Computational Intelligence, Volume 14, Number 3, 1998

ON THE CORRECTNESS AND REASONABLENESS OF COX'S THEOREM FOR FINITE DOMAINS

PAUL SNOW

Concord, New Hampshire

Halpern has recently claimed a counterexample to Cox's Theorem, a well-known existence result for subjective probability distributions, but stated that the counterexample can be defeated by a specific assumption. Cox made this assumption, and so escapes the counterexample. Although Halpern questioned whether the assumption is reasonable for finite sets of sentences, it supports features that distinguish Cox's work from other, more restrictive motivations of probabilism. Paris has recently offered a new proof of Cox's Theorem whose correctness is satisfactory to Halpern, one that depends on a premise consistent with Cox's later work. As with any deductive argument, denial of a premise licenses denial of the conclusion, but Cox's conclusion does follow from premises plainly acceptable to him.

Key words: Cox's Theorem, subjective probability.

1. INTRODUCTION

The physicist Richard Threlkeld Cox claimed that any scheme for managing cardinal degrees of belief which obeyed certain constraints could be modeled using the ordinary calculus of probabilities. This theorem appeared in three of his works: an article in a physics journal (1946), a book (1961), and in a lengthy conference paper (1978). The theorem has achieved influence in the artificial intelligence community through the scholarship of researchers such as Horvitz, Heckerman, and Langlotz (1986).

In 1994, Paris proved a version of the theorem, but reported that an additional assumption beyond what he found in Cox (1946) appeared to be needed. In 1996, Halpern presented a paper at the annual AAAI Conference, claiming to have found a counterexample to Cox's Theorem when applied to finite sets of sentences. Halpern (1996) identifies a particular assumption that would defeat his counterexample, and goes on to question whether that assumption would be reasonable for finite domains. He raises two specific issues: whether infinitely many possible gradations of belief can plausibly arise with only finitely many sentences, and whether a believer ought to recognize and regulate degrees of belief which do not correspond to any sentence of interest.

A brief overview of Cox's Theorem appears in the next section. That Cox did in fact make the assumption required of him by Halpern is shown in the third section. Both questions bearing on the reasonableness of this assumption are then discussed. Finally, Paris' additional assumption is examined and found to be within Cox's intention.

2. COX'S THEOREM

Cox's Theorem concerns the existence of subjective probability distributions and derives the rules that govern them from intuitive properties of credibility without any reference to relative frequencies in an ensemble.

Suppose there were some arbitrary real-valued measure of belief in a proposition conditioned on the truth of a second (not necessarily distinct) proposition. Following Cox, let that measure be called a *likelihood* and be denoted

$$b \mid a$$

Address correspondence to the author at P.O. Box 6134, Concord, NH 03303-6134; e-mail: paulsnow@delphi.com.

which may be read "the degree of belief in sentence b if a were known to be true." Then under assumptions to be discussed, there would also exist some measure of belief, denoted here by $h(b \mid a)$, that satisfies

$$
\begin{gathered}
h(\mathbf{T} \mid a) = 1 \\
h(\mathbf{F} \mid a) = 0 \\
h(\mathbf{T} \mid a) \geq h(b \mid a) \geq h(\mathbf{F} \mid a) \\
h(b \vee c \mid a) = h(b \mid a) + h(c \mid a) - h(b \wedge c \mid a) \\
h(b \wedge c \mid a) = h(b \mid c \wedge a) \times h(c \mid a),
\end{gathered}
$$

where $\mathbf{T}$ is any tautology, $\mathbf{F}$ is any contradiction, b is any sentence, a and c are any sentences where $a \wedge c$ is not believed to be certainly false (so neither is a certainly false), and $h(\ )$ is some order-preserving function over the real numbers. The relationships involving $h(\ )$ are readily recognized as restatements of Kolmogorov's (1933) axioms for a probability measure and his definition of conditional probability.

Cox's principal assumptions were that

$c \wedge b \mid a$ is determined by a function of $c \mid b \wedge a$ and of $b \mid a$
$\neg b \mid a$ is determined by a function of $b \mid a$.

As Cox developed his argument, additional conditions were appended to these as the occasion arose. Sometimes the additional conditions were motivated by an appeal to Boolean algebra, and other times they were just asserted.

