T. SEIDENFELD

REMARKS ON THE THEORY OF CONDITIONAL PROBABILITY: SOME ISSUES OF FINITE VERSUS COUNTABLE ADDITIVITY[1]

1. INTRODUCTION

Why does it matter whether probability is countably additive or merely finitely additive?

Recall the received view of mathematical probability. Let $\mathcal{B}$ be a σ-field of sets of subsets of Ω, with points ω. Ordinary elements of $\mathcal{B}$ are denoted by upper-case letters, 'E', 'F', etc. Kolmogorov's 1933 (1950) axiomatization of probability requires that $\forall(A, B) \in \mathcal{B}$:

AXIOM 1. $0 < P(A) < 1$.

AXIOM 2. $P(\Omega) = 1$.

AXIOM 3. If $A \cap B = \emptyset$ then $P(A) + P(B) = P(A \cup B)$.

Last, σ-additivity is taken by Kolmogorov as an "expedient" (1950, 15):

AXIOM 4. If $(A_i \cap A_j) = \emptyset$ whenever $i \neq j$ $(i, j = 1, 2, \ldots)$ then

$$P(\cup_i A_i) = \sum_i P(A_i).$$

Here are six reasons for considering the theory of finitely additive probability: the theory that results from Axioms 1-3, without requiring Axiom 4.

REASON 1. Measurability precludes non-trivial, σ-additive unconditional probabilities defined over the powerset of Ω when that is an uncountable set. (See Ulam's theorem, e.g., (Jech, 1978, 297)) That is, with the received theory, the domain of probability is restricted to a proper sub-σ-field of an uncountable Ω. However, an application of the Hahn-Banach theorem establishes that probability can always be extended to the power set if it is finitely additive (see Ash (1972)). The interesting work of Dubins and Savage (1976) relies on just this flexibility of finitely additive probability to avoid problems of measurability. However, this maneuver opens the door to an interesting debate over

V.F. Hendricks et al. (eds.), Probability Theory, 167–178.

non-constructive methods in probability since, by an important result due to Solovay, without the axiom of choice, using the axiom of dependent choice instead and a large cardinal assumption, all subsets of the continuum can be made Lebesgue measurable (see (Jech, 1978, 537)).

REASON 2. Limits of relative frequencies need not satisfy the axiom of countable additivity. For example, let the sample space be a countable set, $\Omega = \{\omega_1, \omega_2, \ldots\}$. Define the probability of an event E, $P(E)$, as the limit of relative frequency of E in a denumerable sequence S of repeated trials, with outcomes from Ω, $S = < o_1, o_2, \ldots >$, $o_i \in \Omega$, $i = 1, 2, \ldots$. Then, for each $\omega \in \Omega$, it can be that $o_i = \omega$ for only finitely many values of i, i.e., only finitely often does ω occur in the sequence S. Then, for each $\omega \in \Omega$, $P(\omega) = 0$, in violation of Axiom 4. See Kadane and O'Hagen (1995) for an interesting discussion of how such a finite, but not countably additive probability can be used to model selecting a natural number "at random."

REMARK 1. The class of events for which the limit of frequency is defined need not form even a field (see (Billingsley, 1986, problem 2.15.)). However, the same argument cited above, involving the Hanh-Banach theorem, shows that limits of relative frequencies can be extended to form a finitely additive probability on the powerset of Ω.

REASON 3. Some important decision theories require no more than finite additivity. e.g., deFinetti (1974), and Savage's (1954) theories. However, these theories require a controversial assumption about the state-independent utility for *consequences* (see Seidenfeld and Schervish (1983) and Schervish et al. (1990)).

REASON 4. Two-person, zero-sum games with bounded payoffs that do not have (minimax) solutions using σ-additive probability do, when finitely additive mixed strategies are permitted. For example, Wald's (1950) zero-sum game of picking the bigger integer, with +1 to the winner, −1 to the loser, and 0 in case of a tie, has no value within the class of σ-additive strategies. However (Schervish and Seidenfeld, 1996), the game is "fair," i.e. each

player's minimax strategy has value 0, when purely finitely additive strategies are used. Wald's treatment of statistical games with infinite parameter spaces often leads to maximin strategies for Nature that, as limits of countably additive strategies, are purely finitely additive. These priors are the "least favorable" possible from the statistician's point of view (see, e.g, (Berger, 1985, 350)).

