Constructing a Logic of Plausible Inference:
A Guide to Cox's Theorem

Kevin S. Van Horn

Dept. of Computer Science and Operations Research
North Dakota State University
Fargo, ND 58105

Abstract

Cox's Theorem provides a theoretical basis for using probability theory as a general logic of plausible inference. The theorem states that any system for plausible reasoning that satisfies certain qualitative requirements intended to ensure consistency with classical deductive logic and correspondence with commonsense reasoning is isomorphic to probability theory. However, the requirements used to obtain this result have been the subject of much debate. We review Cox's Theorem, discussing its requirements, the intuition and reasoning behind these, and the most important objections, and finish with an abbreviated proof of the theorem.

Key words: Cox, Bayesian, probability

1 Introduction

In 1946, R. T. Cox wrote a paper [1] discussing systems for plausible reasoning, that is, reasoning about degrees of plausibility, belief, confidence, or credibility. He proposed a handful of intuitively-appealing, qualitative requirements on systems of plausible reasoning, and showed that only those systems isomorphic to probability theory satisfy the requirements. Over the years Cox's arguments have been refined by others [2–5], making explicit some requirements that were only implicit in Cox's original presentation, and replacing some of the requirements with slightly less demanding (and hence less disputable) assumptions than those used in Cox's original proof. We apply the name "Cox's Theorem" to all of these variants.

_____

Email address: `Kevin.VanHorn@ndsu.nodak.edu` (Kevin S. Van Horn).

Orthodox statistics views probabilities only as long-run frequencies of repeatable events [6], in contrast to Bayesian theory, in which probabilities may describe degrees of belief [7] or states of partial knowledge [8] (and hence may be applied even to non-repeatable events). From the orthodox, frequentist standpoint, Bayesians may seem to be misapplying probability theory—after all, why should one expect rules for manipulating relative frequencies to be at all appropriate for manipulating degrees of plausibility? Cox's Theorem answers this objection: the rules of probability theory need not be derived from a definition of probabilities as relative frequencies, but also follow from certain properties one might desire of any system of plausible reasoning.

This paper may be thought of as a tutorial guide to Cox's Theorem. In order to understand the significance, applicability, and limitations of Cox's Theorem, one must know exactly what requirements are used to obtain the result; thus, we follow Paris's lead [4] in carefully and explicitly laying out the requirements we use. We go further and devote a major portion of this paper to discussing why one might find the requirements to be reasonable or desirable, and discussing the most significant objections to these requirements. We then give an abbreviated proof of Cox's Theorem from our requirements. We omit the proofs of solutions to functional equations, instead merely referencing where these may be found in the literature; the emphasis is on how our requirements lead to these equations.

In contrast to previous treatments of Cox's Theorem, but following common working practice among Bayesians, we condition the plausibility of a proposition on a *state of information*, rather than on another proposition. This is important in motivating the universality requirement of Section 7.

2   Degrees of plausibility vs. degrees of truth

Our goal is to develop a system of plausible reasoning that extends classical deductive logic—in particular, the propositional calculus—to deal with propositions that we cannot conclusively prove true nor false. Instead of simply giving up in such cases and providing no information at all about the truth of the proposition, our logic should allow us to compute a degree of plausibility for the proposition. Such plausible reasoning constitutes the vast majority of the reasoning that we as humans do, since outside of mathematics there is little that we can declare true or false with complete certainty.

We stress that we are concerned with degrees of *plausibility*, as opposed to degrees of *truth*. Fuzzy logic [9] (with the exception of possibility theory [10]) and various other multivalued logics deal with the latter, and hence have aims distinct from ours. Failure to distinguish these distinct concepts has in the

past led to unnecessary controversy [11]. On this issue, Dubois and Prade [12] write:

> "... [name omitted] fails to understand the important distinction between two totally different problems... These are the handling of *gradual* (thus non-Boolean) properties whose satisfaction is a matter of degree (even when information is complete) on the one hand, and the handling of uncertainty being induced by incomplete states of knowledge...
>
> "Very often, discussions about fuzzy expert systems or uncertain knowledge base systems get confused because of a lack of distinction between degrees of truth and degrees of uncertainty... This distinction was made by one of the founders of subjective probability theory—De Finetti—but with few exceptions (including ourselves) it has been quite forgotten by the AI community in general."

As an example, one's confidence in the statement P ("Jim is over six feet tall"), after seeing Jim sitting at a table, is a degree of plausibility. In contrast, the statement Q ("Jim is tall") may be somewhat true (if Jim measures five feet eleven inches) or entirely true (if Jim measure seven feet even).

Our logic shall be restricted to statements such as P, which are either true or false, although we may not know which. This does not leave us utterly incapable of dealing with statements such as Q. In some cases, we may take statements involving fuzzy concepts (e.g., "Jim has a beard") and, as an engineering approximation, treat them as either entirely true or entirely false. A more general approach is to recognize that such fuzzy statements usually arise as human utterances, and to turn them into propositions by stating that someone uttered them. For example, we may apply plausible reasoning to the proposition Q' ("Mary said that Jim is tall") or Q'' ("I was told that Jim is tall"), rather than to Q itself, with the goal of deriving the plausibility of another proposition such as P.

3   Preliminary definitions

Before proceeding, let us review some notions from the propositional calculus and provide basic definitions we use throughout this paper. A *proposition* is an unambiguous statement that is either true or false. A *compound proposition* is constructed from other propositions using the unary operator ¬ (negation) or any of the binary operators ∧ (and), ∨ (or), ⇒ (implies), or ⇔ (equivalence). All other propositions are *atomic* propositions: they cannot be decomposed

into other propositions. [1]

Definition 1 *A* state of information *X* *summarizes the information we have about some set of atomic propositions* A*, called the* basis *of X, and their relationships to each other. The* domain *of X is the logical closure of* A*, that is, the union of* A *and all compound propositions that involve only atomic propositions from* A*.*

A state of information is not restricted to containing only deductive information; it can also contain soft information that says nothing with certainty, but still affects one's assignment of plausibilities. We do not at this point pin down exactly what states of information are in a formal sense, but we shall characterize them axiomatically.

Definition 2 *If X is a state of information and A is a proposition in the domain of X, we write* $(A \mid X)$ *for the plausibility we assign to A given the information in X. We write A, X for the state of information obtained from X by adding the additional information that A is true.*

Although our notation is reminiscent of the notation of probability theory, please keep in mind that we are *not* introducing probability theory at this point. Note also one difference between the plausibility $(A \mid X)$ and a conditional probability $P(A \mid B)$ from classical, frequentist probability theory: B is a proposition, while X is a state of information. The comma notation allows us to write things like $(A \mid B, X)$, that is, "the plausibility of A given both X and that B is true."

