The Disappearance of Equation Fifteen, a Richard Cox Mystery

Paul Snow

P.O. Box 6134

Concord, NH 03303-6134 USA

paulusnix@cs.com

Abstract

Halpern has retracted an earlier claim that Cox's Theorem is deductively unsound, but he has renewed and amplified his objections to the reasonableness of the theorem for finite domains. His new argument highlights one functional equation used by Cox in 1946 but which is missing in 1978. The circumstances of its disappearance are explored, along with some of the advances in knowledge since 1946 which account for its absence.

Introduction

Cox's Theorem (1946, 1961, 1978) is a well-known foundational result for subjective probabilities. The theorem concerns the existence of real-valued functions which, when applied to suitable measures, yield ordinary probabilities which obey the usual additivity and product rules. Cox proves the theorem without reliance on any frequentist notions. Since the measures and functions in question exist under what many take to be mild assumptions, the theorem is interpreted as a normative motivation of belief models which feature probability.

In 1996, Halpern (also 1999a) claimed to have found a counterexample to Cox's Theorem, and said that Cox failed to disclose the domains of the variables which had appeared in the functional equations used in the 1946 version of Cox's proof. The bulk of Halpern's 1996 discussion and his counterexample involved the functional equation (8 on page 6 of 1946) which Cox had used to motivate the product rule:

$$F[\, F(\, x, y\,), z\,] = F[\, x, F(\, y, z\,)\,] \qquad \text{(C8)}$$

where the variables correspond to beliefs in selected logically related conditional expressions. Halpern also held that since the domains were of infinite density, that even if this had been disclosed, the theorem would be inapplicable to finite domains with fixed scalar beliefs.

In 1998, Snow pointed out that Cox had made the required disclosure and that Cox had viewed his probabilities to be real-valued variables, rather than specific numbers. Thus, the equations were intended to hold over the disclosed dense domains of these variables. There clearly would be a problem if each of a finite

number of conditional expressions $a \mid b$ had a specific real number attached to it, as Halpern had assumed and used in the construction of his counterexample, but Cox did *not* assume that. On the contrary, Cox would avoid such an arrangement based upon his recorded ideas about the nature of belief modeling and also upon the uses he made of his own result.

Halpern (1999b) withdrew his claim of counterexample relative to Cox, but persisted in his objection against the reasonableness of Cox's Theorem for finite domains. Halpern also made one new claim. A functional equation (15 on page 8 of 1946), which Cox derives using the product rule and whose typical solutions exhibit total probability, might raise some normative difficulties beyond those discussed earlier:

$$xS[\, S(\, y\,)/x\,] = yS[\, S(\, x\,)/y\,] \qquad \text{(C15)}$$

As before, the variables correspond to beliefs in some conditional expressions. Halpern acknowledges that Cox disclosed the domains of the variables in this equation, and does not renew the earlier claims of deductive lapse. The normative issue is whether assuming that this equation holds in a finite domain for all values which x and y could take is "natural."

In one sense, there is no new issue here. Halpern identifies as the source of his dissatisfaction the same domain density concerns as had figured in his earlier discussions of (C8). Anyone is obviously free to disagree with Cox that x and y should "naturally" be treated as variables rather than as specific numbers. But that was true for (C8), too.

On the other hand, (C15) is different from (C8). Equation (C8) expresses something interesting about belief models: functions which combine beliefs in certain logically related expressions might be expected to be associative. In contrast, equation (C15) does not seem to express anything fundamental about belief. Its symmetry is pretty, and it follows just as surely from Cox's assumptions (if not Halpern's) as (C8) does, but in itself it is a means to an end, and uninteresting as a destination in its own right.

Unlike (C8), equation (C15) disappeared from Cox's presentation of his theorem in 1978. Furthermore, Halpern notes that with one exception, none of the commentators on Cox whom he has read discuss (C15). Halpern offers an unsavory interpretation of these facts, and complains that something has been "swept under the carpet."