REASON 5. Textbook, classical statistical methods have (extended) Bayesian models that rely on purely finitely additive prior probabilities. For example, Sir Harold Jeffreys' (1971) Bayes-models of inferences for locations/scale parameters rely on "improper" priors. For example, his theory adopts $d\mu$ the uniform density for Lebesgue measure as the "improper" prior for inference about the mean, μ, based on Normal $\mathbf{N}(\mu, 1)$ data. This prior is "improper" as it does not integrate to 1 (or any quantity) though, of course, Lebesgue meaure is σ-finite. Nonetheless, when it is used in formal Bayes calculations with the likelihood from the Normal $\mathbf{N}(\mu, 1)$ model, the resulting posterior distributions are proper and agree with the nominal Confidence Intervals for the same inference problem. However, this "improper" prior assigns equal (finite) weight to each bounded interval of length k for μ. Stated in terms of probabilities, this "improper" prior gives equal probability to each unit interval for μ. As the parameter space is a countable union of disjoint unit intervals, the "improper" prior density, $d\mu$, corresponds to a (class of) finitely but not countably additive prior probability distribution(s) on μ.

REASON 6. When $P(A) > 0$, $A \in \mathcal{B}$, then conditional probability given A is well defined by

$$P(\cdot|A) = P(\cdot \cap A)/P(A).$$

The focus in the balance of this paper is on some differences between the two theories, with and without the axiom of countable additive, for the development of conditional probability with respect to null events, when $P(A) = 0$.

2. CONDITIONAL PROBABILITY $P(\cdot|A)$ WHEN $P(A) = 0$

Let $< \Omega, \mathcal{B}, P >$ be a measure space. Denote by ω points in Ω. When $P(A) > 0$, $A \in \mathcal{B}$, then conditional probability over $\mathcal{B}$ given A is well defined by the familiar rule, $P(\cdot|A) = P(\cdot \cap A)/P(A)$. Of course, this approach does not provide guidance for conditional probability given P-null (measure 0) events. For that, the received view comes from Kolmogorov's seminal 1933 work. In the usual terminology, with $\mathcal{A}$ a sub-σ-field of $\mathcal{B}$,

$P(\cdot|\mathcal{A})$ is a *regular conditional distribution* [rcd] on $\mathcal{B}$, given $\mathcal{A}$ provided that:

1. For each $\omega \in \Omega$, $P(\cdot|\mathcal{A})(\omega)$ is a probability on $\mathcal{B}$.
2. For each $B \in \mathcal{B}$, $P(B|\mathcal{A})(\cdot)$ is an $\mathcal{A}$-measurable function.
3. For each $A \in \mathcal{A}$, $P(A \cap B) = \int_A P(B|\mathcal{A})(\omega)dP(\omega)$.

That is, $P(B|\mathcal{A})$ is a version of the Radon-Nikodym derivative of $P(\cdot \cap B)$ with respect to $P(\cdot)$.

2.1. *Two limitations in this approach are well documented*

- An rcd may not exist.

The canonical example of a measure space that admits no rcd's is obtained by extending the σ-field of Borel sets on $[0, 1]$ under Lebesgue measure with the addition of one non-measurable set. Denote the original measure space by $< [0, 1], \mathcal{A}, \lambda >$. A familiar maneuver allows an extension to an enlarged measure space, denoted $< \Omega, \mathcal{B}, P >$. However, there is no rcd $P(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}$ given $\mathcal{A}$. (See (Halmos, 1950, 211), (Billingsley, 1986, Exercise 33.13), (Breiman, 1968, 81), (Doob, 1953, 624), or (Loeve, 1955, 370) for variations on this common theme.) Though, for each $B \in \mathcal{B}$, the extended measure space has Radon-Nikodym derivatives $P(B|\cdot)$ satisfying condition 3, above, these resist assembly of pointwise probabilities into a full probability distribution on $\mathcal{B}$, as required by condition 1.

In recent work (Seidenfeld et al., 2000) we show that, quite generally, a measure space admitting rcd's can be extended to another measure space admitting rcd's if and only if the latter lies within the measure completion of the former. This finding hints at other links between (a) the aforementioned problem of nonmeasurable sets for an unconditional probability and (b) the theory of conditional probability distributions when Axiom 4 is imposed. We discuss that in our paper (2000).