In the sequel we assume that X is a state of information and that A, B, C and D are propositions in the domain of X.

4   Representation of plausibility

R1 $(A \mid X)$*, the plausibility A given X, is a single real number. There exists a real number* T *such that* $(A \mid X) \leq$ T *for every X and A.*

Since we use real numbers to measure everything from time and distance to temperature—indeed, to measure any sort of magnitude—it seems quite reasonable to measure degrees of plausibility this way. At this point the only meaning we can assign to these numbers is that higher numbers indicate higher degrees of plausibility, with T being the plausibility of a known true proposition. One might want to represent truth by $+\infty$, or falsity by $-\infty$, but

---

[1] Atomic propositions may still have some internal structure. For example, we might wish to consider the atomic proposition $x = e$ for any numeric expression e.

any such representation of plausibilities can be mapped into a finite interval of the real number line via a continuous, strictly increasing, invertible transformation—for example, $f(x) = \arctan(x)$.

We must mention, however, that R1 is not without controversy. In fact, it is arguably *the* most fundamental distinction between the Bayesian and other approaches to plausible reasoning.

By representing plausibilities with a single real number we implicitly assume that the plausibilities of any two propositions are comparable. That is, given any two propositions A and B and a state of information X, either A and B are equally plausible, or one is more plausible than the other. Some consider this assumption of *universal comparability* unwarranted, and have explored weaker assumptions. Fine [13] gives one summary of approaches that do away with universal comparability, whereas Jaynes [8, Appendix A] argues for universal comparability on pragmatic grounds.

The most common objection to universal comparability is a more fundamental objection to representing one's degree of certainty or belief in a proposition with a single value. In particular, two popular approaches to plausible inference—belief function theories [14,15] and possibility theory [10]—are *two-dimensional* theories in which one's certainty in a proposition is represented by a pair of numbers. Such theories unavoidably lack universal comparability. With regard to belief-function theory, Shafer [14, p. 42] writes:

> "One's beliefs about a proposition A are not fully described by one's degree of belief $\mathrm{Bel}(A)$, for $\mathrm{Bel}(A)$ does not reveal to what extent one doubts A— i.e., to what extent one believes its negation $\bar{A}$. A fuller description consists of the degree of belief $\mathrm{Bel}(A)$ together with the *degree of doubt* $\mathrm{Dou}(A) = \mathrm{Bel}(\bar{A})$."

Describing possibility theory, Dubois and Prade [10, p. 11] likewise write:

> "The possibility (or necessity) of an event, and that of the contrary event, are but weakly linked; in particular, in order to characterize the uncertainty of an event A one needs both of the numbers $\Pi(A)$ and $N(A)$."

One motivation for using a two-dimensional theory is the concern that a one-dimensional theory cannot adequately represent ignorance. Belief-function theories allow one to represent ignorance by allowing one to express a degree of belief that some one out of a set of possibilities is true, without requiring one to subdivide this into assignments of belief to the individual possibilities. Regarding the representation of ignorance with probabilities, Shafer [14, p. 23, 24] writes:

> "Are there or are there not living beings in orbit around the star Sirius?

Some scientists may have evidence on this question, but most of us will profess complete ignorance about it. So if θ_1 denotes the possibility that there is such life and θ_2 denotes the possibility that there is not, we will adopt the vacuous belief function over the set of possibilities $\Theta = \{\theta_1, \theta_2\}$.

"We can also consider the question in the context of a more refined set of possibilities. We might, for example, raise the question of whether there even exist planets around Sirius. We would then have a set of possibilities $\Omega = \{\xi_1, \xi_2, \xi_3\}$, say, where ξ_1 corresponds to the possibility that there is life around Sirius, ξ_2 corresponds to the possibility that there are planets but no life, and ξ_3 corresponds to the possibility that there are not even planets..."

Shafer then points out that if one tries to represent ignorance about the question by assigning equal probabilities to all the possibilities, one runs into an inconsistency: the probability of the proposition A that there are living beings in orbit around Sirius differs depending on whether one examines the set of possibilities Θ or the more refined set of possibilities Ω.

One response to such concerns is that, in practice, one is never completely ignorant. One generally has at least some weak information about the propositions of interest. Still, the notion of complete ignorance is useful as a limiting case, and as an approximation when the available information is so weak as to not be worth the effort of including it in one's analysis.

A better response is that "complete ignorance" is a slippery concept, and we need to be careful to state exactly wherein our ignorance lies. Returning to Shafer's example, the set of possibilities Θ, and the assignment of equal probabilities to these, corresponds to a state of information wherein we are completely ignorant of any factors relevant to A. But the very act of including ξ_1 instead of ξ_1' (life and one or more planets) and ξ_1'' (life but no planets) in the set of possibilities Ω reveals that we are *not* completely ignorant—we are, in fact, aware that A cannot be true unless there are one or more planets in orbit about Sirius. Assignment of equal probabilities to the elements of Ω corresponds to a state of information wherein we know that planets are necessary for life, but that is all we know. There is no inconsistency in arriving at different probabilities for A when using the two different sets of possibilities, as they represent different states of information. [2]

Can a one-dimensional theory adequately represent ignorance? A reasonable discussion of this question would take us well beyond the scope of this paper, but we can make a few comments. Bayesians have not ignored this issue. Jaynes suggests treating each aspect of one's ignorance as defining a transformation of the problem that leaves one's state of information invariant, and

---

[2] We believe this response to Shafer originated with E. T. Jaynes, but we have been unable to track down a specific reference.

using this to derive a probability distribution representing ignorance [16,17]. Though quite elegant, this approach does not always yield a unique solution: too few transformations can yield multiple candidate distributions, and too many can leave no solution at all. Kass and Wasserman have written a useful review of techniques for constructing probability distributions representing ignorance [18]. Their paper illustrates a difficulty Bayesians face in representing ignorance: there are too many choices available, as the issue is not sufficiently resolved to allow widespread agreement on the proper solution. On the other hand, the rapidly growing use of Bayesian methods by working scientists [19] suggests that the issue need not be a serious stumbling block in practice.

Another motivation for two-dimensional theories has been the perception that proper application of Bayesian methods requires knowledge of the "true" probabilities, which are viewed as physical properties. This can lead to theories in which one represents uncertainty as a convex set of probability distributions [20]. Jaynes dismisses such concerns about "true," physical probabilities as examples of the "Mind Projection Fallacy" [21], yet even the Jaynesian viewpoint admits certain states of information that are mathematically equivalent to being uncertain about some "physical" probability.[3] Bayesians deal with uncertain "physical" probabilities by reasoning about the probabilities of various physical probability values. This brings us back to the previous concern, representation of ignorance, and hence this second motivation for considering a two-dimensional theory reduces to the first.