No such thing happened. (C15) disappeared for good reasons, and once lost, stayed lost. That it was superfluous to any point Cox was making in 1978 probably suffices to account for its absence there. But from a larger perspective, the older Cox had the benefit of relevant knowledge which was unavailable to him in 1946. A large part of this was the wide-ranging work of Aczél (1966) on functional equations. Another likely strand would be the twin papers of Schrödinger (1947) on what was the original question on Cox's mind, the role of subjective probability in the laws of physics. And although Cox did not cite this, nor did he need to, he would plausibly be aware of the famous counterexample of Kraft, Pratt, and Seidenberg (1959) which prominently brought the "density issue" to the attention of the mathematical community. Kraft, et al. also showed that additivity can be had more directly than by passing through total probability and the oblique (C15), a point which was reinforced by Aczél a few years later.

In the current paper, the development of these ideas begins with a clarification of Cox's views and some indication of their evolution. We then review Schrödinger's crisp dispatch of total probability without benefit of (C15). Aczél's and Kraft, et al.'s direct approaches to additivity are recalled next, along with the counterexample which seems to dispose of a conjecture discussed by Halpern. The parsimony of assumptions permitted by a direct derivation of additivity is illustrated by a simple recovery of ordinal Bayesian revision, both in typical statistical situations and in cases like one studied by Kyburg (1978).

Cox's Three Theorems

One complication in discussing "Cox's Theorem" as if it were a single thing is that Cox gave a distinct interpretation of his result on each occasion and varied his assumptions to match. Only the first (1946) version began with the assumption that there existed unspecified measures of "reasonable credibility," some of which turn out to be probabilities. In both 1961 and 1978, Cox took probabilities or functions of probabilities as his *starting* point. In 1961, this led to the conclusion that subjective probabilities shared a common mathematical foundation with frequentist ones. In 1978, his emphasis was on the usefulness of subjective probability as a general-purpose tool to support inductive reasoning. What relationship probabilities might enjoy with a possibly larger universe of reasonable measures of belief was explored in 1946, but simply does not come up in the later works.

The content of the theorem also changed subtly over time. It is now routine to say that Cox stated conditions for the existence of *order preserving* functions which transform suitable measures into probability measures. In fact, the order preserving attribute is not claimed in 1946. A basis for the usual statement is Cox's expression of kinship in 1978 with the closely related associativity

equation results developed by Aczél (1966), which do include the order preserving feature.

To the extent that the phrase *Cox's Theorem* is meant to refer to a straightforward existence result, then the order preserving aspect, which Halpern rightly describes as an "informal" understanding, is a reasonable interpretation, apparently satisfactory to Cox himself. However, if one seeks in Cox's work what Halpern calls a "compelling justification" for the use of some selected belief representations, then the absence of ordinal restrictions is crucial if one wishes to identify what representations are being justified.

At no time did Cox hold that the existence of such an order preserving function was necessary for a measure of belief to be reasonable. According to the standards of his 1946 paper, the only time when Cox discussed generic reasonable measures of belief, some possibility measures satisfy his criteria for reasonableness (Snow 2001). Possibility measures, of course, do not display ordinal agreement with any probability distribution other than in their shared special case of the Boolean assignment of zeros and ones. At the same time, many people obviously consider possibilities as quite reasonable. Had Cox implied otherwise, one would be entitled simply to dismiss his views on reasonableness as idiosyncratic.

Cox and Keynes

Another possible source of confusion is that Cox was an advocate of Keynes' (1921) perspective on probability rather than of the now more familiar Bayesianism. This is discussed on page 4 of the 1946 paper and in the preface of the 1961 book. A distinguishing characteristic of Keynes' theory is that not all conditional expressions are comparable to one another in credibility. That is, it is *not* the case that all $a \mid b$ would either be greater than, or less than, or else equal in credibility with all $c \mid d$. One of Cox's achievements was to give Keynes' partial-order doctrine a felicitous mathematical form, *sets* of belief measures. This occurred in the 1961 book. In 1946, he expressed partial-order differently and perhaps more simply as *variable* belief measures.

