International Journal of Approximate Reasoning 80 (2017) 36–51

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Bridging the intuition gap in Cox's theorem: A Jaynesian argument for universality

Aubrey Clayton [a,*,1], Travis Waddington [b]

[a] 6 Cedarwood Rd., Jamaica Plain, MA 02130, United States
[b] Department of Mathematics, SUNY Stony Brook, Stony Brook, NY 11794, United States

ARTICLE INFO

Article history:
Received 13 March 2016
Received in revised form 30 July 2016
Accepted 4 August 2016
Available online 8 August 2016

Keywords:
Cox
Jaynes
Bayesian
Probability

ABSTRACT

Various attempts have been made to patch the holes in Cox's theorem on the equivalence between plausible reasoning and probability via additional assumptions regarding the density of attainable plausibilities (so-called "universality") and the existence of continuous and strictly monotonic functions for manipulating plausibility values. By formalizing an invariance principle implicit in the work of Jaynes and using it to construct a class of elementary examples, we derive these conditions as theorems and eliminate the need for ad hoc assumptions. We also construct the rescaling function guaranteed by Cox's theorem and thus provide a more direct proof and an intuitive interpretation for the theorem's conclusion.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

In a landmark paper [6], R.T. Cox attempted to show that the quantitative rules of probability theory – for centuries justified primarily by arguments concerning relative proportions and frequencies – were derivable as consequences of basic rules for plausible reasoning, in the sense that any way of quantitatively assigning plausibilities to propositions that preserves the rules of logic must be rescalable to be a system of probabilities. E.T. Jaynes took up this idea, combined with the work of Pólya [17], and used it to describe probability as an extension of Aristotelian logic, which formed the foundation of his aggressive program of Bayesian critique of the "orthodox" statistical practices. For example, in arguing against the usual methods of parameter estimation, he writes [13, Preface, p. xxiii]:

> Lacking the necessary theoretical principles, [frequentist methods] force one to 'choose a statistic' from intuition rather than from probability theory, and then to invent ad hoc devices (such as unbiased estimators, confidence intervals, tail-area significance tests) not contained in the rules of probability theory. Each of these is usable within the small domain for which it was invented but, as Cox's theorems guarantee, such arbitrary devices always generate inconsistencies or absurd results when applied to extreme cases.

* Corresponding author.
E-mail address: aubreyiclayton@gmail.com (A. Clayton).
[1] The views expressed in this publication are wholly those of the author. They do not necessarily represent the views of the author's employer or any of its affiliates; and accordingly, such employer and its affiliates expressly disclaim all responsibility for the content and information contained herein.

http://dx.doi.org/10.1016/j.ijar.2016.08.002
0888-613X/© 2016 Elsevier Inc. All rights reserved.

If logic is Bayesian, so the argument goes, *then anything at odds with Bayesian methods is necessarily illogical.* The basic principles required for assigning plausibility values were not formally axiomatized but rather expressed as elementary "desiderata." Using the ordering of [13], these desiderata were grouped in three categories:

I. Plausibilities are represented by real numbers.
II. Plausibility assignments agree with common sense.
III. Plausibility assignments are logically consistent; that is, two different ways of making the same assignment must result in the same numerical value.

For an excellent explanation and background on the history of the Cox–Jaynes program, see Van Horn [21].

However, as has since been explored by Aczél [1], Tribus [20], Dubois and Prade [8], Paris [16], and Halpern [10] among others, Cox's derivation (essentially repeated by Jaynes) was non-rigorous, in that not all of the necessary assumptions required by "common sense" were made explicit, and the necessary remedies were seen to strain the bounds of common sense, to say the least. In particular, Halpern [10] exhibits a counterexample to Cox's theorem on finite domains, and in a subsequent paper, Halpern [11] suggests additional technical conditions that would suffice to make the conclusions of the theorem true.

Halpern deals with the setting of a function applied to pairs of subsets of a domain. Van Horn [21] compiles a similar set of assumptions phrased in terms of the propositional calculus – plausibilities assigned to propositions – and collates them as[2]:

R1. $(A|X)$, the plausibility of proposition A given information X, is a single real number, $0 \le (A|X) \le 1$.
R2. Plausibility assignments are compatible with the propositional calculus:
  1. If AX is logically equivalent to $A'X$ then $(A|X) = (A'|X)$.
  2. If A is a contradiction given X then $(A|X) = 0$; if A is a tautology given X then $(A|X) = 1$.

R3. There exists a nonincreasing function S_0 such that $(\overline{A}|X) = S_0(A|X)$ for all A and consistent X.
R4. There exists a nonempty set of real numbers P_0 with the following two properties:
  1. P_0 is a dense subset of $[0, 1]$. That is, for every pair of real numbers a, b such that $0 \le a < b \le 1$, there exists some $c \in P_0$ such that $a < c < b$.
  2. For every $y_1, y_2, y_3 \in P_0$ there exists some consistent X and propositions A_1, A_2 and A_3 such that $(A_1|X) = y_1, (A_2|A_1X) = y_2$ and $(A_3|A_2A_1X) = y_3$.

R5. There exists a continuous function $F : [0, 1]^2 \to [0, 1]$, strictly increasing in both arguments on $(0, 1]$, such that $(AB|X) = F((A|BX), (B|X))$ for any A, B and consistent X.

The functions S and F describe global rules for calculating plausibilities of compound statements in terms of their constituent parts. For discussion of the reasoning behind this particular form (excluding other variables from the F function, for instance) see Tribus [20].

Similarly to Halpern, Van Horn is then able to prove the following version of Cox's theorem:

Theorem 1.1 *(Van Horn). If $(\cdot|\cdot)$ is a plausibility rule satisfying the above assumptions, then $(\cdot|\cdot)$ is isomorphic to ordinary probability, in the sense that there exists a continuous, strictly increasing function p such that, for every A, B and consistent X,*

1. $p(A|X) = 0$ *iff A is known to be false given the information in X.*
2. $p(A|X) = 1$ *iff A is known to be true given the information in X.*
3. $0 \le p(A|X) \le 1$.
4. $p(AB|X) = p(A|BX)p(B|X)$.
5. $p(\overline{A}|X) = 1 - p(A|X)$.

Arnborg and Sjödin [2] investigate an alternate version of the theorem assuming noninformative refinability and information independence on finite domains. Dupré and Tipler [9] show a version of Cox's theorem after axiomatizing a "strong rescaling" condition on "plausible value" assignments in an algebra. For further discussion of the theorem in the context of scaled Boolean algebras, see Hardy [12].

While the additional assumptions used by Halpern and Van Horn do provide the necessary rigor to resolve the issues with Cox's theorem, they are somewhat lacking in intuitive support (admittedly, according to both authors). We pick up the line of thinking here and attempt to refine these assumptions to have more commonsense reasoning behind them.

[2] We have modified this slightly so that plausibilities are scaled to the unit interval and to be consistent with the notation of Jaynes. Here and throughout, we use the notation "$\overline{A}$" to mean the proposition "NOT A", "$A + B$" = "A OR B", and "AB" = "A AND B".

1.1. Universality

Perhaps the most troubling condition is R4, which Van Horn labels "universality." This is similar in character to Halpern's "Par5" assumption, referencing Paris [16], which Halpern describes as a potentially necessary piece that was missing from Cox's original work. The essential problem, perhaps first identified by Aczél [1], who proved a stronger version of Cox's theorem (eliminating Cox's assumption of differentiability of F) but with additional assumptions along these lines, is that in order to use the solutions to certain functional equations in this context, we require those equations to hold on an entire real domain and not just for the values that can be realized as plausibilities of propositions.

For example, from the assumption that plausibility respects logical equivalence and the logical relationship

$$(AB)C \iff A(BC)$$

we would have for any propositions A, B, C, and D that

$$F((C|ABD), F((B|AD), (A|D))) = F(F((C|ABD), (B|AD)), (A|D))$$

Letting x, y, and z stand in for the arguments of the function F we have

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

This equation, known, for obvious reasons, as the "associativity equation," has a well-known general solution that would make F isomorphic to multiplication, corresponding with the product rule of probability. However, there is some sleight-of-hand in replacing the plausibilities with variables named x, y, and z; the former equation only applies to input values that can be realized for particular propositions A, B, C, and D whereas the latter seems to apply for all values on the real domain $[0, 1]^3$, which is what's actually assumed for the derivation of the solution.

So, some form of the universality assumption is helpful to ensure the set of plausibilities that can be realized in particular ways is dense in the unit cube; this and the continuity assumptions guarantee that any equation one can show to be true for plausibility numbers will continue to hold for all real values. Halpern [11, pp. 430–431] comments that the assumption of density stands out as an artificiality without much intuition to support it:

> The problematic assumption here is Par5... 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.

One of Halpern's principal objections to the density assumption is that it disallows plausibility functions on finite domains W, and therefore plausibility functions that take only finitely many values, and this is a problem for any applications where the domain of interest must be finite. He suggests a resolution via expanding the domain W to include an infinite set of unneeded possibilities and also offers a variant theorem that requires a plausibility function to be defined simultaneously on an infinite set of finite domains Ω in such a way that the functions F and S are fixed, and a variant of the Par5 assumption holds.