We close our discussion of R1 with one final, pragmatic argument in its favor: it is the simplest alternative available. Two-dimensional theories are unavoidably more complex, and in applications to decision-making under uncertainty they lack the simple (and widely used) principle of maximizing expected utility. Thus, R1 seems a desirable property for a system of plausible inference to have, as long as it does not lead to an unsatisfactory theory.

5 Compatibility with the propositional calculus

Definition 3 *We say that A is equivalent to B if* $(A \Leftrightarrow B)$ *is a tautology.*[4]

---

[3] Consider repeated draws, with replacement, from an urn full of black balls and white balls, with the urn thoroughly shaken between draws. Detailed information on the initial positions of the balls and on the shaking and drawing process would allow us to predict each draw with certainty, but in the absence of such information the fraction of white balls determines the probability of drawing a white ball on each turn.

[4] A proposition that is true regardless of the truth or falsity of its atomic propositions, e.g., $(C \vee \neg C)$.

Definition 4 *We say that X is* consistent *if there is no proposition A for which both $(A \mid X) = \mathsf{T}$ and $(\neg A \mid X) = \mathsf{T}$.*

A state of information is consistent if it doesn't assert the truth of two contradictory propositions. $\neg A, A, X$ is an example of an inconsistent state of information.

R2 *Plausibility assignments are compatible with the propositional calculus:*

(1) If A is equivalent to A' then $(A \mid X) = (A' \mid X)$.
(2) If A is a tautology then $(A \mid X) = \mathsf{T}$.
(3) $(A \mid B, C, X) = (A \mid (B \wedge C), X)$.
(4) If X is consistent and $(\neg A \mid X) < \mathsf{T}$, then A, X is also consistent.

Let's consider why one might desire each of the above requirements in turn:

(1) If A and B are equivalent, we should be able to use them interchangeably. [5]
(2) In accordance with the propositional calculus, any tautology is known true, hence our plausibility assignment should reflect this.
(3) If we know that B is true, and we also know that C is true, then we know that $B \wedge C$ is true. Likewise, if we know that $B \wedge C$ is true, we know that B and C individually are true.
(4) If we can't say with certainty that $\neg A$ is true, then there remains a possibility that A is true, thus A must not contradict the information we have.

Paris's treatment of Cox's Theorem has no equivalent to the last two requirements above. This is because he does not make states of information explicit. What we write as $(A \mid B, X)$, Paris writes as $Bel(A \mid B)$, with X implicit. What we write as $(A \mid X)$, Paris would write as $Bel(A \mid B)$ for some tautology B. We have chosen to make states of information explicit and axiomatize them, as this helps to make clear the rationale for the universality requirement we give below. Furthermore, making states of information explicit is of great practical use in avoiding errors arising from unwitting use of different states of information in different parts of an analysis.

6 Negation

R3 *There exists a nonincreasing function S_0 such that $(\neg A \mid X) = S_0(A \mid X)$ for all A and consistent X.*

---

[5] Recall that equivalence here means only that $(A \Leftrightarrow B)$ is a tautology, which is a decidable question.

A and $\neg A$ are just flip sides of the same question. For the propositional calculus, there is a simple functional relation between the truth values of A and $\neg A$: if we know that one is true, then we know that the other is false. Thus it may seem natural to extend this to a functional relation between $(A \mid X)$ and $(\neg A \mid X)$.

If, to the contrary, we allow $(A \mid X)$ and $(\neg A \mid X)$ to vary independently of each other, then we effectively have a two-dimensional theory, as we need *two* numbers to completely characterize our uncertainty about A [14, p. 42][10, p. 11]. Thus we see that R3 goes hand in hand with R1 [22], and all of the arguments we have presented for or against the latter also apply to the former.

R3 also states that S_0 is nonincreasing, i.e., if $(A \mid X) < (A \mid Y)$ then $S_0(A \mid X) \geq S_0(A \mid Y)$. This just states that if we become more certain that A is true (our state of information changes from X to Y), we should not also become more certain that A is false. We note that Cox's original proof has no explicit requirement that S_0 be nonincreasing, instead requiring that it be twice differentiable[6]; however, a close examination of the appendix of Cox's paper reveals that his proof implicitly requires S_0 to be either strictly increasing or strictly decreasing ($S_0'(x) \neq 0$).

The following are immediate consequences of these first requirements:

Proposition 1 *Define* $\mathsf{F} = S_0(\mathsf{T})$. *Then* $\mathsf{F} \leq (A \mid X) \leq \mathsf{T}$ *for all A and consistent X.*

Proof. $(A \mid X) = S_0(\neg A \mid X) \geq S_0(\mathsf{T})$. $\square$

Proposition 2 *If X is consistent and $x = (A \mid X)$ then $x = S_0(S_0(x))$.*

Proof. $x = (\neg\neg A \mid X) = S_0(S_0(A \mid X))$. $\square$

7   Universality

R4 *There exists a nonempty set of real numbers* P_0 *with the following two properties:*

- P_0 *is a dense subset of* (F, T). *That is, for every pair of real numbers a, b*

---

[6] Actually, instead of our S_0 Cox postulates a function s such that $w(\neg A \mid X) = s(w(A \mid X))$ (we define w in Section 10). Given our requirements w is continuous and strictly increasing, so the two approaches are equivalent, but the distinction becomes important if we relax the strictness requirement on F of Section 10 [23].

such that $\mathsf{F} \leq a < b \leq \mathsf{T}$, *there exists some* $c \in \mathsf{P}_0$ *such that* $a < c < b$.
- *For every* $y_1, y_2, y_3 \in \mathsf{P}_0$ *there exists some consistent* X *with a basis of at least three atomic propositions—call them* A_1, A_2 *and* A_3—*such that* $(A_1 \mid X) = y_1$, $(A_2 \mid A_1, X) = y_2$ *and* $(A_3 \mid A_2, A_1, X) = y_3$.

To understand the motivation for R4, recall the purpose of this enterprise: to construct a *universal* system or logic of plausible reasoning, intended as an extension of the propositional calculus. Since the propositional calculus is applicable to *any* problem domain for which we can formulate useful propositions, the same should be true for our logic of plausible reasoning; in particular, the least we can ask is that our logic of plausible reasoning be capable of handling a case where we have three completely unrelated atomic propositions with arbitrary plausibilities.

R4 requires a bit less than that. First of all, we do not require our logic to handle completely arbitrary plausibilities. We allow for the possibility that certain real numbers in the range $[\mathsf{F}, \mathsf{T}]$ simply aren't allowed as plausibilities. For example, we could restrict ourselves to rational values. However, we do require that the set P_0 of allowed plausibility values be dense. We want our theory to have no holes, no entire intervals of forbidden plausibility values, as this would unduly restrict its applicability.