While Cox's 1946 theorem begins with talk of a "*measure* of reasonable credibility," one also finds on page 9, "It is hardly to be supposed that every reasonable expectation should have a precise numerical value." The explanation of this apparent equivocation is that in Cox's view an individual scalar measure is generally inadequate to represent a state of belief. Presumably, it was obvious to his readers in 1946 that an avowed Keynesian would not propose any individual measure as the entirety of a general belief representation. And, as is evident from even casual inspection of (C8) or (C15), Cox is treating the beliefs attached to conditional expressions as variables. That is understandable: Cox is treating these things as variables because for him they are variable.

Cox's identification of the partially ordered beliefs of Keynes with algebraic variables was innovative, but not

unprecedented. Polya (1941), who acknowledged influence by Keynes (although not to the same extent as Cox), had also offered a formalism in which subjective probabilities were treated as algebraic variables rather than as specific real numbers. The idea of incompletely specified probabilities itself is older than Keynes, and can be found throughout the later chapters of Boole (1854), who used a different algebraic notation.

In a similar vein, it is simply untrue that in 1946 Cox contemplates that even the individual measures of credibility would typically assign a real number to *every* pair of sentences in some domain of sentences, contrary to Halpern's report (1999b, at 129). Cox specifically denies this on page 6, where we read that "It is not to be supposed that a relation of likelihood exists between any two propositions." Ironically, Halpern (at 130) goes on to recite what is at stake for a Keynesian in this and in the passage from page 9 quoted earlier. If a real number were assigned to every conditional expression and this assignment was the representation of one's belief, then every conditional expression would be ordinally comparable to every other. That is, as Halpern observes, a strong assumption, one which Cox and Keynes abjured.

On the other hand, while Cox was not a Bayesian writer, he was not an *anti*-Bayesian writer, either. Cox imagined that belief could be represented by sets of probability measures defined by algebraic constraints. Bayesians could imagine that, too. It is, for example, an early step in the application of the maximum entropy method, was enthusiastically pursued for a while by de Finetti (1937), and appears congenial with some Bayesian approaches to practical probability elicitation, robustness, and opinion pooling.

Where Cox would part company with the Bayesians is in the *next* step. An orthodox Bayesian would choose one particular member of the set of probabilities as the representation of his or her belief. Cox would be content not to choose a particular member of the set, but rather would accept the whole set as representative of his belief.

The difference between partially ordered and completely ordered belief is important. For one thing, sets allow *all* possibility measures, not just those which satisfy the 1946 equations, to qualify as reasonable measures of belief by Cox's standards (Snow 2001). The difference also makes for some interesting problems when Cox's Theorem is interpreted as a normative finding in a Bayesian context.

For example, Jaynes (1963) reported that Cox had proven that "the mathematical rules of probability theory given by Laplace" are "unique in the sense that any set of rules in which we represent plausibility by real numbers is either equivalent to Laplace's, or inconsistent." Possibility is a difficult case. Although there are close relationships between it and some probabilities (Benferhat, *et al.* 1997), to describe it as flatly "equivalent" to anything in Laplace is strained or worse. The real difficulty with Jaynes' statement, however, is its omission of Cox's own view that we cannot *in general* represent plausibility by specific real numbers.

And yet, any argument which endorses representing belief by sets of probability distributions could, with the additional requirement that one ought to select a particular distribution, be fashioned into an argument for a kind of Bayesianism. Thus, with some care, a Bayesian would be entitled to cite Cox's Theorem as lending support and comfort to one's position. In the context of Halpern's criticisms, however, it would be important to note that while traffic with specific real numbers might be the *conclusion* of such an argument, it need not be an *assumption* of that part of the argument which leads to the sets, and might profit from some justification in any case.

Schrödinger

In 1946, independently of Cox, Erwin Schrödinger presented an axiomatic derivation of subjective probability free of any reference to frequentist notions. Cox cited Schrödinger (1947) without discussion in his 1961 book. Schrödinger did not use functional equations, nor did his derivation of total probability depend upon the product rule. To achieve his total probability result, he used a very strong assumption: the belief value $b()$ attached to $\neg a$ must be a monotonically decreasing, self-inverse function $n()$ of the value attached to a.