In rejoinder to the existence problem, however, a sufficient condition for rcd's to exist on $\mathcal{B}$ (given any sub σ-field $\mathcal{A}$) is that $\mathcal{B}$ be isomorphic under a 1-1 measurable mapping to the σ-field of a random variable (see, (Billingsley, 1986, T.33.3) or (Breiman, 1968, T. 4.30)).

- As Kolmogorov notes $P(\cdot|\mathcal{A})$ is *not* probability given an *event.*

He illustrates this with the so-called "Borel" paradox. Put simply, it is that $P(\cdot|\mathcal{A})$ is *not* probability given an *event* but, rather, probability given a *σ-field.* Specifically, with $\mathcal{B}_Z$ a σ-field generated by the random variable Z, let sub-σ-fields $\mathcal{A}_X$ and $\mathcal{A}_Y$ be generated by the random variables X and Y, respectively. Suppose that $X = x$ is the same event (in $\mathcal{B}_Z$) as $Y = y$. Nonetheless, when $X(\omega) = x$ the two rcd's, $P(\cdot|\mathcal{A}_X)(\omega)$ and $P(\cdot|\mathcal{A}_Y)(\omega)$, may be different, with sup norm arbitrarily close to 1.

In rebuttal, Kolmogorov points out that between any two conditioning sub-σ-fields, this "paradox" can occur only on a P-null set of points. That is, it is a measure-0 failure, at worst.

2.2. *Improper rcd's.*

The focus for the balance of this section comes from important work by Blackwell and Ryll-Nardzewski (1963) and Blackwell and Dubins (1975)) (theorems reported here that are not given explicit citations can be found in Seidenfeld et al. (2000)).

DEFINITION 1.

- An rcd $P(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}$ given $\mathcal{A}$, is *proper at ω* if $P(\cdot|\mathcal{A})(\omega) = 1$ whenever $\omega \in A \in \mathcal{A}$.
- $P(\cdot|\mathcal{A})(\omega)$ is *improper at ω*, otherwise.
- $P(\cdot|\mathcal{A})$ is proper, if $P(\cdot|\mathcal{A})(\omega)$ is proper at every $\omega \in \Omega$.

Two other useful concepts for the presentation here are these:

- An *$\mathcal{A}$-atom* is the intersection of all elements of $\mathcal{A}$ that contain given point ω of Ω.

- A probability distribution is *extreme* if its range is the two point set $\{0,1\}$.

THEOREM 1. Blackwell and Dubins (1975). When $\mathcal{B}$ is a countably generated σ-field, *no* rcd on $\mathcal{B}$ given $\mathcal{A}$ is proper if there exists *some* extreme probability on $\mathcal{A}$ supported by no $\mathcal{A}$-atom belonging to $\mathcal{A}$.

In other words, provided there exists one extreme probability on $\mathcal{A}$ which is not supported by its $\mathcal{A}$-atoms, then the sub-σ-field $\mathcal{A}$ is anomalous for all rcd's on $\mathcal{B}$ given $\mathcal{A}$ in that they are improper, each and every one!

Now, for our central result about the extent of impropriety in certain rcd's. Assume that $\mathcal{A}$ is an atomic sub-σ-field of $\mathcal{B}$, with $\mathcal{A}$-atoms a. Denote by $a(\omega)$ that $\mathcal{A}$-atom containing the point ω.

THEOREM 2. Let P be an extreme probability on $\mathcal{A}$ that is not supported by any of its $\mathcal{A}$-atoms. If an rcd $P(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}$ given $\mathcal{A}$ exists, there is one where $P\{\omega : P(a(\omega)|\mathcal{A})(\omega)) = 0\} = 1$. And, if $\mathcal{B}$ is countably generated, then this rcd is unique.

Theorem 2 asserts that when $\mathcal{B}$ is countably generated, almost surely with respect to P, the rcd's on $\mathcal{B}$ given $\mathcal{A}$ are maximally improper, in two senses simultaneously:

1. The set of points where propriety fails has measure 1 under P.
2. For $P(a(\omega)|\mathcal{A})(\omega) = 0$ when propriety requires that
$$P(a(\omega)|\mathcal{A})(\omega) = 1.$$

Here are several examples of Theorem 2.