Secondly, if the propositions A_i are completely unrelated, then knowing that one of the propositions is true should not change the plausibility of the others; that is, we would expect that $(A_2 \mid A_1, X) = (A_2 \mid X)$, $(A_3 \mid A_2, A_1, X) = (A_3 \mid A_1, X) = (A_3 \mid X)$, etc. We have not required this, but R4 is consistent with such an interpretation, for it is satisfied by first choosing arbitrary values for the various $(A_i \mid X)$, then using the above equalities.

R4 is not without controversy. Paris [4] highlights its crucial importance in his proof of Cox's Theorem, and Halpern [24] shows that omitting R4 allows one to construct an explicit counterexample to Cox's Theorem. Halpern goes further to argue that R4 is unreasonable for finite domains (those with only a finite set of atomic propositions). Following Cox, Halpern conditions on propositions rather than states of information, and he writes W for the set of pairs of propositions (A, B) (corresponding to plausibility expression $(A \mid B)$) constructed from the set of atomic propositions, with equivalent propositions considered equal. He writes [25]:

> "The problematic assumption here is [R4]... [T]o satisfy [R4], W must be infinite; [R4] cannot be satisfied in finite domains. While 'natural' and 'reasonable' are, of course, in the eye of the beholder, it does not strike me as a natural or reasonable assumption in any obvious sense of the words. This is particularly true since many domains of interest in AI (and other application areas) are finite; any version of Cox's Theorem that uses [R4] is

10

simply not applicable in these domains."

Halpern does mention that one might require R4 because one desires a set of rules that apply to arbitrary domains, but then dismisses this motivation because it "does not allow a notion of belief that has only finitely many gradations" [24,25]. (Why one might desire to have only finitely many gradations he does not say.) One may argue to the contrary that any logic of plausible reasoning applicable to only a single domain is of little value. We (humanity) would have found it difficult to make any significant progress in mathematics if we had been required to come up with new rules of logic for every new domain we wished to investigate. It is the very fact that we have identified widely-applicable rules of logic, to be used in nearly every domain[7], that allows us to reason with confidence when entering new conceptual territory.

Still, let us consider doing as Halpern proposes, and restrict ourselves to one finite set of atomic propositions $\mathcal{A}$ used in one problem domain. Halpern notes that W is then finite, and hence the set of plausibility values $(A \mid B)$ is also finite; but the traditional notation $(A \mid B)$, with A and B both propositions, hides a dependence on one's prior state of information X. Using the notation of this paper, the set of possible plausibility values are all values $(A \mid B, X)$, where X may range over any allowed state of information whose basis is $\mathcal{A}$. Thus, the set of possible plausibility values is finite only if, in addition to restricting ourselves to a single, finite problem domain, we also restrict ourselves to a finite set of possible states of information for that problem domain.

Snow [28] argues against such a restriction and for infinite gradations of plausibility within even a single, finite domain:

> "It often happens that sentences of interest include some that describe events for which there is an 'objective' probability...
>
> "The source of an objective rational measure of belief is external to the cognitive apparatus of the believer. Its value is determined by the vagaries of the real world or by some idealized model of that world. There is no way to tell in advance just which values might arise, and each value may be graduated with arbitrary precision. Any such value can simply be adopted by the believer without recourse to unboundedly precise discrimination between affective states related to credibility..."

For example, consider the proposition A that a particular atom of a radioactive isotope will decay within a particular time period. The plausibility that we assign A will depend on what we determine the half-life of the isotope to be, and has a continuum of potential values before the measurement is made.

---

[7] The predicate calculus (plus the axioms of set theory) seems to be adequate for almost all of mathematics and science. Many feel it is inadequate for certain domains, however, and have proposed various alternatives [26,27].

Let's examine a variant of an example that Snow gives, illustrating a continuum of plausibility values arising from different states of information for the same problem domain, but not relying on any notion of "physical" probabilities. Suppose we have a problem in a finite domain involving a proposition B whose meaning is "point p lies on the vertical leg of right triangle T." Let $X(x, y)$ denote a state of information in which the only thing we know about the location of p is that it lies on the perimeter of T, which has width x and height y. It seems reasonable to assign $(B \mid X(x, y))$ a value near F if $y \ll x$, a value near T if $y \gg x$, and that $(A \mid X(x, y))$ should increase smoothly as y/x increases from 0 to infinity.

Arnborg and Sjödin have investigated an alternative form of Cox's Theorem that does not rely on universality, but instead uses notions of non-informative refinability and information independence for finite domains [29,30]. Other differences include replacing R3 with a requirement that $(A \lor B \mid X)$, where A and B are exclusive, be a strictly increasing function of $(A \mid X)$ and $(B \mid X)$.

One trivial consequence of adding R4 is that our logic does not collapse to a single possible plausibility value:

Proposition 3 $\mathsf{F} < \mathsf{T}$.

Proof. Follows from the fact that P_0 is nonempty. $\square$

8 Properties of S_0

Our requirements to this point imply some additional properties for the function S_0.

Lemma 4 *There exists a continuous, strictly decreasing function $S_1 : [\mathsf{F}, \mathsf{T}] \rightarrow [\mathsf{F}, \mathsf{T}]$ such that $S_1(A \mid X) = S_0(A \mid X) = (\neg A \mid X)$ for all A and consistent X.*

Proof. Let P_1 be the set of all possible plausibility values $(A \mid X)$, where X is consistent, and restrict the domain of S_0 to P_1. If $x_1, x_2 \in \mathsf{P}_1$ and $S_0(x_1) = S_0(x_2) = y$, then $x_1 = S_0(y) = x_2$; hence S_0 is one-to-one, and therefore strictly decreasing. Proposition 2 tells us that P_1 is the range of S_0; combined with the facts that P_1 is a dense subset of $[\mathsf{F}, \mathsf{T}]$ and S_0 is nonincreasing, this implies that S_0 is continuous (any discontinuity would produce a gap in the range of S_0.) The lemma is then proved by defining $S_1(x) = \lim_{y \rightarrow x} S_0(y)$. $\square$

Paris [4] shows that, if we strengthen R4 to require that $\mathsf{P}_0 = [\mathsf{F}, \mathsf{T}]$, then we can dispense with the requirement that S_0 be nonincreasing, as we can derive

this property from the other requirements.