The bluntness of the assumption is to some extent mitigated by the narrow scope of Schrödinger's theory: only *events* (things which happen or do not) can be the subject of probability valuation. The assumption is motivated by the observation that "a conjecture about an event coming true amounts to the same as a conjecture about its not coming true." If the assumption is granted, then total probability emerges from it by simple algebra.

It is difficult to imagine that Cox failed to notice that Schrödinger's argument for his strictly decreasing functional negation was little different from Cox's own sole justification in 1946 for functional negation (with its decreasing character to be derived), "Since $\neg b$ is determined when b is specified, a reasonable assumption, and the least restrictive possible, appears to be that $\neg b \mid a$ is determined by $b \mid a$." (Note that Cox did not restrict his theory to events, and $b \mid a$ here is not any reasonable measure of credibility, but only a measure which obeys the product rule.) By 1961, the conclusion of the quoted passage is simply offered as an axiom (at page 3).

Turning ahead to 1978, by the time Cox takes up total probability, he has *already* assumed that his $b \setminus a$ is a probability (at 133). Complementary negation obtains necessarily. There is nothing for (C15) to prove, and sure enough, it is gone. Moreover, so far as we know, Cox has nothing to say about negation beyond what was satisfactory to Schrödinger to motivate strictly decreasing negation, from which total probability follows by convention. How surprising is it, then, that Cox would simply offer total probability as an assumption, bolstered with some brief remarks and a footnoted reference to his earlier work?

The 1959 Counterexample and Additivity

While Cox's own lack of interest in 1978 regarding (C15) appears sensible, it remains to be explained why later commentators would not be moved to revisit the issue, in light of the uncompelling character of arguments like Schrödinger's (and Cox's own) which attempt to motivate total probability. The short answer is that it has long been known that ordinal agreement with full additivity, of which total probability is a special case, can be had for about the same cost in assumptions as the product rule. And as we shall see in the next section, if additivity is motivated, ordinal Bayesian revision can be explained in many cases with an assumption less strenuous than Cox and Schrödinger's assumption for functional negation.

A motivation for associative disjunction,

$$\text{bel}(\text{ a} \vee \text{b}) = \text{F}[\text{ bel}(\text{ a}), \text{bel}(\text{ b} \wedge \neg\text{a})]$$

would be little different from Cox's arguments for associative conjunction leading to (C8). The case that such an $F()$ should also be increasing in each place was reviewed by de Finetti (1937), and $F()$ would then display the property which he called *quasi-additivity*. Aczél (1966) reported a variety of technical conditions which yield additive solutions for associative functional equations, including order-preserving solutions for the quasi-additive ones. To recover something comparable to Cox (1946), that under gentle conditions associative functional equations have additive solutions in *some* function of the original beliefs, would be straightforward. To the extent that the variables which appear in the functional equations would need dense domains, the same Keynesian considerations would supply them.

Although it is not relevant to Cox's concerns, there is also interest in which discrete domains might support additive order-preserving representations of quasi-additive disjunction. That not all discrete domains do so was established by Kraft, Pratt, and Seidenberg in 1959. They created the following arrangement of the 32 sentences formed from 5 atoms, in ascending order of credibility:

$\varnothing$, a, b, c, a $\vee$ b, **a $\vee$ c**, **d**, a $\vee$ d, b $\vee$ c, e, a $\vee$ b $\vee$ c,
b $\vee$ d, **c $\vee$ d**, **a $\vee$ e**, a $\vee$ b $\vee$ d, **b $\vee$ e**, **a $\vee$ c $\vee$ d**, c $\vee$ e,
b $\vee$ c $\vee$ d, a $\vee$ b $\vee$ e, a $\vee$ c $\vee$ e, d $\vee$ e, a $\vee$ b $\vee$ c $\vee$ d,
a $\vee$ d $\vee$ e, b $\vee$ c $\vee$ e, a $\vee$ b $\vee$ c $\vee$ e, b $\vee$ d $\vee$ e,
c $\vee$ d $\vee$ e, a $\vee$ b $\vee$ d $\vee$ e, a $\vee$ c $\vee$ d $\vee$ e, b $\vee$ c $\vee$ d $\vee$ e,
a $\vee$ b $\vee$ c $\vee$ d $\vee$ e