Along the way, it emerged that the belief-combination function described in the first principal assumption,

$$c \wedge b \mid a = F(c \mid b \wedge a, b \mid a)$$

must be associative. That is,

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

where x, y, and z are real numbers.

3. HALPERN'S COUNTEREXAMPLE

Halpern (1996) constructed a finite set of numbers and showed that for these values there is no associative F function. The method of construction was to take selected values from a finite probability distribution and to create a new belief function by changing some of those values. The new set of values still satisfied Halpern's version of Cox's assumptions. The conclusion of the theorem was then shown to require an associative combination function contrary to fact.

In discussion (at page 1319), Halpern states

> The counterexample given here could be circumvented by requiring that F be associative on all triples (rather than just on the triples (x, y, z) that arise [from specific conditional propositions used in the proof]).

As it happens, immediately after Cox derives the associativity expression (at page 6, Eq. (8) of the 1946 paper), he writes

> The function F must be such as to satisfy Eq. (8) for arbitrary values of x, y, and z.

Thus, Cox did make the required assumption, circumventing the counterexample.

Although not at issue in the counterexample, the domain pertinent to Cox's function which relates complementary propositions, $S(\ )$, must also subtend the entire real interval where belief values are found, not just the specific values corresponding to explicitly evaluated conditional propositions. Halpern points this out, and so it is in Cox. The argument x in Cox's functional equation

$$S[S(x)] = x,$$

where $S(\ )$ relates complementary propositions, "may have any possible value of a likelihood between those of certainty and impossibility" (1946, p. 7).

The sentence in question is poorly phrased. It reads "Thus S must be such a function that $S[S(x)] = x$, where x may have any possible value [in the indicated range]." The scope of the first word, *thus*, is not transparent. In fact, only the phrase that appears before the comma is motivated from the text that precedes it. The statement about the values of x is a new subject, asserted without any supporting discussion. This phrasing may cause confusion, suggesting that the now crucial words state a conclusion rather than an assumption. There simply is no argument, however, for which that statement could be the conclusion.

Cox has conformed to the requirements of deductive soundness demanded by Halpern. This does not quiet the matter entirely, however, since the possibility remains that the domains of belief functions *ought* to be restricted if the result is to be relevant to finite sets of propositions.

As Halpern states the objection (immediately after the passage quoted earlier, p. 1319),

> However, if we really are interested in a single domain, the motivation for making requirements on the behavior of F on belief values that do not arise is not so clear.

Halpern also notes elsewhere on the same page that Cox's Theorem "disallows a notion of belief that takes on only finitely many or even countably many gradations."

4. INFINITE GRADATIONS WITHIN FINITE DOMAINS

It often happens that sentences of interest include some that describe events for which there is an "objective" probability. That is, the believer is convinced there is some physically grounded probability model that permits the calculation of a reliable numerical relative frequency for the event at issue. In such a case, many statisticians subscribe to the view that this objective probability of an event is a rational measure of the belief in a sentence which asserts that the event will occur (Edwards 1969).

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. Exquisite precision per se does not prevent manipulation of a value within a finite symbol system.

For example, consider a right triangle each of whose legs is one unit long. Suppose a point is to be chosen at random from that triangle according to a uniform distribution. Let the question of interest be whether the chosen point lies along the hypotenuse, or one leg, or the other leg; a finite universe of discourse. The probability that the point lies on the hypotenuse is

$$\sqrt{2}/(2+\sqrt{2}),$$

a number that is different from any integral ratio, and yet that is exactly represented by the finite-length symbolic expression just given.

Any requirement that degrees of belief must be rational in the arithmetic sense amounts to a prohibition against adopting this quantity as the degree of belief for the specified event. And yet the expression raises no difficulties of any kind for a formal manipulator of symbols of finite capacity. Any requirement that there be only a finite number of possible degrees of belief would restrict which right triangles one could entertain beliefs about in similar problems with different dimensions.