9 Conjunction

R5 *There exists a continuous function* $F : [\mathsf{F}, \mathsf{T}]^2 \to [\mathsf{F}, \mathsf{T}]$, *strictly increasing in both arguments on* $(\mathsf{F}, \mathsf{T}]^2$, *such that* $(A \wedge B \mid X) = F((A \mid B, X), (B \mid X))$ *for any A, B and consistent X.*

Let's first examine why it might be reasonable to require $(A \wedge B \mid X)$ to be a function of $(B \mid X)$ and $(A \mid B, X)$ only. The obvious candidates on which $(A \wedge B \mid X)$ might depend are $(A \mid X)$, $(B \mid X)$, $(A \mid B, X)$, and $(B \mid A, X)$. There are fifteen different subsets of these four values from which we might compute $(A \wedge B \mid X)$; however, $A \wedge B$ is equivalent to $B \wedge A$, and this symmetry reduces the number of distinct candidates to nine. In a similar fashion as Tribus [5], we try to rule out as many of these nine candidates as we can:

(1) $(A \wedge B \mid X) = F(A \mid X)$. Suppose that A is a tautology and B any atomic proposition. Then $A \wedge B$ is equivalent to B, and $(A \mid X) = \mathsf{T}$, so $F(\mathsf{T}) = (B \mid X)$. But we can choose X so as to make $(B \mid X)$ be any desired value in the infinite set P_0, so we have a contradiction.
(2) $(A \wedge B \mid X) = F(A \mid B, X)$. Suppose that A and B are the same atomic proposition. Then $(A \wedge B \mid X) = (A \mid X)$ and common sense dictates that $(A \mid B, X) = \mathsf{T}$, hence $F(\mathsf{T}) = (A \mid X)$, yielding a contradiction.
(3) $(A \wedge B \mid X) = F((A \mid X), (A \mid B, X))$. Suppose that A is a tautology and B any atomic proposition. Then $F(\mathsf{T}, \mathsf{T}) = (B \mid X)$, yielding a contradiction.
(4) $(A \wedge B \mid X) = F((A \mid B, X), (B \mid A, X))$. Suppose that A and B are the same atomic proposition. Common sense dictates that $(A \mid B, X) = (B \mid A, X) = \mathsf{T}$, so $F(\mathsf{T}, \mathsf{T}) = (A \mid X)$, yielding a contradiction.
(5) $(A \wedge B \mid X) = F((A \mid X), (B \mid X))$. Let A be any atomic proposition and let B be $\neg A$. Then $\mathsf{F} = F((A \mid X), S_1(A \mid X))$, hence $F(x, S_1(x)) = \mathsf{F}$ for all $x \in \mathsf{P}_0$. If instead A and B are the same atomic proposition, we obtain $F(x, x) = x$ for all $x \in \mathsf{P}_0$. These equalities lead to the following three undesirable consequences:
   (a) *F must be discontinuous.* To see this, assume that F is continuous. Then $F(x, S_1(x)) = \mathsf{F}$ for all $x \in [\mathsf{F}, \mathsf{T}]$ and $F(x, x) = x$ for all $x \in [\mathsf{F}, \mathsf{T}]$. Since S_1 is a continuous and strictly decreasing function whose domain and range are both $[\mathsf{F}, \mathsf{T}]$, it has a fixed point in the interior of this range; that is, there exists some $\mathsf{F} < x_0 < \mathsf{T}$ such that $x_0 = S_1(x_0)$. Then $x_0 = F(x_0, x_0) = F(x_0, S_1(x_0)) = \mathsf{F}$, which contradicts $x_0 > \mathsf{F}$.
   (b) *It is not allowed for a proposition and its negation to be equally plau-*

sible. For if $x = (A \mid X) = (\neg A \mid X)$, we have $x = F(x, x) = F(x, S_1(x)) = \mathsf{F}$ and so $x = S_1(x) = \mathsf{F}$; but $S_1(\mathsf{F}) = \mathsf{F}$ is not possible, since S_1 is strictly decreasing and its range is $[\mathsf{F}, \mathsf{T}]$.

(c) *There are plausibility values* $x, y > \mathsf{F}$ *such that* $F(x, y) = \mathsf{F}$. This is undesirable because it means that if $(A \mid X) = x$ and $(B \mid X) = y$, then $(A \wedge B \mid X)$ must be known false with absolute certainty, even when A and B are entirely unrelated propositions, each individually possible. This consequence arises as follows. Choose any arbitrary $x \in \mathsf{P}_0$ and define $y = S_1(x)$. Since $S_1(\mathsf{T}) = \mathsf{F}$ and S_1 is strictly decreasing, we have $y > \mathsf{F}$. Then $F(x, y) = F(x, S_1(x)) = \mathsf{F}$.

We are left with the following possibilities:

(1) $(A \wedge B \mid X) = F((A \mid B, X), (B \mid X))$.
(2) $(A \wedge B \mid X) = F((A \mid B, X), (B \mid X), (A \mid X))$.
(3) $(A \wedge B \mid X) = F((A \mid B, X), (B \mid X), (B \mid A, X))$.
(4) $(A \wedge B \mid X) = F((A \mid B, X), (B \mid X), (B \mid A, X), (A \mid X))$.

Tribus claims to rule out (3) and (4) by showing that for these candidates one cannot avoid dealing with plausibility expressions that involve some inconsistent state of information Y, and hence there is unavoidable ambiguity. We find this argument unconvincing, as there is no problem with simply picking an arbitrary value for $(A \mid Y)$ when Y is inconsistent. From deductive logic we know that one can prove anything from inconsistent premises, and so there is even a good argument for defining $(A \mid Y) = \mathsf{T}$ for all propositions A and inconsistent Y. Tribus also claims that one can rule out (2), but leaves this as an exercise for the student.

At this point we have to admit that there is no completely compelling reason for choosing any particular one of the four remaining candidates. However, (1) seems intuitively appealing to many people (it has not engendered any controversy of which we are aware); furthermore, the other candidates merely add additional arguments to those used in (1)—whatever we decide, we are going to need at least $(B \mid X)$ and $(A \mid B, X)$—and the simpler candidate is to be preferred, all else being equal.

Jaynes [8, Chapter 2] gives the following intuitive rationale for (1):

"In order for $A \wedge B$ to be a true proposition, it is necessary that B is true. Thus the plausibility $(B \mid X)$ should be involved. In addition, if B is true, it is further necessary that A should be true; so the plausibility of $(A \mid B, X)$ is also needed. But if B is false, then of course $A \wedge B$ is false independently of whatever one knows about A, as expressed by $(A \mid \neg B, X)$; if the robot reasons first about B, then the plausibility of A will be relevant only if B is true. Thus, if the robot has $(B \mid X)$ and $(A \mid B, X)$ it will not need $(A \mid X)$. That would tell it nothing about $A \wedge B$ that it did not have already."

(We have replaced Jaynes's notation with our own in the above quote.)

Let us now consider the remaining requirements on F.