We agree with Halpern's comment that the latter is, in fact, more in the spirit of Jaynes, who objected to the idea of an infinite domain of anything, except as a well-defined mathematical limit of finite domains. According to Jaynes [13, chapter 1, pp. 8–9], we are to imagine we're programming a "robot" with basic rules to compute plausibilities of propositions based on conditioning information:

> In order to direct attention to constructive things and away from controversial irrelevancies, we shall invent an imaginary being. Its brain is to be designed by us, so that it reasons according to certain definite rules. These rules will be deduced from simple desiderata which, it appears to us, would be desirable in human brains; i.e. we think that a rational person, on discovering that they were violating one of these desiderata, would wish to revise their thinking... Our robot is going to reason about propositions.

Something left unspecified, though, is what the domain of those propositions is supposed to be and whether it's infinite. Jaynes would likely argue that it's finitely limited, say by the number of available electrons in the universe needed to store the assumptions in binary, but would allow that for the purposes of mathematical reasoning we should allow it to be infinite, imagining, for example, that we could always tack on an additional clause to a given proposition to form a new proposition. In any event, the underlying calculation engine of the robot should remain fixed for all possible inputs.

Therefore, as Van Horn [22] and Snow [19] argue, to follow Jaynes, our orientation should be that of imagining a single plausibility rule applied simultaneously to all domains. If we're constructing our robot to reason about all manner of propositions, either as statements to assign plausibilities to or as conditioning information, it should not be artificially limited from the start to a certain finite set, and the rules it uses should apply uniformly to any chosen domain.

We assume that S and F are fixed functions satisfying:

Assumption 1.2.

1. $(\overline{A}|B) = S((A|B))$ *for all propositions* A *and states of information* B.
2. $(AB|C) = F((A|BC), (B|C))$ *for all propositions* A, B *and states of information* C.

By asserting the existence of these functions, we mean only that once the relevant "input" plausibilities are known, the corresponding "outputs" are determined, so that, for example, if $(A|B) = (A'|B')$ for any propositions A, A' and information B, B' then $(\overline{A}|B) = (\overline{A'}|B')$. Later on, we will address the properties that S and F are assumed to have as functions and consider extending their domains.

If the background conditioning information X is itself a contradiction, then both A and $\overline{A}$ are derivable from X, so the numerical values above will not be uniquely determined, causing our robot to "crash." For the remainder, then, we assume that any conditioning information is consistent,[3] and we assume that $(\cdot|\cdot)$ is a fixed plausibility rule that satisfies assumptions R1 and R2.

While we do not take up their defense in the present work, we note that assumptions R1 and R2 are not without controversy. Assumption R1 has been criticized by Shafer [18], among others, as being too limited a description of one's belief about a proposition. A perhaps troubling consequence of representing plausibility by a single real number is that any two propositions are comparable; we must always be able to say which is more plausible now matter how disparate they seem. Various multi-dimensional alternatives have been proposed in which propositions are only partially ordered; for a summary and discussion see Van Horn [21] or Jaynes [13, appendix A].

The equivalence described in assumption R2 (and which we will employ throughout this paper) is logical equivalence within classical logic. That is, we assume a formal logical language (capital letters denoting propositions) with bivalent semantics and system of deductive rules satisfying the usual axioms of classical logic including the Law of Excluded Middle, double negative elimination, De Morgan's Laws, etc. Thus, we consider propositions P and Q equivalent if they are derivable from each other, or, equivalently, if $P \iff Q$ is a tautology. Cox's theorem as we explore it, therefore, is a purely mathematical result, generalizing the classical idea of deduction to allow for "partial deduction" – so that a conclusion may not follow deductively from a set of premises but may be said to possess a degree of plausibility given those premises.

Ultimately, Jaynes's ambition was to apply this system of plausible reasoning to natural language propositions in the context of scientific reasoning in the real world, exposing the theory to thorny philosophical issues concerning the vagueness and ambiguity of natural language propositions, or, at a minimum, requiring a deliberate restriction of the domain of reasoning to what he describes as "Aristotelian propositions," i.e., those of "an unambiguous meaning and... of the simple, definite logical type that must be either true or false" [13, Chapter 1, p. 9]. For some discussion of possible extensions of probabilistic reasoning beyond the realm of "Aristotelian propositions," see Jaynes [13, chapter 18].

As far as our robot is concerned, these rules of equivalence are "hard-wired," so that, for example, the equivalence $AB \iff BA$ is always "known" to the robot. If, within the context of a given problem, we expect the robot to understand additional connections between propositions, then those connections must be made explicit in the assumed background information X. For example, we may consider such propositions as

A = "Event E lasts 2 days."
B = "Event E lasts 48 hours."

Naturally we expect A and B to return the same plausibility, so we must be sure to include the fact that a day is 24 hours long in our background set.

Another consequence of these rules is that any statement whose truth value is derivable from our background assumptions must return a plausibility of 1 or 0. So, for example, the robot will eventually return a value of 1 or 0 for statements such as "$2^{2^{61}-1} - 1$ is prime." though it may take an unfathomably long time to do so, whereas we in everyday mathematical conversation might assign various plausibilities ranging from 0 to 1 to such a proposition. Our uncertainty must therefore be of a different "type" than the uncertainty due to "partial deduction" as assumed here.

Colyvan [5] gives a thorough summary of these philosophical difficulties and argues convincingly against overstating the universal applicability of a Cox–Jaynes system of plausible reasoning to all situations of reasoning with uncertainty, in light of the restrictions described above.

In full generality, the logical information on the right-hand side of the plausibility line can be arbitrary. Our goal in discussing the universality condition will be to exhibit a special class of examples that result in plausibility assignments that are dense in the unit cube in an appropriate sense; it happens that for this class of examples it will suffice to have conditioning information expressed as a proposition (perhaps expressed as a conjunction of sub-propositions). We will indicate in what follows whether we have assumed the background information to be propositional.

1.2. Continuity

In a subtle way, the continuity assumption R5 is also employing the same trick as the above, in that it assumes the functional relationship that holds among plausibilities with respect to conjunction is extendable to all real numbers via a continuous function.[4] Why should this hold in general?

---

[3] Equivalently, for the examples described below, that it is satisfiable.

[4] The continuity of S follows from its monotonicity and the property $S(S(x)) = x$, but F must be assumed to be continuous from the start.

The difficulty lies in the fact that the natural domain of the plausibility function is the space of propositions, which does not come equipped with a topology. Therefore, all continuity assumptions for F must be mediated through the subspace topology which the space $P = \{((A|BX), (B|X))|A, B, X\}$ inherits from $[0, 1]^2$. For example, to state a continuity property that applies only on the relevant range of plausibilities, we might choose:

Given any A, B, X; for all $\epsilon > 0$ there exists $\delta > 0$ such that for all A', B', X' with $|(B'|X') - (B|X)| < \delta$ and $|(A'|B'X') - (A|BX)| < \delta$ we have $|(A'B'|X') - (AB|X)| < \epsilon$.

Indeed this seems consistent with what Van Horn has in mind, since in defending the intuition behind the continuity assumption for F, he writes [21, p. 16]:

> 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 AB to suddenly jump.

The assumption does sound convincingly intuitive; however, it alone does not guarantee that F is extendable to a continuous function on all of $[0, 1]^2$. It might be the case that F has an apparent discontinuity at a certain point $(x, y) \in [0, 1]^2$ but there are no propositions A, B, X with $(A|BX) = x, (B|X) = y$, and so we can't *realize* the discontinuity by choosing appropriate propositions to serve as counterexamples. For example, the function $F(x, y) = .5xy$ for $xy < .5$ and $F(x, y) = xy$ for $xy > .5$ has a discontinuous ridge along the curve $xy = .5$ but if all proposition triples A, B, X map to plausibility points in $\{xy < .5\} \cup \{xy > .5\}$ then the function F is *locally* continuous for all propositions in the sense of Van Horn's argument but not extendable to a continuous function in the sense required by his assumption. Worse still, this set of plausibilities is dense in $[0, 1]^2$, so this problem is not even resolved by the already controversial universality assumption.

As we will describe in Section 3, what's missing here is a *uniform* continuity assumption on the plausibility values, which in retrospect is unsurprising given that we desire that the function be extendable to a continuous – and therefore necessarily uniformly continuous – function on the compact space $[0, 1]^2$.

1.3. Strict monotonicity

As Van Horn remarks, the assumption that F is strictly increasing in each variable is also crucially important to Cox's proof. In its defense he argues [21, p. 16]:

> 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 AB must not become less plausible in this case. It accords with many people's intuition that AB must, in fact, be considered more plausible in this case, but there are others who disagree with this stronger requirement.

Even putting aside the possible disagreement, this too falls somewhat short of a complete defense. The premise of the hypothetical, that an update in information leaves the other "no less plausible," is not specific enough to tease out the actually problematic case. Suppose our initial state of information is denoted by X and the updated state of information X', and let $x = (A|BX), x' = (A|BX'), y = (B|X)$, and $y' = (B|X')$. Then this intuitive line of argument, in its "stronger" form, would support the claim that $x' > x$ and $y' \geq y$ implies $F(x', y') > F(x, y)$. But the specific sub-case we need to worry about, to block potential counterexamples such as $F(x, y) = min(x, y)$, is actually $x' > x$ and $y' = y$.