5. UNDERDETERMINED BELIEF CONSTRAINT SYSTEMS

Situations of the kind discussed in the preceding section, featuring sources of belief external to the believer, are natural enough to a physicist. These are also the situations where frequencies in an ensemble are usually discussed. Cox might have been content, then, with presenting a theory of truly subjective estimation of the objective probabilities found in those problems. It is clear from his 1978 paper, however, that Cox was *not* content with such a limited view of subjective probability. On the contrary, he wished to consider "What song the Sirens sang, or what name Achilles assumed when he hid himself among women," subject matter where useful objective frequencies are few.

In the sixth chapter of his 1961 book, Cox acknowledges that except in gambling or statistical problems, precise cardinal degrees of belief may be unattainable. For other beliefs that stem mainly from the intuitions of the believer, such as the predicted outcomes of one-time events, Cox states that only "approximations or judgments of more or less" may be possible. Cox goes on to argue that qualitative probabilistic reasoning can and should be applied to such matters nonetheless.

That advice presents a logical difficulty. Cox's Theorem as originally presented relies heavily on the cardinal character of the degrees of belief which are supposed to exist. Merely qualitative beliefs simply will not support Cox's method of proof. What, then, is his warrant for introducing qualitative probabilities?

Cox models qualitative beliefs as systems of simultaneous probabilistic inequalities. The particular example he discusses involves personal beliefs about a political election with several candidates. Natural language expressions of belief are translated into algebraic relationships. For instance, one candidate is more likely than another to be nominated for the office on background information h,

$$\text{nominated}(c) \mid h < \text{nominated}(c') \mid h,$$

and, if nominated, candidate c has at best an even chance of winning,

$$\text{elected}(c) \mid \text{nominated}(c) \wedge h \leq \neg\text{elected}(c) \mid \text{nominated}(c) \wedge h.$$

The resulting system yields infinitely many solutions, as consistent constraint systems often do in practice. Each solution is a tuple of *cardinal* values—that is, individual probability values—and so comports with Cox's Theorem. The *ordinal* effect is achieved by the agreement among the solutions about the ordering of the candidates' prospects, and even about numerical probability bounds on those prospects.

To model a finite set of beliefs in this way demands the good behavior of uncountably many possible, but not necessarily experienced, degrees of belief. Within the solution set, alternative degrees of belief in the same proposition are as dense as the real numbers. Nevertheless, an exact description of the set can be stated using a finite number of symbols.

A distinguishing feature of Cox's approach to probability is that it supports ordinal models (Snow 1995). Alternative foundations for subjective probability based on gambling arguments were already available in Cox's time (for example, De Finetti 1937). Even apart from the difficulties of settling bets about the Sirens' choral repertoire, Cox's Theorem readily generalizes to the set-based Bayesianism advocated by other thinkers influential in the artificial intelligence community (Kyburg & Pittarelli 1996). A ban on infinitely many and arbitrarily dense potential degrees of belief would prohibit any belief model that depended on general convex (or other "solid") sets of probability distributions.

Some people might find Cox's generality undesirable, preferring instead a motivation that coincided exactly with an orthodox Bayesianism where only one probability distribution represented belief. Even within the orthodox Bayesian community, however, many scholars advocate arriving at that one probability distribution by means of entropy maximization. In this technique, beliefs are first described by an algebraic constraint system in which probability values appear as variables. The solution of that system which has the greatest entropy value is then selected. The fruit of this labor is a single specific probability distribution, but one whose derivation requires degrees of belief to be treated as variable quantities en route.

It also happens that there are *nonprobabilistic* accounts of belief popular among artificial intelligence researchers which can be modeled by convex sets of probability distributions, for example, the Dempster–Shafer formalism (Kyburg 1987), or the practice of default entailment by Possibility Theory (for discussion, see Benferhat, Dubois, & Prade 1997). Even if there were no other use for set-based models, exploring the relationships between probability distributions and competing formalisms for representing beliefs is an interesting thing to do (e.g., see Horvitz et al. 1986), and reason enough to tolerate the models' existence.

6. BELIEFS OUTSIDE A FINITE DOMAIN

It remains to be discussed why a single probability distribution might constrain belief values that do not correspond to any sentence of interest. One way such constraints arise occurs when the set of interesting sentences is being constructed, prior to the main activity of analyzing beliefs about the sentences which have found themselves inside the set. The believer, or some agent acting on the believer's behalf, must choose which sentences are included and which are not. A particularly effective criterion for omitting sentences is that they are neither interesting in themselves nor relevant to any sentence that is interesting.