EXAMPLE 1. (See (Billingsley, 1986, E33.11)). Let $\Omega = [0,1]$, let $\mathcal{B}$ = the Borel subsets of Ω, and let P be Lebesgue measure. Let $\mathcal{A}$ be the sub-σ-field of all countable and co-countable sets in $[0,1]$. Clearly, $P(A) = 0$ or $P(A) = 1$, for each $A \in \mathcal{A}$. Equally obvious, $P(A) = 0$ for each countable set A. Note that the $\mathcal{A}$-atoms, which in fact belong to $\mathcal{A}$, are just the points of Ω, $\{x : 0 \leq x \leq 1\}$. Hence, according to Theorem 2, the rcd on $\mathcal{B}$ given $\mathcal{A}$ satisfies:

$$P\{x : P(x|\mathcal{A})(x) = 0, \text{ for } 0 \leq x \leq 1\} = 1.$$

EXAMPLE 2. (See (Blackwell and Dubins, 1975, 742)). Let $\Omega = \{0,1\}^{\aleph_0}$ – the sample space of infinite binary sequences. Let $\mathcal{B}$ = the Borel subsets of Ω. And let P be the product measure, corresponding to independent "fair" coin flips. Let $\mathcal{A}$ be the tail σ-field for this process. Then, by the Kolmogorov 0-1 law, for $A \in \mathcal{A}$, $P(A) = 0$ or $P(A) = 1$. $\mathcal{A}$ is atomic. The $\mathcal{A}$-atoms, a, are countable sets of points, $\omega \in \Omega$, where ω', $\omega \in a$ if they differ in at most finitely many places. Since each $\mathcal{A}$-atom is a countable set, $P(a) = 0$; hence, P is not supported by any of its $\mathcal{A}$-atoms. Thus, $P\{\omega : P(a(\omega)|\mathcal{A})(\omega) = 0\} = 1$.

Example 2 has a generalization to i.i.d. binomial "weighted" coin flipping, $P_\theta(\{1 \times \{0,1\} \times \ldots\}) = \theta$, $0 < \theta < 1$, included in the next result, Theorem 3, dealing with symmetric measures, as covered by deFinetti's representation theorem for exchangeable processes.

EXAMPLE 3. (Continuing example 2). Let $\Omega = \{0,1\}^{\aleph_0}$; let $\mathcal{B}$ = the Borel subsets of Ω; and let P be a symmetric probability, in the sense of Hewitt and Savage (1955) defined as follows. Let T be an arbitrary finite permutation of the positive integers, i.e., a permutation of the coordinates of Ω that leaves all but finitely many places fixed. For $B \in \mathcal{B}$, given T, define the set $T^{-1}B = \{\omega : T(w) \in B\}$.

> P is called a *symmetric probability* if $P(T^{-1}B) = P(B)$, for each $B \in \mathcal{B}$ and T. If $B = T^{-1}B$ for all (finite) permutations T, B is called a *symmetric event.*

Let $\mathcal{A}$ be the sub-σ-field of $\mathcal{B}$ generated by the class $\mathbf{T}$ of all finite permutations of the coordinates of Ω, i.e., $\mathcal{A}$ is the σ-field of the symmetric events. $\mathcal{A}$ is atomic, with $\mathcal{A}$-atoms comprised by a countable set of sequences, each pair of sequences in the same atom differing by some finite permutation of its coordinates. In all but two cases the $\mathcal{A}$-atoms are countably infinite sets; the two exceptions are the two constant sequences.

THEOREM 3. Each rcd $P(\cdot|\mathcal{A})$ on $\mathcal{B}$ given $\mathcal{A}$, for a symmetric probability P, satisfies $P\{\omega : P(a(\omega)|\mathcal{A})(\omega) = 0\} = 1$, provided that $P(<0,0,0,\ldots>) = P(<1,1,1,\ldots>) = 0$.

REMARK 2. The proof of Theorem 3 relies on three considerations:

1. *The Hewitt-Savage 0-1 law for i.i.d. probability* P_θ *over the field of symmetric events – to establish that such an i.i.d probability is extreme and is not supported by* $\mathcal{A}$*-atoms whenever* $0 < \theta < 1$. *Hence, by Theorem 2, the rcd's for* $P_\theta(\cdot|\mathcal{A})(\omega)$ *on* $\mathcal{B}$ *given* $\mathcal{A}$ *are maximally improper.*
2. *DeFinetti's representation of symmetric probabilities as a mixture of i.i.d probabilities* P_θ*;*
3. *A representation of the rcd for a symmetric probability* $P(\cdot|\mathcal{A})(\omega)$ *on* $\mathcal{B}$ *given* $\mathcal{A}$ *as a mixture, depending upon* ω*, of the rcd's* $P_\theta(\cdot|\mathcal{A})(\omega)$.