- *F must be strictly increasing.* Suppose that our state of information changes so as to make either B more plausible or make A (assuming B) more plausible, while leaving the other no less plausible. Surely $A \wedge B$ must not become *less* plausible in this case. It accords with many people's intuition that $A \wedge B$ must, in fact, be considered *more* plausible in this case, but there are others who disagree with this stronger requirement.
- *F must be continuous.* If one's state of information changes so that either B becomes infinitesimally more plausible, or A (assuming B) becomes infinitesimally more plausible, many would find it quite unnatural and counterintuitive for the plausibility of $A \wedge B$ to suddenly jump.

Cox's 1946 paper does not explicitly require that F be strictly increasing, instead requiring that it be twice differentiable; however, a careful examination of the appendix to his paper reveals that the proof implicitly assumes F to be either strictly increasing in both arguments or strictly decreasing in both arguments on $(\mathsf{F}, \mathsf{T}]^2$.

It is possible to slightly weaken our requirements on F without losing our main result (that our system must be isomorphic to probability theory) [31]. Unfortunately, we cannot get by with requirements nearly as weak as those on S_0. In particular, the requirement that F be strictly increasing is essential; if we relax this to a requirement that F be merely nondecreasing, then $F(x, y) = \min(x, y)$ is consistent with our requirements [22,23], giving us a system not isomorphic to probability theory.

10   The Product Rule

We have now presented all of the requirements we impose on our system of plausible reasoning. We now proceed to show that those requirements force our system of plausible reasoning to be isomorphic to probability theory, giving us the Bayesian approach to plausible reasoning. Our proof borrows from Jaynes [3] and Paris [4].

We begin by deriving some properties of F. The simple fact that "$\wedge$" is associative turns out to have great ramifications. It ensures that F is associative, which in turn limits the possibilities for F to functions that are isomorphic to multiplication.

Lemma 5 $F(x, F(y, z)) = F(F(x, y), z)$ *for all* $x, y, z \in [\mathsf{F}, \mathsf{T}]$.

15

Proof. Let us consider $(A \wedge B \wedge C \mid X)$, where X is consistent. Applying R5 and R2, we obtain

$$
\begin{aligned}
&(A \wedge (B \wedge C) \mid X) \\
={}& F[(A \mid (B \wedge C), X), \, (B \wedge C \mid X)] \\
={}& F[(A \mid B, C, X), \, F[(B \mid C, X), \, (C \mid X)]]
\end{aligned}
$$

Using an alternate grouping, we obtain

$$
\begin{aligned}
&((A \wedge B) \wedge C \mid X) \\
={}& F[(A \wedge B \mid C, X), \, (C \mid X)] \\
={}& F[F[(A \mid B, C, X), \, (B \mid C, X)], \, (C \mid X)].
\end{aligned}
$$

But $A \wedge (B \wedge C)$ is equivalent to $(A \wedge B) \wedge C$, so equating the two plausibilities above and applying R4 yields

$$
F(x, F(y, z)) = F(F(x, y), z)
$$

for all $x, y, z \in \mathsf{P}_0$. Since F is continuous and P_0 is dense, the equation holds for all $x, y, z \in [\mathsf{F}, \mathsf{T}]$. $\square$

The functional equation we have obtained for F is one with a long history extending back to the early 19th century. Here is its solution.

Lemma 6 (Aczél) *Let a and b, with $a < b$, be real numbers. Suppose that $f : (a, b]^2 \to (a, b]$ is a continuous function, strictly increasing in both arguments, and satisfies the associativity equation*

$$
f(x, f(y, z)) = f(f(x, y), z)
$$

for all $x, y, z \in (a, b]$. Then there exists some continuous, strictly increasing function g such that

$$
g(f(x, y)) = g(x) + g(y)
$$

for all $x, y \in (a, b]$.

Proof. This theorem is found on p. 256 of Aczél's book [2], in a slightly different and more general form. $\square$

We now obtain the familiar product rule from probability theory.

Lemma 7 *There exists a continuous, strictly increasing, nonnegative function w such that*

$$
w(A \wedge B \mid X) = w(A \mid B, X)w(B \mid X)
$$

for every A, B and consistent X.

Proof. R5 and Lemma 5 allow us to apply Lemma 6, with $f = F$, $a = \mathsf{F}$, and $b = \mathsf{T}$. Define $w(x) = \exp(g(x))$; then $w(F(x,y)) = w(x)w(y)$ for $x, y \in (\mathsf{F}, \mathsf{T}]$. Since w is increasing, continuous, and nonnegative on $(\mathsf{F}, \mathsf{T}]$, $\lim_{x \to \mathsf{F}} w(x)$ exists and is nonnegative. Defining $w(\mathsf{F})$ to be this limit makes w increasing, continuous, and nonnegative on $[\mathsf{F}, \mathsf{T}]$. By continuity of w and F, we then have $w(F(x,y)) = w(x)w(y)$ for $x, y \in [\mathsf{F}, \mathsf{T}]$, and combined with R5 this gives us the lemma. $\square$

The function w rescales plausibilities to what one might call "proto-probabilities." Lemma 7 states that proto-probabilities obey the product rule of probability theory. The next lemma states that proto-probabilities have the same range of values, and represent truth and falsity in the same way, as probabilities.

Lemma 8 $w(\mathsf{F}) = 0$, $w(\mathsf{T}) = 1$, *and* $0 < w(x) < 1$ *for* $\mathsf{F} < x < \mathsf{T}$.

Proof. For $x > \mathsf{F}$ we have $w(x) > w(\mathsf{F}) \geq 0$. Choose any A, tautology D, and consistent X with $(A \mid X) > \mathsf{F}$; then

$$w(A \mid X) = w(D \wedge A \mid X) = w(D \mid A, X)w(A \mid X) = w(\mathsf{T})w(A \mid X).$$

Dividing both sides of the above equation by $w(A \mid X)$ gives $w(\mathsf{T}) = 1$, and hence $w(x) < 1$ for $x < \mathsf{T}$.

Suppose that $w(\mathsf{F}) = z > 0$. Then $0 < z < 1$, hence $z < \sqrt{z} < 1$. Choose A_1, A_2, consistent X, and $x \in \mathsf{P}_0$ such that $w^{-1}(z) < x < w^{-1}(\sqrt{z})$ and $(A_1 \mid X) = (A_2 \mid A_1, X) = x$. Then

$$w(A_1 \wedge A_2 \mid X) = w(x)^2 < z = w(\mathsf{F})$$

and hence $(A_1 \wedge A_2 \mid X) < \mathsf{F}$, a contradiction. So $w(\mathsf{F}) = 0$. $\square$

11 The Sum Rule

Having established that F must amount to multiplication under the mapping from plausibilities to proto-probabilities, we now investigate what form S_1 must take. We begin by using w to construct what will turn out to be a mapping from plausibilities to probabilities, and examine the behavior of S_1 under this mapping.