The question of how to assess relevance engaged Cox, who treated the subject at length both in 1961 and 1978. Cox's work uses entropy measures, and emphasizes the circumstances in which the relevance of entire sets of sentences to each other can be evaluated. For the present purposes, it is enough to consider how individual sentences might be found irrelevant to others, and to pursue that subject using a single probability distribution without discussing entropy measures. The two approaches are entirely consistent with each other, however.

The usual warrant for labeling a sentence as irrelevant is that the believer estimates that its truth or falsehood does not change the credibility ordering of interesting sentences conditioned on any sentence that might be true. Within a probabilistic belief management regime, if x is irrelevant to interesting a and b on the assumption that c is true, and $c \wedge x$ is not certainly false, then this condition may be written as

$$p(a \mid c \wedge x) \geq p(b \mid c \wedge x) \quad \text{iff } p(a \mid c) \geq p(b \mid c). \tag{1}$$

An obvious necessary consequence of the inequality on the left is that

$$p(a \wedge x \mid c) \geq p(b \wedge x \mid c). \tag{2}$$

This is a constraint on the degrees of belief that can be attached to sentences $a \wedge x$ and $b \wedge x$, which are not to be included in the interesting domain because of (1). Similar constraints attach to conjunctions involving every sentence that is omitted from the interesting set on grounds of irrelevance.

The commitment binds when the believer assents to expression (1). Either the believer is using probability or else is relying on some other ground for assessing relevance. If only probability is being used, including Kolmogorov's definition of conditional probability, then (2) follows.

If the interesting set is finite, then there is no principled bound on the number of well-formed nonequivalent sentences which are omitted. Since the point of the exercise is to ignore x, it might be expedient to view all expressions like $p(a \wedge x \mid c)$ as variables, and simply cede the entire closed unit interval to them collectively without further ado.

Of course, the believer need take no notice of such matters. If the believer is a Bayesian, however, then the constraints are in the deductive closure of the believer's corpus of beliefs. If any are not, then some formalism other than probability has been used. There can be no correct statement of sufficient conditions for the exclusive use of ordinary probability as the means of belief management on finite sets which fails to place these constraints in the believer's deductive closure.

7. PARIS'S ASSUMPTION

This section considers briefly the criticisms of Cox's 1946 paper made by Paris (1994). Paris finds value in Cox's contribution, but takes issue with the rigor of the proof presented. In order to furnish rigor, Paris himself proves a version of Cox's Theorem, reporting that he was unable to dispense with an assumption missing from the 1946 paper.

Paris labeled this assumption Co5. When adapted to Cox's notation with degrees of belief scaled to the closed unit interval, it holds that

> For any $0 \leq \alpha, \beta, \gamma \leq 1$, and $\varepsilon > 0$ there are sentences a, b, c, and d where $a \wedge b \wedge c$ is not contradictory, such that each of
>
> $$|(d \mid a \wedge b \wedge c) - \alpha|, |(c \mid a \wedge b) - \beta|, |(b \mid a) - \gamma|$$
>
> is less than ε.

This is obviously a specific form of the requirement that degrees of belief may be found anywhere within the range of belief values. A finite set of sentences, each with a fixed belief value, cannot accomplish this.

Paris's criticism of the 1946 paper is fair. Cox was silent on the cardinality of his set of sentences, and in the vicinity of the proof he discussed only fixed (if unspecified) sentences, each with a functionally related degree of belief. Variable probabilities were discussed in 1946 after presentation of the main theorem (on pages 10 and 11), but in the context of point estimation for physically grounded limiting frequencies, without reliance on the theorem. It is in later work that Cox introduced plainly identified variable degrees of subjective belief, as in the 1961 book where the degrees are explicitly algebraic variables appearing in a system of simultaneous constraints. In his 1978 paper, one finds (at page 153) the direct statement that the mature Cox's conception of probability generally is indeed one of "a variable relation, capable of every degree between fixed limits."

In each of his three works, essentially the same theorem appears with strikingly similar proofs. It would be very difficult to argue that Co5 was beyond Cox's intention based on his complete life's work.