One could argue that intuition would suggest that the plausibility of the conjunction should increase in this case as well, but in order for this to be tested for given values x, x', y there must exist propositions A, B and conditioning information X, X' such that $x = (A|BX), x' = (A|BX')$, and $y = (B|X) = (B|X')$. What if this situation is never actually realized for any A, B, X, and X'? We may have a plausibility rule, such as the aforementioned $F(x, y) = min(x, y)$, that *appears* to satisfy the strictly increasing property but only *vacuously* so. And even despite the set of realizable plausibilities being dense in $[0, 1]^2$ it would not be possible to extend this function to one that is strictly increasing in both variables on the entire square.

Once again, *attempting to rephrase the functional assumptions only in terms of how they behave with respect to propositions is apparently insufficient, and yet it's only when reasoning about propositions that we can make any appeals to common sense*. This is the "intuition gap" that we believe has troubled many workers in this field and the one we will attempt to bridge in the present paper.

To begin with, we weaken the assumptions of R3 and R5 so that they only apply to propositions and are not strict:

Assumption 1.3 *(S is nonincreasing). For all propositions A, A' and states of information B, B' if $(A|B) \leq (A'|B')$ then $(\overline{A}|B) \geq (\overline{A'}|B')$.*

Assumption 1.4 *(F is nondecreasing). For all propositions A, A', B, B' and states of information C, C' if $(A|BC) \leq (A'|B'C')$ and $(B|C) \leq (B'|C')$ then $(AB|C) \leq (A'B'|C')$.*

With these basic assumptions in place, the plan for the rest of this paper is as follows: first, we explore and attempt to formalize a notion of "equivalence" that Jaynes uses in deriving his group invariance principle but that is not already made explicit in either Cox's or Van Horn's assumptions. We then use this to derive independence properties that hold among plausibility assignments for any problem of "symmetric ignorance," i.e., one consisting of mutually exclusive and exhaustive propositions that are indistinguishable in the Jaynesian sense and must therefore have equal plausibility assignments. To illustrate the point, we consider the example of drawing a ball from an urn given a set of symmetric assumptions concerning the balls and the urn. We postulate the existence of a similar problem for any given size and make appeals to common sense regarding their plausibilities.

Finally, we amend the continuity assumptions of the functions S and F so that they are uniform and only apply on the range of attainable plausibilities. By using the class of examples described above, we are able to show the existence of strictly monotonic and continuous extensions as well as Van Horn's universality condition. As a bonus, in the course of proving these conditions, we construct the function p rescaling plausibilities to be probabilities. In doing so, we provide a much more direct and elementary proof of Cox's theorem without the functional equation apparatus. Thus, we put these assumptions, and Cox's theorem, on a more solidly intuitive foundation.

As Shafer [18] remarks, the idea of building up probability rules by first considering mutually exclusive and equally likely cases is not a new one. It is certainly present in the work of Keynes [14] who coined the term "principle of indifference," and arguably originates with the classical idea of probability going back to Bernoulli [3], de Moivre [7], and Laplace [15], who began his *Théorie Analytique des Probabilités* with the definition:

> The Probability for an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.

One difference in the modern version is that we do not assume from the start that the probabilities for such a problem – which we call an "urn" for short – are equal to the corresponding ratios. Nor do we claim them from symmetries of the physical objects involved in the problem or observed frequencies of occurrence. Rather, we show through arguments concerning the symmetries of the *assumptions comprising our state of knowledge* that the plausibility assigned to such an event is expressible as a *function* of the ratio. After rescaling, we are free to take this function to be the identity, and so recover the classical definition merely as a convention for quoting probabilities.

Shafer describes this as also being the approach of Bernstein [4], who arrived at the same conclusion through a different set of axioms regarding equally likely cases. We claim only three modest innovations here:

1. We formalize an invariance principle with respect to translation that *necessitates* equal plausibility assignments for certain propositions given a symmetry in the background assumptions. This clarifies the notion of "indifference" used by Jaynes and forms the first example of his group invariance arguments for assigning prior probabilities, with the group in question being the symmetric group S_n, and avoids the circular definition of probability in terms of "likelihood." We go a step further by demonstrating how these problems also give an elementary class of *independent* propositions that have useful properties with respect to conjunction.
2. We add the word "uniform" to the continuity assumptions for the negation and conjunction plausibility functions and demonstrate the importance of this assumption in showing the density of urn-drawing plausibilities. In doing so, we prove Van Horn's universality assumption, likely necessary for Cox's result to hold true, and we rid ourselves of gratuitous assumptions on the continuity and strict monotonicity of the extension of F to the unit square. We therefore extend the classical idea of probability and that of Bernstein to problems that do not admit an obvious decomposition into some set of equally likely atomic sub-propositions but that can be approximated by such problems in the limit.
3. We explicitly construct the rescaling function that Cox's theorem promises and avoid having to solve the associativity equation. In the process, we give a very satisfying intuitive meaning to what the rescaling function accomplishes – that plausibilities for urn-drawing problems are expressible as a function of their ratios, and the rescaling function is simply the inverse – and thus harmonize Cox's point of view with that of Bernstein.

2. Principle of indifference

In reformulating Cox's desiderata for plausible reasoning, Jaynes [13, chapter 1, p. 19], elaborates slightly on the consistency requirement in his desideratum IIIc:

> The robot always represents equivalent states of knowledge by equivalent plausibility assignments. That is, if in two problems the robot's state of knowledge is the same (except perhaps for the labeling of the propositions), then it must assign the same plausibilities in both.

However, the desideratum seems to be stated so vaguely – perhaps on purpose, to obscure the controversy it will later cause – that one could hardly object. What is meant by "equivalent states of knowledge"?

One kind of equivalence we can imagine is that two sets of conditioning information may be cosmetically different yet remain logically equivalent. For example, if $X = BC$ and $X' = CB$ then we should require $(A|X) = (A|X')$ for any A. Arguably this is already covered by the requirement that $(\cdot|\cdot)$ be compatible with the propositional calculus; to be explicit, let us expand assumption R2 to include consistency with propositional logic on the right-hand side of the conditioning line:

Assumption 2.1. *If proposition X is logically equivalent to X' then for any proposition A, $(A|X) = (A|X')$.*

In the course of introducing his group invariance principle, however, Jaynes appeals to the "equivalence" in a different way, in the sense of the background assumptions of a problem being "indifferent" to two propositions [13, chapter 2, p. 39]:

> Now suppose that information B is indifferent between propositions A_1 and A_2; i.e., if it says something about one, it says the same thing about the other, and so it contains nothing that would give the robot any reason to prefer either one over the other.

And he concludes that the propositions must have equal plausibility in this case. Here we are required to reckon with what a proposition X "says about" proposition A_1 and whether it "says the same thing" about proposition A_2 – terms without exact definitions. This version of the requirement is in some ways the most important one for creating actual numerical examples, because it leads to what Jaynes describes as the "principle of indifference," following Keynes [14], which allows one to start computing probabilities for real propositions.

To make this more clear, let us consider an example like the one in which Jaynes first claims the indifference applies:

Example 2.2. Suppose we have the background information and propositions:

$X =$ "We are given an urn with two balls in it." AND "Ball #1 is in the urn." AND "Ball #2 is in the urn." AND "We draw a ball from the urn."
$A_1 =$ "We draw ball #1."
$A_2 =$ "We draw ball #2."

It seems that the assumptions X are indifferent to the propositions A_1 and A_2, since every statement included in X about the relevant objects in A_1 and A_2 (that is, "ball #1" and "ball # 2," respectively) also applies equally well to the other, and the propositions themselves assert the same things about those objects, just that "we draw" them. Can we make this "sameness" more precise?

Suppose in general that we have a problem involving two propositions A_1 and A_2 with background assumptions X, and we are attempting to assign the plausibilities $(A_1|X)$ and $(A_2|X)$. Suppose furthermore that the background information X can be built up, using the logical connectives, from the propositional symbols A_1 and A_2.

Consider then a different problem which is obtained by translating the propositional symbols into different symbols A_1' and A_2' by means of some bijection T so that $A_1' = T(A_1)$ and $A_2' = T(A_2)$. Let $X' = T(X)$ be the result of applying this translation to the symbols comprising X. Without making any more assumptions, we may assert that whatever X "has to say" about A_1 is also now exactly what X' "has to say" about A_1', since, effectively, all we have done is translate the statements embedded in X and A_1 into a different set of labels where the symbol A_i' means what we previously understood by A_i.

Let us codify this, then, as an additional assumption regarding plausibilities:

Assumption 2.3 *(Translation-invariance). If propositions X and A are expressible using the logical connectives on a background set of propositional symbols S, T is a bijection between this set and another set S', and $X' = T(X)$, $A' = T(A)$ are the results of applying the translations to X and A, respectively, then $(A'|X') = (A|X)$.*

However, suppose that the dictionary bijection is actually a permutation of the symbols S, with $T(A_1) = A_2$ and $T(A_2) = A_1$; that is, $A_1' = A_2$ and $A_2' = A_1$. If it is also the case that background assumption X' is logically equivalent to X, then we may assert that whatever X "has to say" about A_2 is also what X' "has to say" about A_2, since the X's are logically equivalent to each other. This will come about, for example, if X is logically symmetric with respect to the action of T, e.g., statements of the form $X = A_1 + A_2$; permuting the symbols merely reorders them, but the disjunctions are logically equivalent. It now follows in this circumstance that $(A_2|X) = (A_1'|X') = (A_1|X)$, which is the principle of indifference as Jaynes intends it.

