symmetric and increasing. For the function F, the numbers f_i must be negative, since we assumed $F(x,y) < \min(x,y)$ and there is an equation $f_i + f_L = f_j$ as well as inequalities $f_j < f_i < f_L$ in our system. The f_i can thus be taken as log probabilities. For the function G, the f_i must for analog reasons be positive. They can be taken as probabilities after some normalizing linear scaling. Assume thus that the plausibility measure has been scaled taking G to $+$. The distributivity equation (Claim 1) is transformed to a family of Cauchy equations, $F(x+y,z) = F(x,z) + F(y,z)$. By Lemma 13[1] and Claim 1, the solution has the form $F(x,z) = xc(z)$ for some monotone function c. But since $F(1,z) = z$ we must have $c(z) = z$ on the domain, *i.e.*, $F(x,y) = xy$.

Proof of Theorem 5. Define the function family $\phi_m(x)$ inductively by $\phi_0(x) = 1$, and $\phi_{m+1}(x) = x \circ \phi_m(x)$, for $m \geq 0$. The family is well-defined since the range of $\circ$ is included in its domain. It is not necessarily continuous. Now, $\phi_{m_1}(x) \circ \phi_{m_2}(x) = \phi_{m_1+m_2}(x)$ and $\phi_{m_1}(\phi_{m_2}(x)) = \phi_{m_1 m_2}(x)$, and $\phi_m(x)$ is a strictly monotone function of x when $m \neq 0$, going from 0 to 1 as x does so. The symmetry assumption also leads to the law $\phi_m(x \circ x') = \phi_m(x) \circ \phi_m(x')$. For a rational number $r = p/q$ and $x, y \in D$, we define $y = \phi_r(x)$ if $\phi_q(y) = \phi_p(x)$. Now, ϕ_r is a partial strictly monotone function. Select some $c \in D - \{0,1\}$ and let $f(r) = \phi_r(c)$. As a function of r, $f(r) = \phi_r(c)$ is a strictly monotone decreasing partial function. We extend f by defining its inverse f^{-1} to a total function from D into the reals by a direct limit process of interval bisection: For $x \in D - \{0,1\}$, define $f^{-1}(x)$ as the limit of p_i/q_i, when $q_i = 2^i$ and $\phi_{p_i+1}(c) < \phi_{q_i}(x) \leq \phi_{p_i}(c)$. There is always a unique solution sequence p_i, because $\phi_0(c) = 1$, so the right inequality can be satisfied, and the left inequality can be satisfied because the separability condition makes $\lim_{n\to\infty} \phi_n(x)$ independent of x if $0 < x < 1$. The sequence p_i/q_i will converge to a real number, since the sequence p_j/q_j for $j > i$ is contained in the interval $[p_i/q_i, (p_i+1)/q_i]$, whose length goes to 0. Thus, $f^{-1} : (D - \{0\}) \to [0, \infty)$ is a strictly monotone total function on D (strictly because of separability). Its range is closed under addition, as can be seen by comparing two bisection sequences $\{p_i\}$ and $\{p'_i\}$ converging to $f^{-1}(x) = r$ and $f^{-1}(y) = r'$ with the one, $\{p''_i\}$, for $d = x \circ y$ converging to $f^{-1}(x \circ y) = r''$.

The relations $\phi_{p_i+1}(c) < \phi_{q_i}(x) \leq \phi_{p_i}(c)$ and $\phi_{p'_i+1}(c) < \phi_{q_i}(y) \leq \phi_{p'_i}(c)$ imply, because of symmetry of $\circ$, $\phi_{q_i}(d) \leq \phi_{p_i+p'_i}(c)$ and $\phi_{q_i}(d) > \phi_{p_i+p'_i+2}(c)$. This implies that $p''_i = p_i + p'_i$ or $p''_i = p_i + p'_i + 1$ for all i and thus $r'' = r + r'$. But then $f^{-1}(d) = f^{-1}(x \circ y) = f^{-1}(x) + f^{-1}(y)$, or $x \circ y = f(f^{-1}(x) + f^{-1}(y))$ for all $x, y \in (D - \{0\})$, which finishes the proof.