Halpern discusses Paris's assumption Co5 under the name of Assumption A4, and states (at page 1319), "[i]f we also assume A4 (or something like it), we can then recover Cox's theorem." In other words, Halpern accepts Paris's proof, albeit with the reservations about reasonableness already discussed.

8. CONCLUSIONS

The record absolves Cox in the matter of the purported counterexample. If Cox's discursive, decidedly unsyllogistic, expository style excited doubts about the rigor of his demonstration, then Paris repaired whatever lapse there may have been. The patch Paris fashioned is harmonious with Cox's writings after 1946, and appropriate for the applications which Cox made of his own theorem. Thus, there is no counterexample, and no lively controversy about the deductive correctness of Cox's conclusions.

Motivational questions can never be as crisply resolved as counterexamples can. Is it reasonable to assume that belief regulation must extend to merely possible, unused stations on the scale of credibility? Obviously not, if the believer's intuition says not.

There is some hope, however, of judging whether Cox himself made a reasoned choice in the issue. Cox's repeated rationale for presenting his theorem in 1946 was to provide a foundation for subjective probability without frequentist arguments. Cox articulated the defect he found in them.

> The difficulty of the frequency theory of probability may now be summarized. There is a field of probable inference which lies outside the range of that theory. The derivation of the rules of probability by ordinary algebra from the characteristics of the ensemble cannot justify the use of these rules in this outside field. Nevertheless, the use of these rules in this field is universal and appears to be a fundamental part of our reasoning. Thus the frequency theory is inadequate in the sense that it fails to justify what is conceived to be a legitimate use of its own rules.

In context, these "rules" are Kolmogorov's axioms and definitions. To remedy the situation required a theory of probability which does justify what is conceived to be a legitimate use of those rules. One legitimate use is, at least to Cox's satisfaction, treating degrees of belief as algebraic variables. The choices he made are consonant with the uses he pursued.

REFERENCES

BENFERHAT, S., D. DUBOIS, and H. PRADE. 1997. Possibilistic and standard probabilistic semantics of conditional knowledge. *In* Proceedings AAAI, pp. 70–75.

COX, R. T. 1946. Probability, frequency, and reasonable expectation. American Journal of Physics, **14**, 1–13.

COX, R. T. 1961. The Algebra of Probable Inference. Johns Hopkins Press, Baltimore.

COX, R. T. 1978. Of inference and inquiry: An essay in inductive logic. *In* The Maximum Entropy Formalism. *Edited by* R. D. Levine and M. Tribus. MIT Press, Cambridge, MA, pp. 119–167.

DE FINETTI, B. 1937. La prevision, ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincare, **7**:1–68. English translation by H. E. Kyburg, Jr. 1964. *In* Studies in Subjective Probability. *Edited by* H. E. Kyburg, Jr. and H. Smokler. John Wiley, New York.

EDWARDS, A. W. F. 1969. Statistical methods in scientific inference. Nature, **222**:1233–1237.

HALPERN, J. Y. 1996. A counterexample to theorems of Cox and Fine. *In* Proceedings AAAI, pp. 1313–1319.

HORVITZ, E. J., D. E. HECKERMAN, and C. P. LANGLOTZ. 1986. A framework for comparing alternative formalisms for plausible reasoning. *In* Proceedings AAAI, pp. 210–214.

KOLMOGOROV, A. N. 1933. Grundbegriffe der wahrscheinlichkeitrechnung. Ergebnisse der Mathematik, vol. 3. English translation by N. Morrison. 1956. Foundations of the Theory of Probability. Chelsea, New York.

KYBURG, H. E., Jr. 1987. Bayesian and non-Bayesian evidential updating. Artificial Intelligence, **31**:271–294.

KYBURG, H. E., Jr., and M. PITTARELLI. 1996. Set-based Bayesianism. IEEE Transactions on Systems, Man, and Cybernetics, **26**:324–339.

PARIS, J. B. 1994. The Uncertain Reasoner's Companion, A Mathematical Perspective. Cambridge University Press, Cambridge UK, Ch. 3.

SNOW, P. 1995. An intuitive motivation of Bayesian belief models. Computational Intelligence, **11**:449–459.