Returning to the above example, if we let A_1 label the proposition "We draw ball #1." and A_2 label the proposition "We draw ball # 2.", then we note that we can equivalently write our background information X as

$$X = (A_1 + A_2)\overline{A_1 A_2}$$

That is, we draw exactly one ball. Letting T be the permutation $T(A_1) = A_2, T(A_2) = A_1$, we see that after applying T we have $X' = (A_2 + A_1)\overline{A_2 A_1}$ and so X and X' are logically equivalent. The conditions given above are satisfied, and it follows that $(A_2|X) = (A_1|X)$. Note that it does not immediately follow that $(A_1|X) = (A_2|X) = 1/2$; for that we will require the sum rule after rescaling plausibilities to be genuine probabilities.

We could have phrased our background assumptions asymmetrically, which would have preserved the translation property $(A_1|X) = (A'_1|X')$ but not the indifference; for example:

Example 2.4. $X =$ "We are given an urn with two balls in it." AND "Ball #1 is in the urn *on the bottom*." AND "Ball #2 is in the urn *on the top*." AND "We draw a ball from the urn." AND "When we draw a ball, we draw the ball on the top."

That is, letting A_1 and A_2 as before and D_T be the proposition "We draw the ball on top.", we can encode our background information as

$$X = (A_1 + A_2) \cdot \overline{A_1 A_2} \cdot (D_T \iff A_2) \cdot D_T$$

After doing the translation ($A_1 \leftrightarrow A_2$), we can see that whatever conclusions we now draw about ball #2 are the same ones we would have drawn about ball #1, but the background assumptions X and $X' = (A_2 + A_1)\overline{A_2 A_1}(D_T \iff A_1)D_T$ are no longer equivalent to each other, so we do not conclude that A_1 and A_2 have the same plausibility. Note that we cannot even force the logical symmetry by further translating "bottom" to mean "top" (i.e., translating between D_T and $\overline{D_T}$) because we have a symmetry-breaking assumption that would make the translation not logically equivalent.

Let us formalize this example in a general definition:

Definition 2.5. Call a set of propositions $X; A_1, \ldots, A_N$ an "N-urn" if the following hold:

1. Propositions $X, A_1, \ldots, A_N$ are built up, using the logical connectives, from a common set of background propositional symbols S.
2. A_i, A_j are mutually exclusive given X for all $i \neq j$.
3. $A_1, \ldots, A_n$ are exhaustive given X; that is, given X the proposition $A_1 + \ldots + A_N$ is certain.
4. For all i, j, A_i and A_j are indistinguishable given X in the sense given above; that is, there exists a permutation T of the symbols S such that $T(A_i) = A_j$, $T(A_j) = A_i$, $T(A_k) = A_k$ for all $k \neq i, j$, and such that $T(X)$ is logically equivalent to X.

It then follows that for all i, j, $(A_i|X) = (A_j|X)$.

Given any distinct indices $i(1), \ldots, i(n)$, after composing the permutations that interchange A_1 with $A_{i(1)}$, A_2 with $A_{i(2)}$, and so on, we have by the translation-invariance assumption that

$$(A_1 + \ldots + A_n|X) = (A_{i(1)} + \ldots + A_{i(n)}|X)$$

for any $1 \leq n \leq N$ as well. Therefore the disjunction of any two subsets of equal size within a given urn will have equal plausibilities.

In general there may not be an obvious translation from one N-urn to another, because the background assumptions of each may include additional information that is particular to that urn. For example, the following are both 2-urns:

$$X; A_1, A_2$$

$$X = (A_1 + A_2)\overline{A_1 A_2}$$

$$X'; B_1, B_2$$

$$X' = (A_1 + A_2 + A_3 + A_4)\overline{A_1 A_2} \cdot \overline{A_1 A_3} \cdot \overline{A_1 A_4} \cdot \overline{A_2 A_3} \cdot \overline{A_2 A_4} \cdot \overline{A_3 A_4}$$

$$B_1 = A_1 + A_2$$

$$B_2 = A_3 + A_4$$

(For the latter, consider the permutation with $A_1 \leftrightarrow A_3$ and $A_2 \leftrightarrow A_4$.) But no obvious bijection of symbols takes X to X'.

However, if $X; A_1, \ldots, A_N$ is an N-urn, then considering permutations that fix A_1 while interchanging A_i and A_j for $i, j > 1$ shows that $\overline{A_1}X; A_2, \ldots, A_N$ is an $(N-1)$-urn and similarly for other A_i. By applying this fact along with our assumptions on the functions S and F we can see the plausibility of similar corresponding propositions constructed out of the atomic propositions $A_1, \ldots, A_N$ and $A'_1, \ldots, A'_N$ for two different urns must also be equal:

Proposition 2.6 *(Urn equivalence). If $X; A_1, \ldots, A_N$ and $X'; A'_1, \ldots, A'_N$ are N-urns, then for any n, $1 \leq n \leq N$*

$$(A_1 + \ldots + A_n|X) = (A'_1 + \ldots + A'_n|X')$$

Proof. By induction on N. In the case $N = 1$ both A_1 and A'_1 are certain and so $(A_1|X) = (A'_1|X') = 1$.

For the case $N = 2$, $n = 2$ follows similarly to the above. For $n = 1$ note that if $X; A_1, A_2$ is a 2-urn then $(A_1|X) = (A_2|X)$ implies

$$(A_1|X) = (\overline{A_2}|X) = S((A_2|X)) = S((A_1|X))$$

and similarly

$$(A'_1|X') = S((A'_1|X'))$$

So both $(A_1|X)$ and $(A'_1|X')$ are fixed points of the function S. However, by Assumption 1.3 this fixed point is unique, since if any two plausibility values $x < y$ are fixed points then

$$x < y \Rightarrow S(x) \geq S(y) \Rightarrow x \geq y,$$

a contradiction. Therefore, $(A_1|X) = (A'_1|X')$.

Assume the result holds for all M-urns with $M \leq N$.

Let $1 \leq n \leq N + 1$ be given, and let $X; A_1, \ldots, A_{N+1}$ and $X'; A'_1, \ldots, A'_{N+1}$ be $(N + 1)$-urns.

If $n > (N + 1)/2$ note that

$$(A_1 + \ldots + A_n|X) = S((A_{n+1} + \ldots + A_{N+1}|X)) = S((A_1 + \ldots + A_m|X))$$

where $m = N + 1 - n < (N + 1)/2$, so it suffices to consider $n \leq (N + 1)/2$.

In the special case $n = (N + 1)/2$, let

$$B_1 = A_1 + \ldots + A_n$$
$$B_2 = A_{n+1} + \ldots + A_{2n}$$
$$B'_1 = A'_1 + \ldots + A'_n$$
$$B'_2 = A'_{n+1} + \ldots + A'_{2n}$$

After composing the permutations that interchange A_i with A_{n+i} for all i and those that interchange A'_i with A'_{n+i} we see that $X; B_1, B_2$ and $X'; B'_1, B'_2$ are 2-urns, so $(B_1|X) = (B'_1|X')$.

If $n < (N + 1)/2$ note that

$$\begin{aligned}(A_1 + \ldots + A_n|X) &= (\overline{A_{n+1} + \ldots + A_{N+1}}|X)\\ &= (\overline{A_{2n+1} + \ldots + A_{N+1}} \cdot \overline{A_{n+1} + \ldots + A_{2n}}|X)\\ &= F(S((A_{2n+1} + \ldots + A_{N+1}|\overline{A_{n+1} + \ldots + A_{2n}}X)), S((A_{n+1} + \ldots + A_{2n}|X)))\\ &= F(S((A_{2n+1} + \ldots + A_{N+1}|\overline{A_{n+1} + \ldots + A_{2n}}X)), S((A_1 + \ldots + A_n|X)))\end{aligned}$$

and similarly

$$(A'_1 + \ldots + A'_n|X') = F(S((A'_{2n+1} + \ldots + A'_{N+1}|\overline{A'_{n+1} + \ldots + A'_{2n}}X')), S((A'_1 + \ldots + A'_n|X')))$$

The first arguments inside the F functions on the right-hand sides are equal by the induction hypothesis, since both $\overline{A_{n+1} + \ldots + A_{2n}}X$ and $\overline{A'_{n+1} + \ldots + A'_{2n}}X'$ are $(N + 1 - n)$-urns. Call this common value p.