It is easily verified that the rank of any disjunctive sentence is an associative function of the ranks of its disjuncts, and this function is strictly increasing in each place. No ordinally agreeing additive $p()$ exists, since:

p(a) + p(c) < p(d)
p(a) + p(d) < p(b) + p(c)
p(c) + p(d) < p(a) + p(e)
p(b) + p(e) < p(a) + p(c) + p(d)

but the first three inequalities imply that $p(a) + p(c) + p(d) < p(b) + p(e)$, contradicting the fourth inequality.

This appears to dispose of a conjecture which Halpern attributes to a former student, that distinct degrees of belief might suffice for a finite discrete version of Cox's Theorem with order preservation. All the beliefs in the Kraft, *et al.* example are distinct, so distinct degrees of belief seem unhelpful. Kraft, *et al.* went on to state a sufficient condition for finite additive agreement, and research into the matter continues (Fishburn 1996).

Bayes' Rule and Kyburg's Parrot

If one has achieved a motivation of additivity, and selected a single probability as the representative of belief, then ordinal agreement with a form of Bayesian revision can be motivated with a weaker additional assumption than the strictly increasing associative conjunction used by Aczél to motivate the product rule. The argument is also simple, and with no functional equations, no density issue would come up. (A longer version for sets can also be spun.)

The additional assumption was offered by de Finetti (1937), and abstracts in ordinal form an earlier principle which Boole (1854) adopted and attributed to the astronomer W.F. Donkin. The assumption is that $p(a \mid e) \geq p(b \mid e)$ just when $p(ae) \geq p(be)$. The intuition seems acceptable to wider audience than the probabilist community. It is consistent, for example, with the orderings in the revision scheme for possibility of Dubois and Prade (1986).

The revision argument itself is straightforward, and a form of it has been used by DeGroot (1970). In the usual Bayesian case, common in statistical and Bayes nets work, there is a prior commitment to a comprehensive joint probability distribution which anticipates the possible evidence, at least in principle. The observation of a piece of evidence corresponds to the discard of all joint events contradictory to the observation, and Bayesian revision corresponds to a normalization of the probabilities of the surviving events. Obviously, the ordinal relationships are just those captured in de Finetti's principle. The discarded events conjoined with a contradictory observation would have prior and posterior probability of zero, also consistent with the principle. If one were only claiming ordinal agreement in the first place, then the task is accomplished.

The argument also works at least in a behavioral sense for what at first glance appears to be a polar opposite case, that of unforeseen evidence. That is, one has additive beliefs over some domain of interest, and something happens (outside the domain) which changes those beliefs, now also additive. The situation recalls that proposed by Kyburg (1978), where belief change may proceed not by conditioning, but by any mechanism at all, including the memorable image of adopting new additive beliefs by consulting with a parrot.

If we modify Kyburg's conditions to incorporate de Finetti's principle, then no observable violation of Bayes rule will occur even though formally no conditioning occurs to link earlier and later belief states. While there is no comprehensive corpus of beliefs which

includes the new observation before it is seen, de Finetti's intuition does suggest one prior constraint on foresight: $p(a) = 0$ implies that $p(a \mid e) = 0$. That is, one imagines that $p(ae)$ is never greater than $p(a)$ on typical belief-modeling grounds, and that de Finetti's assumption applies at least to this modest extent. Of course, even Bayesians sometimes learn that the "impossible" is true, but then they must rely on something other than simple conditioning in such cases.

If prior zeroes are respected, then the posterior probability values $q()$ are the same as would have been achieved had one performed a Bayesian revision of the prior values $p()$ by something proportional to $q(a)/p(a)$ for those atoms a where $p(a) > 0$. The specialized version of de Finetti's principle ensures that when $p(a) = 0$, then $q(a) = 0$, and so any multiplier at all could be imputed to such an atom. One never need consider what might have been other than what was, although it is easy enough to impute a hypothetical value to the evidence's negation to impose "total probability" on "conditional probabilities" based upon the imputed "likelihoods."