To the extent that the rcd's for these examples are maximally improper, they cannot serve as coherent conditional probability distributions given the conditioning sub-σ-field—their conditional probability distributions are not supported by their conditioning events! Thus, we see that the received σ-additive theory of regular conditional distributions suffers some faults that prevent it from serving as a fully general account of degrees of belief for hypothetical reasoning when conditioning on certain (non-countably generated) sub-σ-fields.

3. SOME FINITELY ADDITIVE CONDITIONAL PROBABILITY

In dealing with finitely additive conditional probabilities $P(\cdot|\cdot)$, especially for conditional probability given events of unconditional probability 0, we adopt the following restriction:

Principle of Conditional Coherence (see Dubins (1975)):

For all pairs of events, A and B such that $A \cap B \neq \emptyset$,

1. $Q(\cdot) = P(\cdot|A)$ is a finitely additive probability.
2. $Q(A) = 1$, and $Q(\cdot|B) = P(\cdot|A \cap B)$.

When $P(A \cap B) > 0$, this principle applies, trivially. It formalizes the idea deFinetti (1974) had of conditional probability *given an event*, rather than *given a field*. That is, $P(\cdot|A)$ does not depend

upon how A^c is partitioned. Hence, there is no "Borel" paradox to solve. Moreover, by clause (2), each coherent finitely additive conditional probability is proper. Last, Dubins (1975) shows that each unconditional finitely additive probability P carries a full set of coherent conditional probabilities.

What, then, is special about finitely additive conditional probability? Consider the following:

EXAMPLE 4. (attributed to P. Lévy, by deFinetti): Let P be a finitely additive probability on the denumerable set of all pairs $< s,t >$, for s and t positive integers, with the following two restrictions:

1. $P(< s,t >) = 0$, *so that* P *is* 0 *on finite sets of pairs—*P *is purely finitely additive;*
2. $P(< s,t > |B) = 0$ *if* B *is an infinite set, so that the conditional* $P(\cdot|B)$ *is also purely finitely additive when* B *is an infinite set.*

Define the events:

$E = \{< s,t >: s > t\}$,

$S_m = \{< s,t >: s = m\}$ $(m = 1, \ldots)$

and

$T_n = \{< s,t >: t = n\}$ $(n = 1, \ldots)$.

Then, $P(E|S_m) = 0$ *for* $m = 1, \ldots$, *yet* $P(E|T_n) = 1$ *for* $n = 1, \ldots$.

Evidently the following principle is invalid for the finitely additive probability P, though it is satisfied by all σ-additive probabilities. Let $\Pi = \{h_n : n = 1, \ldots\}$ be an exhaustive partition and assume that P is defined on the field $\mathcal{B}$.

- Principle of ($\aleph_0$-)Conglomerability for Events (de Finetti):

$\forall(A \in \mathcal{B})$: If $c_1 = P(A|h_n) = c_2$ $(n = 1, \ldots)$, then $c_1 = P(A) = c_2$.

Note that in Example 4, conglomerability fails in at least one of the two orthogonal partitions, $\Pi_1 = \{S_m : m = 1, \ldots\}$ and $\Pi_2 = \{T_n : n = 1, \ldots\}$. Where conglomerability fails, there the finitely additive probability P fails also to satisfy condition 3 for regular conditional probabilities. That is, then $P(A) \neq$

$\int_\Omega P(A|h_n)dP(h)$, and P is not *disintegrable* in the partition Π. In an earlier study (Schervish et al., 1984) we showed that:

THEOREM 4. Each merely finitely additive probability fails conglomerability in some denumerable partition.

REMARK 3. Unless the probability assumes only finitely many values, there exists such a partition where conglomerability fails, whose elements each have positive prior probability, unlike in the Lévy example.

Thus, as highlighted here, the trade-off between the two theories of conditional probability can be summarized this way. With a countably additive probability, when conditioning on sub-σ-fields that are not countably generated (e.g., the field of symmetric events), the risk is that the rcd's may be maximally improper. By contrast, with a finitely additive probability, though conditional probability always is proper – the conditional probability is coherent – the risk is that the conditional probability will fail to be conglomerable (will fail to be disintegrable) in the partition of interest.