We have that both $(A_1 + \ldots + A_n|X)$ and $(A'_1 + \ldots + A'_n|X')$ are fixed points of the function

$$x \mapsto F(p, S(x))$$

However, by the monotonicity of S and F, this fixed point is again unique, since if $x < y$ are distinct fixed points then

$$x < y \Rightarrow S(x) \geq S(y) \Rightarrow F(p, S(x)) \geq F(p, S(y)) \Rightarrow x \geq y,$$

a contradiction. It follows that $(A_1 + \ldots + A_n|X) = (A'_1 + \ldots + A'_n|X')$. □

When reasoning about the plausibility of the propositions in an N-urn, we would expect strictly less restrictive statements to return strictly larger plausibility numbers. For example, considering the propositions $P = A_1 + \ldots + A_{n-1}$ and $Q = A_1 + \ldots + A_n$, we see that on background information X, the former proposition implies the latter and this implication is "strict," in the sense that truth value assignments exist for $A_1, \ldots, A_N$ that are consistent with X making P false while Q is true. In some sense the latter must therefore be "more plausible." For these simple examples, then, we assert a reasonable monotonicity of the plausibilities corresponding to groups of propositions.

Assumption 2.7. *For any N-urn, $X; A_1, \ldots, A_N$, we have $(A_1|X) > 0$ and $(A_1 + \ldots + A_n|X) > (A_1 + \ldots + A_{n-1}|X)$ for all n.*

By induction we also have $(A_1 + \ldots + A_n|X) > (A_1 + \ldots + A_m|X)$ for all $n > m$. As above, it also follows that

$$(A_{i(1)} + \ldots + A_{i(n)}|X) > (A_{j(1)} + \ldots + A_{j(m)}|X)$$

for $n > m$ and distinct indices $(i(1), \ldots, i(n))$, $(j(1), \ldots, j(m))$.

That is, the plausibility of the disjunction of a subset of the propositions $A_1, \ldots, A_N$ depends only on the size of the subset and is a strictly increasing function of that size.

Effectively, propositions making up an N-urn describe a process for choosing an integer i, $1 \le i \le N$, the index of the "result" proposition A_i, in such a way that the conditioning information is indifferent to the possible values of i, and they must therefore be equally plausible. This can be further considered as choosing *independently* a pair of integers (m, n) if we arrange the proposition set A_1, ..., $A_{M \cdot N}$ into an $M \times N$ grid and define propositions corresponding to the rows and columns:

Let $X; A_{1,1}, \ldots, A_{M,N}$ be an $(M \cdot N)$-urn. Let $C_1, \ldots, C_N$ be the "column" propositions

$$C_j = A_{1,j} + \ldots + A_{M,j}$$

and let R_1, ..., R_M be the "row" propositions

$$R_i = A_{i,1} + \ldots + A_{i,N}$$

Then each of the sets $(C_1, \ldots, C_N)$ and $(R_1, \ldots, R_M)$ are mutually exclusive and exhaustive given X. Composing the permutations that interchange A_{1,j_1} with A_{1,j_2}, A_{2,j_1} with A_{2,j_2}, and so on, we see that $X; C_1, \ldots, C_N$ is an N-urn, and similarly $X; R_1, \ldots, R_M$ is an M-urn.

The latter property remains unchanged even after conditioning on C_j for any j, $1 \le j \le N$, since, for example, applying the permutation T that interchanges rows R_1 with R_2 via:

$$T(A_{1,k}) = A_{2,k}$$
$$T(A_{2,k}) = A_{1,k}$$

transforms the column C_j to the logically equivalent $A_{2,j} + A_{1,j} + \ldots + A_{M,j}$, and similarly for any pair of rows. Therefore, $C_j X; R_1, \ldots, R_M$ is an M-urn, and by Proposition 2.6, $(R_i|C_j X) = (R_i|X)$.

Similarly, $(C_j|R_i X) = (C_j|X)$, and so the R_i and C_j propositions are independent in the usual sense.

Definition 2.8. Let us refer to this rearrangement $X; R_1, \ldots, R_M; C_1, \ldots, C_N$ as an "(M, N)-urn pair," but with the implicit understanding that all we have accomplished is an arrangement of propositions from the original $(M \cdot N)$-urn into a $M \times N$ grid with useful independence properties. Similarly, from an $(M \cdot N \cdot L)$-urn we may construct an $M \times N \times L$ array of dimension 3 (indexing the propositions by "row, column, plane" coordinates, for example) and refer to an "(M, N, L)-urn triple," etc. In particular, any N-urn is also an $(N, 1)$-urn pair and an $(N, 1, 1)$-urn triple and so on.

Similarly to the situation above for urn pairs, for an (M, N, L)-urn triple

$$X; R_1, \ldots, R_M; C_1, \ldots, C_N; P_1, \ldots, P_L$$

the R, C, and P propositions are jointly independent, in the sense that

$$(R_i|C_j P_k X) = (R_i|X)$$
$$(C_j|R_i P_k X) = (C_j|X)$$
$$(P_k|R_i C_j X) = (P_k|X)$$

Thus, from an elementary set of indistinguishable and equally plausible propositions we may also construct elementary sets of independent propositions, which remain indistinguishable even when conditioning on a proposition from the other set.

Several other properties of urns and urn pairs are then readily attainable from the already established consistency theorems:

Proposition 2.9. *For each pair of positive integers M, N assume $X_{M,N}$; R_1, ..., R_M; C_1, ..., C_N is an (M, N)-urn pair as above with underlying individual propositions $A_{1,1}, \ldots, A_{M,N}$. For (m, n) with $1 \le m \le M$, $1 \le n \le N$, let $R_{\le m}$ denote the proposition $R_{\le m} = R_1 + \ldots + R_m$ and $C_{\le n} = C_1 + \ldots + C_n$. For the special case of $N = 1$ we adopt the notation $X_M = X_{M,1}$, $A_i = A_{i,1}$ and $A_{\le m} = A_1 + \ldots + A_m$.*

Then we have

1. $(R_{\le m}|X_{M,N}) = (R_{\le m}|C_{\le n} X_{M,N})$. *That is, $R_{\le m}$ is conditionally independent from $C_{\le n}$ under $X_{M,N}$.*
2. $(R_{\le m}|X_{M,N}) = (R_{\le m}|X_{M,1}) = (A_{\le m}|X_M)$ *for all N.*
3. $(R_{\le m} C_{\le n}|X_{M,N}) = (A_{\le (m \cdot n)}|X_{M \cdot N})$.

4. $(A_{\leq m}|X_M) = (A_{\leq (c\cdot m)}|X_{c\cdot M})$ *for all* $c \in \mathbb{N}$.
5. $(A_{\leq m}|X_M) < (A_{\leq m'}|X_{M'}) \iff m/M < m'/M'$.
6. $(\overline{A_{\leq m}}|X_M) = (A_{\leq (M-m)}|X_M)$.
7. $(\overline{A_{\leq m}}|X_M) < (\overline{A_{\leq m'}}|X_{M'}) \iff m/M > m'/M'$.
8. $(R_{\leq m}C_{\leq n}|X_{M,N}) < (R_{\leq m}C_{\leq n'}|X_{M,N'}) \iff n/N < n'/N'$.
9. $(R_{\leq m}C_{\leq n}|X_{M,N}) < (R_{\leq m'}C_{\leq n}|X_{M',N}) \iff m/M < m'/M'$.

Proof.

1. Follows from the fact that $(C_{\leq n}X_{M,N})$; $R_1, \ldots, R_M$ is an M-urn.
2. Follows from the fact that both $X_{M,N}$; $R_1, \ldots, R_M$ and $X_{M,1}$; $R_1, \ldots, R_M$ are M-urns.
3. Follows since the (M, N)-urn pair $X_{M,N}$; $R_1, \ldots, R_M$; $C_1, \ldots, C_N$ is also an $(M \cdot N)$-urn with propositions $A_{1,1}, \ldots, A_{M,N}$, so both sides are the plausibilities of a disjunction of $m \cdot n$ propositions of an $(M \cdot N)$-urn.
4. Given a $(c \cdot M)$-urn $X_{c\cdot M}$; $A_1, \ldots, A_{c\cdot M}$, let $A'_1 = A_1 + \ldots + A_c$, $A'_2 = A_{c+1} + \ldots + A_{2c}, \ldots, A'_M = A_{(M-1)c+1} + \ldots + A_{Mc}$ and note that $X_{c\cdot M}$; $A'_1, \ldots, A'_M$ is an M-urn and $A'_{\leq m} \iff A_{\leq (c\cdot m)}$.
5. Let $x = (A_{\leq m}|X_M)$ and $x' = (A_{\leq m'}|X_{M'})$. By (4) we also have $x = (A_{\leq (m\cdot M')}|X_{M\cdot M'})$ and $x' = (A_{\leq (m'\cdot M)}|X_{M\cdot M'})$ for an $(M \cdot M')$-urn $X_{M\cdot M'}$; $A_1, \ldots, A_{M\cdot M'}$. It follows that $x < x' \iff m \cdot M' < m' \cdot M$.
6. Follows from the fact that $\overline{A_{\leq m}} \iff (A_{m+1} + \ldots + A_M)$, so both sides are the plausibilities of a disjunction of $M - m$ propositions in an M-urn.
7. Follows from the argument above and the fact that $(\overline{A_{\leq (m\cdot M')}}|X_{M\cdot M'}) = (A_{m\cdot M'+1} + \ldots + A_{M\cdot M'}|X_{M\cdot M'}) > (A_{m'\cdot M+1} + \ldots + A_{M\cdot M'}|X_{M\cdot M'}) \iff m \cdot M' < m' \cdot M$.
8. Follows since $(R_{\leq m}C_{\leq n}|X_{M,N}) = (A_{\leq (m\cdot n\cdot N')}|X_{M\cdot N\cdot N'})$ and $(R_{\leq m}C_{\leq n'}|X_{M,N'}) = (A_{\leq (m\cdot n'\cdot N)}|X_{M\cdot N\cdot N'})$ for an $(M \cdot N \cdot N')$-urn. So $(R_{\leq m}C_{\leq n}|X_{M,N}) < (R_{\leq m}C_{\leq n'}|X_{M,N'}) \iff m \cdot n \cdot N' < m \cdot n' \cdot N \iff n/N < n'/N'$.
9. Follows similarly to the above. □

This and the preceding discussion show that plausibilities for subsets of urn propositions are determined only by the size of the subset and the size of the urn, and in fact the key quantity is the ratio of the one to the other. For the sake of simplicity, we adopt the following notation:

Definition 2.10. For any rational number $0 < m/M \leq 1$, let $U(m/M) = (A_{\leq m}|X_M)$ for any M-urn X_M; $A_1, \ldots, A_M$. Let A_0 denote a contradiction and $U(0) = 0$.