[1] This lemma assumes the function to be defined on a dense interval, whereas we have a partial function assumed to be extensible. This technicality is solved in [3].

Proof of Theorem 8

We show how a closed plausibility model is embedded in an ordered field. We have the closed model defined by the ordered domain D with smallest and largest elements 0 and 1. The functions F, G and S, where F and G satisfy the rules of $\cdot$ and $+$ for a commutative ring on their domains of definition, 0 and 1 are unit elements of G and F, respectively. Let $D^+ = (D - \{0\})$.

The function $S : D \to D$ maps its argument x to the solution y of the equation $G(x, y) = 1$. The function F is defined on D^2, and G on $\{(x, y) : (x, y) \in D^2 \wedge x \leq S(y)\}$. We must extend D while defining the rules for $\circ$ and $\oplus$ as extensions of F and G. We do this in three steps, using the standard technique of defining an extension as the quotient of a set of pairs by an equivalence relation, and indicating which element of the extension that corresponds to each element of the original domain. We use the notation $[a]_\sim$ for the equivalence class of $\sim$ containing a. It is easy, in each step, to define the functions $\circ$ and $\oplus$ on the extensions and verify that they are indeed functions and extensions, that their laws are preserved, as well as to verify that no two elements of the old domain become equivalent in the new domain. Details of this verification are omitted. First, note that for every sequence $(a_i)_1^n$ there is a non-zero c_n depending only on n such that $G(F(c_n, a_1), G(F(c_n, a_2), \ldots F(c_n, a_n) \ldots)$ is defined (for any non-trivial plausibility value e, choose $c = \min(e, S(e))$ and $c_n = c^{\lceil \log n \rceil}$).

The first embedding step introduces non-negative rationals and is similar to the standard quotient construction for integral domains. Let $D^{(1)} = (D \times D^+/\sim$, where, for $a, c \in D$ and $b, d \in D^+$, $(a, b) \sim (c, d)$ iff $F(a, d) = F(b, c)$. Use notation $[a, b]$ for $[(a, b)]_\sim$. An element $d \in D$ is identified with $[d, 1] \in D^{(1)}$. Define $<$, $\circ$ and $\oplus$ as total functions by $[a, b] < [c, d]$ iff $F(a, d) < F(c, b)$, $[a, b] \circ [c, d] = [F(a, c), F(b, d)]$, and $[a, b] \oplus [c, d]) = [G(F(a, d, e), F(b, c, e)), F(b, d, e)]$ for suitably small e. The rational number $2/3$ is identified with $[G(x, x), G(x, G(x, x))]$ for some $0 < x < c_3$, and the other non-negative rationals are similarly defined. In this extension the rules for a field are satisfied, except that we have not yet an additive inverse (or negative values).

Our second embedding step introduces subtraction and negative values: Let $D^{(2)} = (D^{(1)} \times (D^{(1)})/\approx$, where, for $a, b, c, d \in D^{(1)}$, $(a, b) \approx (c, d)$ iff $a \oplus d = b \oplus c$. Use notation $[[a, b]]$ for $[(a, b)]_\approx$. An element $d \in D^{(1)}$ is identified with $[[d, 0]] \in D^{(2)}$. Define $<$, $\cdot$ and $+$ in this extension by $[[a, b]] < [[c, d]]$ iff $a \oplus d < c \oplus b$, $[[a, b]] \cdot [[c, d]] = [[a \circ c \oplus b \circ d, a \circ d \oplus b \circ c]]$, and $[[a, b]] + [[c, d]]) = [[a \oplus c, b \oplus d]]$. The structure $(D^{(2)}, <, \cdot, +, 1, 0)$ is now an ordered ring and indeed an ordered integral domain (because of strict monotonicity). Like all ordered integral domains it can be embedded in an ordered field[11, Ch V.2, Theorem 6]. This field embeds the closed plausibility model. This finishes the embeddability proof.