Definition 5 *We define* $S(x)$ *to be* $p(S_1(p^{-1}(x)))$, *where*

- $p(x) = w(x)^r$,
- $r = -(\log 2)/(\log w(\alpha))$, *and*

17

- α is the unique fixed point of S_1, i.e., $S_1(\alpha) = \alpha$ and $\mathsf{F} < \alpha < \mathsf{T}$.

Lemma 9 p^{-1} is well defined and has the following properties:

(1) p is continous and strictly increasing.
(2) $p(\mathsf{F}) = 0$, $p(\mathsf{T}) = 1$, and $0 < p(x) < 1$ for $\mathsf{F} < x < \mathsf{T}$.
(3) $p(A \wedge B \mid X) = p(A \mid B, X)p(B \mid X)$ for all A, B, and consistent X.

S is also well defined, and $S(1/2) = 1/2$.

Proof. S_1 has a fixed point because it is continuous and maps the interval $[\mathsf{F}, \mathsf{T}]$ onto itself. The fixed point is unique because S_1 is strictly decreasing, hence α is well-defined and $\mathsf{F} < \alpha < \mathsf{T}$. Then $0 < w(\alpha) < 1$, and so $-\infty < \log w(\alpha) < 0$. This furthermore implies that $0 < r < \infty$. Combined with the continuity and strictness of w, this gives (1). Item (2) follows from the facts that $r > 0$ and w has the same properties. Item (3) follows from $(ab)^r = a^r b^r$ for all a, b, and the corresponding property for w. Item (1) implies that p is one-to-one, hence the inverse p^{-1} is well defined.

Substituting in the definition of r gives $p(\alpha) = w(\alpha)^r = 1/2$. Hence,

$$S(1/2) = p(S_1(p^{-1}(1/2))) = p(S_1(\alpha)) = p(\alpha) = 1/2.$$

$\square$

We now restate R3, R4, and Proposition 2 in terms of p and S.

Proposition 10 The following are true:

- $p(\neg A \mid X) = S(p(A \mid X))$ for every A and consistent X.
- S is continuous and strictly decreasing.
- Let $\mathsf{P} = p(\mathsf{P_0})$. Then P is a dense subset of $(0, 1)$, and for every $y_1, y_2, y_3 \in \mathsf{P}$ there exists some consistent X with a basis of three atomic propositions— call them A_1, A_2 and A_3—such that $p(A_1 \mid X) = y_1$, $p(A_2 \mid A_1, X) = y_2$ and $p(A_3 \mid A_2, A_1, X) = y_3$.
- $S(S(x)) = x$ for all $0 \leq x \leq 1$.

Proof. These follow straightforwardly from our previous results. $\square$

As with F, we now derive a functional equation for S from purely logical considerations.

Lemma 11 For all $0 < x \leq y < 1$, $yS(x/y) = S(x)S(S(y)/S(x))$.

18

Proof. For any $u, y \in \mathsf{P}$ we can find propositions A and B and a consistent X such that $y = p(B \mid X)$ and $u = p(A \mid B, X)$. Let $x = uy = p(A \wedge B \mid X)$. Note that $0 < x, y, u < 1$. Then

$$yS(x/y) = yS(u) = p(B \mid X)p(\neg A \mid B, X) = p(\neg A \wedge B \mid X).$$

Define $C \equiv (\neg A \vee \neg B)$ and $D \equiv (A \vee \neg B)$. Since $\neg B$ is equivalent to $D \wedge C$, we have

$$S(y) = p(D \wedge C \mid X)$$
$$S(x) = p(C \mid X).$$

These last two equalities then give

$$
\begin{aligned}
S(x)S\left(\frac{S(y)}{S(x)}\right) &= S(x)S\left(\frac{p(D \wedge C \mid X)}{p(C \mid X)}\right) \\
&= S(x)S(p(D \mid C, X)) \\
&= p(C \mid X)p(\neg D \mid C, X) \\
&= p(\neg D \wedge C \mid X) \\
&= p(\neg A \wedge B \mid X) \\
&= yS(x/y).
\end{aligned}
$$

The third step relies on the fact that $p(\neg C \mid X) = x < 1$ and hence C, X is consistent (R2.4). Since P is dense and S is continuous, the equality then holds for all $0 < y < 1$ and $0 < u \le 1$, hence for all $0 < x \le y < 1$. $\square$

Lemma 12 Let $s : [0, 1] \to [0, 1]$ be a strictly decreasing and continuous function, with $s(0) = 1$, $s(1) = 0$, and $s(1/2) = 1/2$. If s satisfies both of the functional equations

- $s(s(x)) = x$,
- $ys(x/y) = s(x)s(s(y)/s(x))$

for all $0 < x \le y < 1$, then $s(x) = 1 - x$ for all $0 \le x \le 1$.

Proof. Lemmas 3.10 through 3.15 of Paris [4] amount to a proof of this assertion. $\square$

We now obtain the sum rule of probability theory:

Lemma 13 $p(\neg A \mid X) = 1 - p(A \mid X)$ for all A and consistent X.

Proof. Follows directly from Lemma 12, Lemma 11, Proposition 10, and Lemma 9. $\square$

We can now summarize all of these results into one theorem that states that our plausibilities must obey the rules of probability theory after mapping by the function p.

Theorem 14 *There exists a continuous, strictly increasing function p such that, for every A, B and consistent X,*

(1) $p(A \mid X) = 0$ iff A is known to be false given the information in X.
(2) $p(A \mid X) = 1$ iff A is known to be true given the information in X.
(3) $0 \leq p(A \mid X) \leq 1$.
(4) $p(A \wedge B \mid X) = p(A \mid X)p(B \mid A, X)$.
(5) $p(\neg A \mid X) = 1 - p(A \mid X)$ if X is consistent.

Items 1–5 of Theorem 14 are just the basic rules of probability theory. Since p is invertible we lose no information by working with probabilities only instead of the original plausibilities; thus, any system of plausible reasoning that is not isomorphic to probability theory must necessarily violate one of the requirements we have presented. But how do we know that our requirements are not contradictory? How do we know that there is *any* system of plausible reasoning (that is, choice for F, T, F, S_0, P_0, $(\cdot \mid \cdot)$, and definition of a state of information) that satisfies all of our requirements? The set-theoretical approach to probability theory may be taken as an existence proof that our requirements are not contradictory, by taking states of information to be probability distributions, and defining A, X to be the probability distribution obtained from X by conditioning on the set of values for which A is true. In the terminology of mathematical logic, set-theoretical probability theory then becomes the model theory for our logic, a tool to enable us to construct consistent sets of axioms (plausibility assignments from which we derive other plausibilities).

Jaynes's approach to Bayesian inference is built entirely on Theorem 14 and (what amounts to) R2, rejecting any result not derivable from these.[8] Results for continuous domains are accepted only when they are the well-defined limit of results for finite domains [8, Chaps. 2, 15, and App. B].