Restating some of Proposition 2.9 in these terms gives:

Proposition 2.11.

1. $U(m/M) = U((c \cdot m)/(c \cdot M))$ *for any* $c \in \mathbb{N}$, *so* U *is well-defined.*
2. $U(m/M) = (R_{\leq m}|C_{\leq n}X_{M,N})$ *any* n, N.
3. $S(U(m/M)) = U(1 - m/M)$.
4. *For any* M, N*-urn pair* $X_{M,N}$; $R_1, \ldots, R_M$; $C_1, \ldots, C_N$, $(R_{\leq m}C_{\leq n}|X_{M,N}) = U((m \cdot n)/(M \cdot N))$. *Note that by the definition of* F *and the independence property for urn pairs,* $(R_{\leq m}C_{\leq n}|X_{M,N}) = F(U(m/M), U(n/N))$ *and so we have*

$$F(U(m/M), U(n/N)) = U((m \cdot n)/(M \cdot N))$$

5. U *is strictly increasing on* $\mathbb{Q} \cap (0, 1)$.
6. $(p, q) \mapsto F(U(p), U(q))$ *is strictly increasing in each variable on* $(\mathbb{Q} \times \mathbb{Q}) \cap (0, 1)^2$.

3. Cox's theorem

We are now ready to supply a set of intuitive assumptions that will have the universality condition along with the technical requirements on the function F, as well as Cox's theorem, as consequences.

As described in the last section, if a set of propositions form an N-urn, then many of the properties of their plausibility assignments are forced by the translation-invariance and consistency assumptions. But in order to make use of these examples we would need to know that such propositions are out there. If we specified the domain of reasoning for the robot, we could easily exhibit them; for example, if we were allowed to talk about balls and urns as before we could let:

X = "Ball #1 is in the urn." AND … AND "Ball #N is in the urn." AND "We draw a ball."
A_1 = "We draw ball #1.", …, A_N = "We draw ball #N."

Or if we were restricted to set-theoretic language, we could have

$$X = x \in \{x_1, \ldots, x_N\}$$

$$A_1 = \text{“}x = x_1.\text{”}, \ldots, A_N = \text{“}x = x_N.\text{”}$$

This is where the details of the robot's "programming language" begin to matter. As in the example in Section 2, all that is required for the existence of N-urns is a way of labeling atomic propositions with an unstructured set of labels corresponding to the positive integers and a way of constructing the background proposition

$$X = (A_1 + \ldots + A_N) \prod_{1 \le i < j \le N} \overline{A_i A_j}$$

This background conditioning information specifies *only* that exactly one of the constituent propositions is true, so that it is invariant under any permutation of labels and is therefore indifferent to the particular selection propositions. We assume that the individual plausibilities $(A_i|X)$ are known to the robot for any such problem as well as the rules S and F that allow plausibilities such as $(A_{\le i}|X)$ to be calculated, as described above. It does not seem much to ask for such a case of symmetric ignorance to be within the robot's realm of reasoning for any given size.

Let us postulate, therefore, that such systems exist within our assumed framework of plausibility assignments:

Assumption 3.1. *For any positive integer N, there exists an N-urn X; A_1, ..., A_N with assigned plausibilities $(A_1|X)$, ..., $(A_N|X)$.*

According to the assumption of translation-invariance and the argument of the previous section, we therefore have $(A_1|X) = \ldots = (A_N|X)$.

3.1. Continuity

We assume the functions S and F are defined on the range of attainable plausibility numbers, and we begin by modifying their continuity assumptions so that the continuity is uniform. To avoid the potential traps described in the introduction, we state this continuity condition only in terms of propositions:

Assumption 3.2 *(Uniform continuity for negation). For all $\epsilon > 0$ there exists $\delta > 0$ such that for all states of information A, A' and propositions B, B' with $|(B|A) - (B'|A')| < \delta$ we have $|(\overline{B}|A) - (\overline{B'}|A')| < \epsilon$.*

Assumption 3.3 *(Uniform continuity for conjunction). For all $\epsilon > 0$ there exists $\delta > 0$ such that for all propositions A, B, A', B' and states of information C, C' with $|(B|C) - (B'|C')| < \delta$ and $|(A|BC) - (A'|B'C')| < \delta$ we have $|(AB|C) - (A'B'|C')| < \epsilon$.*

By combining these assumptions we can also derive a useful continuity property for disjunction:

Lemma 3.4 *(Uniform continuity for disjunction). For all $\epsilon > 0$ there exists $\delta > 0$ such that for all propositions A, B and states of information C with $(B|\overline{A}C) < \delta$ we have $|(A + B|C) - (A|C)| < \epsilon$.*

Proof. Let $\epsilon > 0$. Let δ_1 satisfy Assumption 3.2 for ϵ. Let δ_2 satisfy Assumption 3.3 for $\epsilon = \delta_1$, and let δ satisfy Assumption 3.2 for $\epsilon = \delta_2$.

Let A, B, C with $(B|\overline{A}C) < \delta$. Letting $B' = \overline{C}$ we have $(B'|\overline{A}C) = 0$. Therefore $|(B|\overline{A}C) - (B'|\overline{A}C)| < \delta$ and so

$$|(\overline{B}|\overline{A}C) - (C|\overline{A}C)| = |(\overline{B}|\overline{A}C) - (\overline{B'}|\overline{A}C)| < \delta_2$$

Applying Assumption 3.3 with $(\overline{A}|C)$ as the second argument gives

$$|(\overline{A} \cdot \overline{B}|C) - (\overline{A}C|C)| < \delta_1$$

Since $(\overline{A}C|C) = (\overline{A}|C)$, applying Assumption 3.2 gives

$$|(A + B|C) - (A|C)| = |(\overline{\overline{A} \cdot \overline{B}}|C) - (\overline{\overline{A}}|C)| < \epsilon \qquad \square$$

Finally, we make one more commonsense continuity-like assertion for N-urns:

Assumption 3.5 *(Density of N-urns near 0). With the function U as in Proposition 2.11, we have $U(1/N) \to 0$ as $N \to \infty$. To be precise, for all $\epsilon > 0$ there exists N_0 such that if X; $A_1, \ldots, A_N$ is an N-urn with $N > N_0$ then $(A_1|X) < \epsilon$.*

That is, as a multitude of indistinguishable but mutually exclusive and exhaustive propositions increases, the plausibility of any given one of them gets arbitrarily close to zero. This seems to us to be completely consistent with a commonsense idea of what plausibility is, provided that plausibilities are represented by real numbers. Imagining for the moment that the space of propositions did have a kind of topology, it seems natural to assume that the propositions A_1 given N-urn assumptions X would be made to converge to a contradiction, i.e., an impossible proposition, as N went to infinity. Thus, asserting that their plausibilities converge to 0 amounts to a continuity property of the plausibility function.

This assumption will then guarantee that we have available propositions with which to exhibit other continuity properties, and the density of plausibility numbers in a neighborhood of 0 will imply density everywhere.

3.2. Universality

Theorem 3.6. *For all γ, $0 < \gamma < 1$ for all $\epsilon > 0$ there exists A, X such that $|(A|X) - \gamma| < \epsilon$.*

Proof. Given $\epsilon > 0$ let $\delta_1 < \epsilon$ be as in Assumption 3.2 for ϵ and let δ be as in Lemma 3.4 for $\epsilon = \delta_1$. Then let M be such that for any integer $N \geq M/2$, $U(1/N) < \delta$.

Let $X; A_1, \dots, A_M$ be an M-urn. Let A_0 be a contradiction.

Considering indices i with $0 \leq i \leq M/2$ we have $\overline{A_{\leq i}}X;\ A_{i+1}, \dots, A_M$ is an $(M-i)$-urn with $M - i \geq M/2$ and so by assumption $(A_{i+1}|\overline{A_{\leq i}}X) < \delta$. Since $A_{\leq(i+1)} \iff A_{\leq i} + A_{i+1}$, from Lemma 3.4 we have $|(A_{\leq(i+1)}|X) - (A_{\leq i}|X)| < \delta_1 < \epsilon$.