Recall that each countably additive space $< \Omega, \mathcal{B}, P >$ can be extended to a coherent finitely additive probability on the powerset of Ω. In case the rcd's $P(\cdot|\mathcal{A})$ on $\mathcal{B}$ given $\mathcal{A}$ are improper, the relevant question is whether there exists a finitely additive extension Q of P that is conglomerable (i.e., disintegrable) in the partition of the $\mathcal{A}$-atoms. If so, then one may blend the familiar properties of P, with the coherence of Q. That is, when $P(B) > 0$, then one may calculate using P's (improper) rcd's in the usual way, $P(B) = \int_\Omega P(B|\mathcal{A})(\omega)dP(\omega)$, and know that this agrees with $Q(B)$. However, for small events when $P(B) = 0$, then the propriety of Q's conditional probabilities ensures that, e.g., $Q(B|a(\omega))$ is coherent. In such a case, say that Q is a useful finitely additive extension of P.

Dubins (1977) shows that useful finitely additive extensions of Lebesgue measure exist for the case of the tail-field, Example 2. It is an open question, I believe, under what conditions such useful finitely additive probabilities exist generally.

Department of Philosophy
Carnegie Mellon University
USA

NOTES

[1]The paper is based on collaborated research with J.B. Kadane and M.J. Schervish, Department of Statistics, Carnegie Mellon University. That research was supported by NSF Grant DMS 9801401.

REFERENCES

Ash, R. B. (1972). *Real Analysis and Probability*, Academic Press, New York.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edn, Springer, New York.

Billingsley, P. (1986). *Probability and Measure*, 2nd edn, John Wiley, New York.

Blackwell, D. and Dubins, L. E. (1975). On existence and non-existence of proper, regular, conditional distributions, *Annals of Probability* **3**: 741–752.

Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions, *Annals of Mathematical Statistics* **34**: 223–225.

Breiman, L. (1968). *Probability*, Addison-Wesley, Reading MA.

deFinetti, B. (1974). *The Theory of Probability*, John Wiley, New York.

Doob, J. (1953). *Stochastic Processes*, John Wiley, New York.

Dubins, L. (1975). Finitely additive conditional probabilities, conglomerability, and disintegrations, *Ann. Prob.* **3**: 89–99.

Dubins, L. (1977). Measurable, tail distintegrations of the Haar integral are purely finitely additive, *Proc. Amer. Math. Soc.* **62**: 34–36.

Dubins, L. and Savage, L. (1976). *Inequalities for Stochastic Processes: How to gamble if you must*, Dover, New York.

Halmos, P. (1950). *Measure Theory*, van Nostrand, New York.

Hewitt, E. and Savage, L. (1955). Symmetric measures on Cartesian products, *Amer. Math. Soc. Trans.* **80**: 470–501.

Jech, T. (1978). *Set Theory*, Academic Press.

Jeffreys, H. (1971). *The Theory of Probability*, 3rd edn, Oxford Univ. Press, Oxford.

Kadane, J. and O'Hagen, A. (1995). Using finitely additive probability: Uniform distributions on the natural numbers, *J. Amer. Stat. Assoc.* **90**: 626–631.

Kolmogorov, A. (1950). *Foundations of the Theory of Probability*, Chelsea, New York.

Loeve, M. (1955). *Probability Theory*, van Nostrand, New York.

Savage, L. (1954). *The Foundations of Statistics*, John Wiley, New York.

Schervish, M. and Seidenfeld, T. (1996). A fair minimax theorem for two-person (zero sum) games involving finitely additive strategies, in *Bayesian Analysis in Statistics and Econometrics*. Eds. D. Berry, C. Chaloner, and J. Geweke. New York: John Wiley.

Schervish, M., Seidenfeld, T. and Kadane, J. (1984). The extent of non-conglomerability of finitely additive probability, *Z. War.* **66**: 205–226.

Schervish, M., Seidenfeld, T. and Kadane, J. (1990). State-dependent utilities, *J. Amer. Stat. Assoc.* **85**: 840–847.

Seidenfeld, T. and Schervish, M. (1983). A conflict between finite additivity and avoiding dutch book, *Phil. Sci.* **50**: 398–412.

Seidenfeld, T., Schervish, M. and Kadane, J. (2000). Improper regular conditional distributions, Technical Report, Dept. of Statistics, Carnegie Mellon University.

Wald, A. (1950). *Statistical Decision Functions*, John Wiley, New York.