Finally, we note that it is convenient, but not necessary, to extend R2 to include the following:

- If X is inconsistent then $(A \mid X) = \mathsf{T}$.
- If $(B \mid X) = \mathsf{F}$ then $(A \mid B, X) = \mathsf{T}$.

---

[8] Assignment of the prior probabilities from which inference proceeds is a separate issue, to which Jaynes applies transformation invariance arguments [16] and the principle of maximum entropy [32].

Adding these rules often allows us to drop requirements such as "X is consistent" or "$p(B \mid X) > 0$" from the statements of theorems. These rules are analogous to the result from the propositional calculus that anything can be proven from inconsistent premises. Conditional probabilities can be defined in set-theoretical probability theory so as to make these rules true, and hence the extended axiom set remains consistent.

13 Conclusion

We have discussed a set of qualitative requirements on systems of plausible reasoning that many find intuitively appealing, and have shown how these requirements allow only systems isomorphic to probability theory. We cannot make a compelling case for all of these requirements, however, and there remains disagreement as to the desirability of several of them. The most important area of disageement is perhaps over R1/R3: whether a single number suffices to completely specify one's uncertainty in a proposition.

Other requirements may be proposed, leading to different results. Smets proposes a set of requirements that necessarily lead to belief function theory [33,34]. Relaxing the strictness requirement on F, combined with a relaxation of R3 in the spirit of Cox's original paper, allows linear possibility distributions [23].

References

[1] R. T. Cox, Probability, frequency and reasonable expectation, American Journal of Physics 17 (1946) 1–13.

[2] J. Aczél, Lectures on Functional Equations and Their Applications, Academic Press, 1996.

[3] E. T. Jaynes, How does the brain do plausible reasoning, Tech. Rep. 421, Stanford University Microwave Laboratory, reprinted in [35], pp. 1–23 (1957).

[4] J. B. Paris, The Uncertain Reasoner's Companion: A Mathematical Perspective, Cambridge University Press, 1994.

[5] M. Tribus, Rational Descriptions, Decisions and Designs, Pergamon Press, 1969.

[6] W. Feller, An Introduction to Probability Theory and Its Applications, John Wiley & Sons, 1950.

[7] J. M. Bernardo, A. F. M. Smith, Bayesian Theory, John Wiley & Sons, 1994.

[8] E. T. Jaynes, Probability theory: The logic of science, `http://bayes.wustl.edu/etj/prob.html` (1996).

[9] L. A. Zadeh, G. J. Klir, B. Yuan, Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lofti A. Zadeh, Advances in Fuzzy Systems—Applications and Theory, World Scientific Publishing Company, 1996.

[10] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, 1988.

[11] C. Elkan, The paradoxical success of fuzzy logic, IEEE Expert 9 (4) (1994) 3–8.

[12] D. Dubois, H. Prade, Partial truth is not uncertainty: Fuzzy logic versus possibilistic logic, IEEE Expert 9 (4) (1994) 15–19.

[13] T. L. Fine, Theories of Probability, Academic Press, 1973.

[14] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

[15] P. Smets, The transferable belief model and other interpretations of dempster-shafer's model, in: P. P. Bonissone, M. Henrion, L. N. Kanal, J. F. Lemmer (Eds.), Uncertainty in Artificial Intelligence 6, Elsevier Science Publishers, 1991, pp. 375–383.

[16] E. T. Jaynes, Prior probabilities, IEEE Transactions on System Science and Cybernetics SSC-4 (1968) 227–241, reprinted in [36], pp. 114–130.

[17] E. T. Jaynes, The well-posed problem, Foundations of Physics 3 (1973) 477–493, reprinted in [36], pp. 131–148.

[18] R. E. Kass, L. Wasserman, Formal rules for selecting prior distributions: A review and annotated bibliography, Tech. Rep. 583, Dept. of Statistics, Carnegie-Mellon University, available at `http://www.stat.cmu.edu/www/cmu-stats/tr/` (1993).

[19] D. Malakoff, Bayes offers a 'new way' to make sense of numbers, Science 286 (1999) 1460–1464.

[20] H. E. Kyburg Jr., Bayesian and non-Bayesian evidential updating, Artificial Intelligence 31 (1987) 271–293.

[21] E. T. Jaynes, Probability theory as logic, in: P. F. Fougere (Ed.), Proceedings, Maximumum Entropy and Bayesian Methods, Kluwer Academic Publishers, 1990, also available in a substantially revised and extended form at `http://bayes.wustl.edu/etj/node1.html`.

[22] D. Dubois, H. Prade, The logical view of conditioning and its application to possibility and evidence theories, International Journal of Approximate Reasoning 4 (1) (1990) 23–46.

[23] P. Snow, The reasonableness of possibility from the perspective of Cox, Computational Intelligence 17.

22

[24] J. Y. Halpern, A counterexample to theorems of Cx and Fine, Journal of Artificial Intelligence Research 10 (1999) 67–85.

[25] J. Y. Halpern, Cox's theorem revisited, Journal of Artificial Intelligence Research 11 (1999) 429–435, technical addendum.

[26] D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, vol. II: Extensions of Classical Logic, D. Reidel Publishing Company, 1984.

[27] D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, vol. III: Alternatives to Classical Logic, D. Reidel Publishing Company, 1986.

[28] P. Snow, On the correctness and reasonableness of Cox's Theorem for finite domains, Computational Intelligence 14 (3) (1998) 452–459.

[29] S. Arnborg, G. Sjödin, Bayes rules in finite models, in: Proceedings of the European Conference on Artificial Intelligence, 2000, pp. 571–575.

[30] S. Arnborg, G. Sjödin, On the foundations of Bayesianism, in: A. Mohammad-Djarafi (Ed.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 20th Int'l Workshop, American Institute of Physics, 2001, pp. 61–71.

[31] S. Arnborg, G. Sjödin, A note on foundations of Bayesianism, `http://www.sics.se/arc/papers/99-3-arnborg.ps`.

[32] E. T. Jaynes, Information theory and statistical mechanics I, Physical Review 106 (1957) 620–630.

[33] P. Smets, Qauntifying beliefs by belief functions: An axiomatic justification, in: Proceedings of the 13th Int'l Joint Conf. on Artificial Intelligence, Morgan Kaufmann, 1993, pp. 598–603.

[34] P. Smets, The normative representation of quantified beliefs by belief functions, Artificial Intelligence 92 (1997) 229–242.

[35] G. J. Erickson, C. R. Smith (Eds.), Maximum Entropy and Bayesian Methods in Science and Engineering, Kluwer Academic Publishers, 1988.

[36] E. T. Jaynes, Papers on Probability, Statistics and Statistical Physics, D. Reidel Publishing Company, 1983.