For those indices i with $M/2 \leq i < M$, note by Proposition 2.6 we have $(\overline{A_{\leq i}}|X) = (A_{\leq(M-i)}|X)$ and similarly $(\overline{A_{\leq(i+1)}}|X) = (A_{\leq(M-i-1)}|X)$. Therefore since $M - i \leq M/2$ it follows from the preceding argument that $|(\overline{A_{\leq i}}|X) - (\overline{A_{\leq(i+1)}}|X)| < \delta_1$. Applying Assumption 3.2 gives $|(A_{\leq(i+1)}|X) - (A_{\leq i}|X)| < \epsilon$.

Combining these observations, we have that the numbers $U(i/M) := (A_{\leq i}|X)$ form an increasing sequence with $U(0) = 0$, $U(1) = 1$, and $U((i+1)/M) - U(i/M) < \epsilon$ for all i. □

From there, the rest of Van Horn's universality assumption $R4$ follows as well:

Theorem 3.7.

1. *Let $P_0^1 = \{U(m/M)|m, M\}$. Then P_0^1 is dense in* $[0, 1]$.
2. *With $P_0^2 := P_0^1 \times P_0^1$ and $P_0^3 := P_0^1 \times P_0^1 \times P_0^1$, we have that P_0^2 is dense in* $[0, 1]^2$ *and P_0^3 is dense in* $[0, 1]^3$.
3. *With $P^1 = \{(A|X)|A, X\}$, $P^2 = \{((A|BX), (B|X))|A, B, X\}$, and $P^3 = \{((A|BCX), (B|CX), (C|X))|A, B, C, X\}$, we have that P^1 is dense in* $[0, 1]$, *P^2 is dense in* $[0, 1]^2$, *and P^3 is dense in* $[0, 1]^3$.

Proof. (1) is just a restatement of the proof of Theorem 3.6 and (2) follows immediately. From the independence property of urn pairs, we also know that any pair $(U(m/M), U(n/N)) = ((R_{\leq m}|C_{\leq n}X_{M,N}), (C_{\leq n}|X_{M,N}))$ and similarly for urn triples, and so (3) follows from the fact that $P_0^i \subseteq P^i$ for $i = 1, 2, 3$. □

3.3. Strict monotonicity

Thus far, we have not made strict monotonicity assumptions on S or F apart from noting how they behave on the subset of plausibilities attainable by urns. From Proposition 2.11 we have strict monotonicity on the space of urn plausibilities, since as we have shown:

$$S(U(p)) = U(1-p)$$

$$F(U(p), U(q)) = U(p \cdot q)$$

And the function U is strictly monotonic.

The existence and uniqueness of a continuous and strictly monotonic extension of S from the dense subset P^1 to $[0, 1]$ follows readily. The proof is elementary, but we include it to demonstrate the importance of the uniform continuity assumption:

Theorem 3.8. *There exists a function S_1 on* $[0, 1]$ *such that*

1. *S_1 is continuous.*
2. *$S_1(x) = S(x)$ for all $x \in P^1$.*
3. *S_1 is strictly decreasing.*

Proof. Let $x \in [0, 1]$ and let (x_n) in P^1 be a sequence with $x_n \to x$. Since S is uniformly continuous on P^1, $(S(x_n))$ is Cauchy, hence convergent. Define $S_1(x) = \lim_{n\to\infty} S(x_n)$.

To see that this is well-defined, suppose there exist two sequences (x_n) and (y_n) in P^1 converging to x with $\lim_{n\to\infty} S(x_n) = s$ and $\lim_{n\to\infty} S(y_n) = t$. Let $\epsilon > 0$ and find N_1 such that $n > N_1$ implies $|S(x_n) - s| < \epsilon$ and $|S(y_n) - t| < \epsilon$. From the assumption of uniform continuity, find $\delta > 0$ such that for any $a, b \in P^1$ with $|a - b| < \delta$ we have $|S(a) - S(b)| < \epsilon$.

Find N_2 such that $n > N_2$ implies $|x_n - x| < \delta/2$ and $|y_n - x| < \delta/2$. Then, taking $n > max(N_1, N_2)$ gives $|x_n - y_n| < \delta$, hence $|S(x_n) - S(y_n)| < \epsilon$ and so $|s - t| \leq |S(x_n) - s| + |S(y_n) - t| + |S(x_n) - S(y_n)| = 3\epsilon$. Since this is true for all ϵ, we have $s = t$.

A similar argument shows that S_1 is continuous and nonincreasing. To see that S_1 is strictly decreasing, note that for any $0 \leq x < x' \leq 1$ we can find p, p' in P_0^1 with $x < p < p' < x'$. Since S is strictly decreasing on P_0^1 (and agrees with S_1 there) we then have $S_1(x) \geq S_1(p) > S_1(p') \geq S_1(x')$. □

Because S operates on a subset of a one-dimensional space, strict monotonicity of the continuous extension comes easily. For the function F some additional care must be taken, since it can happen that a function of two variables is uniformly continuous and strictly increasing in each variable on a dense subset of $[0, 1]^2$ yet not extendable to a strictly increasing function on all of $[0, 1]^2$. For example, letting the set

$$R = T(\mathbb{Q} \times \mathbb{Q}) \cap [0, 1]^2$$

where T is the orthogonal rotation about the origin by 30° (or any other angle taking the x-axis to a line with irrational slope) gives that R is dense in $[0, 1]^2$, but no two distinct points in R lie on the same horizontal or vertical line. Therefore, the function $F(x, y) = min(x, y)$ is strictly increasing in each variable on R, but only vacuously so because there don't exist any two points in R with a common x or y value.

In our case, however, the dense subset P_0^2 is better behaved, and we are able to use our ability to calculate values of F on P_0^2 to derive strict monotonicity of the continuous extension.

Lemma 3.9. *The function U of Proposition 2.11 is uniformly continuous in the subspace topology and extendable to a continuous function U_1 on $[0, 1]$, which is strictly increasing with $U_1(0) = 0$ and $U_1(1) = 1$.*

Proof. To see that U is uniformly continuous, let $\epsilon > 0$ and find M as in the proof of Theorem 3.6 such that $U((i+1)/M) - U(i/M) < \epsilon/2$ for all i, $0 \leq i < M$. Given any $x < y$ with $y - x < 1/M$ there exists $c \in \mathbb{N}$ such that $c/M < x < y < (c+2)/M$, and so $U((c+2)/M) < U(c/M) + \epsilon/2 + \epsilon/2$.

Therefore $0 \leq U(y) - U(x) \leq \epsilon$.

Since U is uniformly continuous and strictly increasing on the dense subset $\mathbb{Q} \cap [0, 1]$, we can extend it to a continuous U_1 on $[0, 1]$ by taking limits as in Theorem 3.8. It follows that U_1 is nondecreasing on $[0, 1]$, and, because $U_1(x) \leq U(p) < U(q) \leq U_1(y)$ for any $x < p < q < y$ with $p, q \in \mathbb{Q}$, U must be strictly increasing. □

Theorem 3.10. *The function F defined on P^2 is extendable to a continuous function F_1 on $[0, 1]^2$. The function F_1 is nondecreasing on $[0, 1]^2$ and strictly increasing on $(0, 1]^2$.*

Proof. F is uniformly continuous on P^2 by Assumption 3.3 and strictly increasing on the subset P_0^2 by Proposition 2.11, which is dense in $[0, 1]^2$ by Theorem 3.7, so the extension to $[0, 1]^2$ is uniquely defined and nondecreasing, as in Lemma 3.9. Let $0 < x, y, y' \leq 1$ be given with $y < y'$. It suffices to show that $F_1(x, y) < F_1(x, y')$.

Since the function U_1 is strictly increasing from $[0, 1]$ onto $[0, 1]$ we have $U_1^{-1}(y) < U_1^{-1}(y')$. Fix n, N such that $U_1^{-1}(y) < n/N < (n+1)/N < U_1^{-1}(y')$. Take M large enough such that $m' := \lfloor U_1^{-1}(x) \cdot M \rfloor > n$. Then with $m = m' + 1$ and $n' = n + 1$ we have $m'/M \leq U_1^{-1}(x) < m/M$ and so $U_1(m'/M) \leq x < U_1(m/M)$ and $y < U_1(n/N) < U_1(n'/N) < y'$.

Meanwhile, $m/m' = 1 + 1/m' < 1 + 1/n = n'/n$ implies $m' \cdot n' > m \cdot n$ and therefore

$$U_1\left(\frac{m' \cdot n'}{M \cdot N}\right) > U_1\left(\frac{m \cdot n}{M \cdot N}\right)$$

As we observed in Proposition 2.11,

$$F(U(m/M), U(n/N)) = U((m/M) \cdot (n/N))$$
$$F(U(m'/M), U(n'/N)) = U((m'/M) \cdot (n'/N))$$

Therefore, $F_1(x, y) \leq F(U(m/M), U(n/N)) < F(U(m'/M), U(n'/N)) \leq F_1(x, y')$.