Nothing in the above disputes anything in Kyburg. The moral is that it is difficult for a believer with additive beliefs to behave inconsistently with Bayesian teaching. In cases where evidence is foreseen and planned for, a Bayesian need motivate little more than additivity to explain one's credal practices. The full-length paper discusses how de Finetti's principle can also help motivate additivity in light of Kraft, Pratt and Seidenberg's work.

Conclusions

Because many people have relied upon Cox's Theorem, there was great interest in the unfounded claim of its deductive unsoundness. While that has been cleared up, a principled normative dissent also deserves attention, since normative appeal is among the theorem's chief virtues. Because of its apparently unilluminating character and its unexplained disappearance from view, equation (C15) may have seemed like an apt place to probe.

With respect to Cox's original paper, (C15) is derived, not assumed, and the density of the domains of its variables reflects a distinctive assumption about the underdetermined character of *all* the belief-related variables throughout the argument. It has been widely known for four decades that a finite set of fixed scalars will not serve in such an argument. The question of whether beliefs are "naturally" fixed or variable has also been discussed for a long time, and one might have wished the liveliness of the issue to have been acknowledged.

As to the later Cox and those who have come after, the progress of knowledge has simply mooted (C15).

References

Aczél, J. 1966. *Lectures on Functional Equations and Their Applications*. New York: Academic Press.

Boole, G. 1854. *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. London: MacMillan.

Benferhat, S.; Dubois, D.; and Prade, H. 1997. Possibilistic and Standard Probabilistic Semantics of Conditional Knowledge. In *Proceedings of the AAAI Conference*, 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*. Baltimore: Johns Hopkins Press.

Cox, R. T. 1978. Of Inference and Inquiry: An Essay in Inductive Logic. In *The Maximum Entropy Formalism*. R. D. Levine and M. Tribus (eds.) Cambridge, MA: MIT.

De Finetti, B. 1937. La Prévision, Ses Lois Logiques, Ses Sources Subjectives. English translation by H.E. Kyburg, Jr. In *Studies in Subjective Probability*. H.E. Kyburg, Jr. and H. Smokler (eds.). New York: John Wiley.

De Groot, M.H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.

Dubois, D. and Prade, H. 1986. Possibilistic Inference Under Matrix Form. In *Fuzzy Logic and Knowledge Engineering*. H. Prade and C.V. Negoita (eds.). Berlin: Verlag TUV Rhineland.

Fishburn, P. C. 1996. Finite Linear Qualitative Probability. *Journal of Mathematical Psychology* 40:64-77.

Halpern, J. Y. 1996. A Counterexample to Theorems of Cox and Fine. In *Proceedings of the AAAI Conference*, 1313-1319.

Halpern, J. Y. 1999a. A Counterexample to Theorems of Cox and Fine. *Journal of Artificial Intelligence Research* 10:67-85.

Halpern, J. Y. 1999b. Cox's Theorem Revisited. *Journal of Artificial Intelligence Research* 11:129-135.

Jaynes, E.T. 1963. Review of *The Algebra of Probable Inference*. *American Journal of Physics* 31:66-67.

Keynes, J. M. 1921. *A Treatise on Probability*. London: Macmillan.

Kraft, C.H.; Pratt, J.W.; and Seidenberg, A. 1959. Intuitive Probability on Finite Sets. *Annals of Mathematical Statistics* 30:408-419.

Kyburg, H. 1978. Subjective Probability: Criticism, Reflection, and Problems. *Journal of Philosophical Logic* 7:157-180.

Polya, G. 1941. Heuristic Reasoning and the Theory of Probability. *American Mathematical Monthly* 48:450-465.

Schrödinger, E. 1947. The Foundation of the Theory of Probability, I and II. *Proceedings of the Royal Irish Academy* A51:51-66 and 141-146.

Snow, P. 1998. On the Correctness and Reasonableness of Cox's Theorem for Finite Domains. *Computational Intelligence* 14:452-459.

Snow, P. 2001. The Reasonableness of Possibility from the Perspective of Cox. *Computational Intelligence* 17. Forthcoming.