A similar argument shows F is increasing in its first variable. □

Now that we have shown that the assumptions of Van Horn's presentation of Cox's theorem are satisfied, we note that we don't actually need the remainder of the proof, because we have already explicitly constructed the function $p := U_1^{-1}$ implementing the isomorphism between our plausibility rule $(\cdot|\cdot)$ and a probability:

Theorem 3.11 *(Cox's theorem). If $(\cdot|\cdot)$ is a plausibility rule with range in $[0, 1]$ satisfying Assumptions R1 and R2 as well as Assumptions 2.1 and 2.3, S and F are functions satisfying Assumptions 1.2, 1.3, 1.4, 3.2, and 3.3, and the domain of reasoning allows for arbitrarily large N-urns as in Assumption 3.1 satisfying Assumptions 2.7 and 3.5, then $(\cdot|\cdot)$ is rescalable to ordinary probability in the sense that there exists a strictly increasing function $p : [0, 1] \to [0, 1]$ such that:*

1. $p(A|X) = 1 - p(A|X)$ *for all propositions* A *and states of information* X.
2. $p(AB|X) = p(A|BX)p(B|X)$ *for all propositions* A, B *and states of information* X.

Proof. Let U_1 be the continuous extension of the function U defined on rational numbers as in Proposition 2.11 and Lemma 3.9, and let $p = U_1^{-1}$. Then

1. $p((\overline{A}|X)) = p(S(A|X)) = U_1^{-1}(U_1(1-q)) = 1 - q = 1 - p(A|X)$ whenever $(A|X) = U(q)$ for some $q \in \mathbb{Q}$.
2. $p(AB|X) = p(F(A|BX, B|X)) = U_1^{-1}(U_1(q \cdot q')) = q \cdot q' = p(A|BX) \cdot p(B|X)$ whenever $(A|BX) = U(q)$ and $(B|X) = U(q')$.

By Theorem 3.7 these plausibilities are dense in $[0, 1]$ and $[0, 1]^2$, respectively, so the conclusions follow for all A, B, X. $\square$

This also provides a sense of intuition about the "scaling" function p, which brings the Cox–Jaynes system more in line with other treatments of probability. Assuming, in the Cox formulation, a rule of associating plausibilities to propositions exists and the value $(A|X)$ is known, we imagine constructing an "urn-drawing" type problem of choosing a ball of a particular subset from the urn, which, for some given ratio q of the size of the subset to that of the urn, has a (nearly) equivalent plausibility. By the density of urn-drawing probabilities, we know this is possible to within any tolerance. The corresponding ratio q is then roughly what we mean when say the "probability."

Assuming two such plausibilities $(A|BX)$ and $(B|X)$ are given, we imagine constructing an "urn pair" type problem with (nearly) equivalent plausibilities, i.e., we find ratios q and q' such that the plausibility of drawing one of the fraction of q balls is within a given tolerance of $(A|BX)$ and, in an independent draw, one of the fraction of q' balls is close to $(B|X)$. We equivalently reason that these results can be organized into groups so that the urn pair can be thought of as a single urn-draw. According to the symmetry properties of this kind of problem, we know that the joint event of drawing one of the q proportion and one of the q' proportion now has plausibility that is the same as the plausibility of drawing one of $q \cdot q'$. In other words, probabilities obey the product rule.

In doing so, we have merely exploited the fact that propositions A, B with probabilities $p(A|BX)$ and $p(B|X)$ can, in some sense, be translated into independent propositions C, D in a special class with background information X', $p(C|DX') = p(C|X') = p(A|BX)$ and $p(D|X') = p(B|X)$. Then, for these independent propositions, we have shown explicitly that $p(CD|X') = p(C|X')p(D|X')$. Because of our general assumption that $p(AB|X)$ is determined uniquely as some function of $p(A|BX)$ and $p(B|X)$ we then have $p(AB|X) = p(CD|X') = p(C|X')p(D|X') = p(A|BX)p(B|X)$. The idea of changing the conditional dependence of A given BX into an independent proposition with that given probability is, in our view, the real insight of Cox's theorem.

4. Conclusion

We have reinforced the intuitive justifications for the Cox–Jaynes model of probability as a necessary consequence of elementary desiderata for plausible reasoning, by providing an intuitive set of assumptions that imply the more troubling ad hoc conditions of universality, continuity, and strict monotonicity as consequences. To do so, we have formalized an assumption already present in Jaynes's reasoning concerning "equivalent states of knowledge" to mean that (1) translating proposition labels in a problem does not change plausibility assignments, (2) as a result, situations of symmetric ignorance concerning indistinguishable propositions must lead to equal plausibilities, and (3) that those numerical values are constant from problem to problem. By then postulating the existence of such problems of any given size, such as drawing a ball from an urn or selecting one particular element of a set given no other assumptions, as well as monotonic behavior of the plausibilities of such propositions, we are able to exhibit a dense subset of the unit cube that is attainable using plausibilities of real propositions.

The most controversial of these assumptions is likely that as the size of the problem (number of available outcomes) gets larger, the plausibility of any particular indistinguishable statement can be made arbitrarily small, but we see this as an intuitive requirement for any system of reasoning with plausibility represented as a real number. In some sense, the proposition of selecting a particular element of very large set – say finding a particular grain of sand on a beach or a particular atom in the solar system – is "close to" an impossible proposition, and therefore if plausibility is continuous, as a sequence of very unlikely propositions converge to the impossible, the plausibility should converge to that of a contradiction, i.e., the value 0.

As a trade-off, we are able to dispense with artificial conditions about the continuity and strict monotonicity of extensions of the negation and conjunction plausibility functions to the unit interval or unit square. Instead, we have rephrased these as uniform continuity assumptions *on the range of attainable plausibilities where they apply* and monotonicity assumptions on the same range that are not strict. Therefore, we are able to support these assumptions with commonsense reasoning about the behavior of plausibilities assigned to propositions, rather than starting with technical conditions on the functions involved as Van Horn [21] does. By making use of our urn-like examples, though, we are able to show that the continuous extensions do exist and that they are strictly monotonic. So we return to the same place that Van Horn's assumptions started, but we arrive there as a consequence of reasoning about a chosen class of elementary examples, rather than postulating them by fiat. Along the way, we have explicitly constructed the rescaling function that turns plausibilities into genuine probabilities and so are able to take a shortcut to the end of Cox's theorem without having to deal with solving functional equations. We are left with a much more intuitive interpretation of the rescaling function and Cox's theorem as a whole.

References

[1] J. Aczél, Lectures on Functional Equations and Their Applications, vol. 1, Academic Press, New York, 1966, edited by Oser, Hansjorg.
[2] S. Arnborg, G. Sjödin, On the foundations of Bayesianism, in: AIP Conference Proceedings, Institute of Physics Publishing LTD, 2001, pp. 61–71.
[3] J. Bernoulli, Ars Conjectandi, Impensis Thurnisiorum, fratrum, 1713.
[4] S.N. Bernstein, Attempt at an axiomatic foundation of probability theory, Mem. Math. Soc. Charkov 15 (1917) 209–274 (in Russian), list of publ. number 6 item 3.
[5] M. Colyvan, The philosophical significance of cox's theorem, Int. J. Approx. Reason. 37 (1) (2004) 71–85.
[6] R.T. Cox, Probability, frequency and reasonable expectation, Am. J. Phys. 14 (1) (1946) 1–13.
[7] A. de Moivre, The Doctrine of Chances, H. Woodfall, 1738.
[8] D. Dubois, H. Prade, The logical view of conditioning and its application to possibility and evidence theories, Int. J. Approx. Reason. 4 (1) (1990) 23–46.
[9] M.J. Dupré, F.J. Tipler, New axioms for rigorous Bayesian probability, Bayesian Anal. 4 (3) (2009) 599–606.
[10] J.Y. Halpern, A counterexample to theorems of Cox and Fine, J. Artif. Intell. Res. 10 (1999) 67–85.
[11] J.Y. Halpern, Technical addendum: Cox's theorem revisited, J. Artif. Intell. Res. 11 (1999) 429–435.
[12] M. Hardy, Scaled Boolean algebras, Adv. Appl. Math. 29 (2) (2002) 243–292.
[13] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[14] J.M. Keynes, A Treatise on Probability, Cambridge University Press, 1921.
[15] P.S. Laplace, Théorie analytique des probabilités, V. Courcier, 1820.
[16] J.B. Paris, The Uncertain Reasoner's Companion: A Mathematical Perspective, Cambridge University Press, 1994.
[17] G. Pólya, Mathematics and Plausible Reasoning: vol 2: Patterns of Plausible Inference, Oxford University Press, 1954.
[18] G. Shafer, Comments on "constructing a logic of plausible inference: a guide to Cox's theorem", by Kevin S. Van Horn, Int. J. Approx. Reason. 35 (1) (2004) 97–105.
[19] P. Snow, On the correctness and reasonableness of Cox's theorem for finite domains, Comput. Intell. 14 (3) (1998) 452–459.
[20] M. Tribus, Rational Descriptions, Decisions and Designs, Pergamon Unified Engineering Series, Elsevier, 1969.
[21] K.S. Van Horn, Constructing a logic of plausible inference: a guide to Cox's theorem, Int. J. Approx. Reason. 34 (1) (2003) 3–24.
[22] K.S. Van Horn, Response to Shafer's comments, Int. J. Approx. Reason. 35 (1) (2004) 107–110.