June 6, 2013

ON THE EQUIVALENCE OF CONGLOMERABILITY AND DISINTEGRABILITY FOR UNBOUNDED RANDOM VARIABLES.*

BY MARK J. SCHERVISH, TEDDY SEIDENFELD AND JOSEPH B. KADANE
Carnegie Mellon University

We extend a result of Dubins (1975) from bounded to unbounded random variables. Dubins (1975) showed that a finitely additive expectation over the collection of bounded random variables can be written as an integral of conditional expectations (disintegrability) if and only if the marginal expectation is always within the smallest closed interval containing the conditional expectations (conglomerability). We give a sufficient condition to extend this result to the collection of all random variables that have finite expected value and whose conditional expectations are finite and have finite expected value.

1. Introduction. In discussions of the foundations of probability, a long-standing topic of debate is whether to require, beyond being finitely additive, that probabilities are countably additive. Specifically, we take the following three axioms to constitute the theory of finitely additive probability. Let $\{\Omega, \mathcal{B}\}$ be a measurable space. For all $A, B \in \mathcal{B}$,

Axiom 1: $0 \leq P(A) \leq 1$.

Axiom 2: $P(\Omega) = 1$.

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

Countable additivity, which is taken by Kolmogorov (1956, p. 15) as an "expedient", requires the following. Let $\{A_i\}_{i=1}^{\infty}$ be elements of $\mathcal{B}$.

*SHORT TITLE: CONGLOMERABILITY AND DISINTEGRABILITY

MSC 2000 *subject classifications.* Primary 60A05; secondary 28C05.
Key words and phrases. finite additivity, law of total probability, Daniell integral, coherence.

Axiom 4: If $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Call a probability *merely finitely additive* if it satisfies the first three axioms but fails the fourth one. In this paper, we assume only the first three axioms.

One major distinction between merely finitely additive probabilities and countably additive probabilities involves the theory of conditional probability. We take it as non-controversial that conditional probability satisfies this product rule:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A).$$

And when the conditioning event has positive probability, $P(A) > 0$, we can use this rule to fix conditional probability by unconditional probability: $P(B|A) = P(A \cap B)/P(A)$.

However, when the conditioning event is null, $P(A) = 0$, the countably additive theory denies that conditional probability is defined given such an event A . Rather, conditional probability is understood through the Radon-Nikodym theorem as a solution to an integral equation with respect to a sub-σ-field.

DEFINITION 1. Let $\mathcal{A}$ be a sub-σ-field of $\mathcal{B}$. Then $P(\cdot|\mathcal{A})(\cdot)$ is a *regular conditional probability* on $\mathcal{B}$, given $\mathcal{A}$ if

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, and
3. for each $A \in \mathcal{A}$, $P(A \cap B) = \int_A P(B|\mathcal{A})(\omega) dP(\omega)$.

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

Kolmogorov (1956, Section 5.2) points out that regular conditional probabilities admit the so-called "Borel Paradox". To summarize the paradox, let $\mathcal{A}$ and $\mathcal{A}'$ be the σ-fields generated by two different random variables X and X' respectively. The paradox finds an event A, a value x of X, and a value x' of X' such that the two events $\{X = x\}$ and $\{X' = x'\}$ are identical, but $P(A|X = x) \neq P(A|X' = x')$. That is, with $B = \{X = x\} = \{X' = x'\}$, we find that $P(A|B)$ depends on which σ-field we choose for conditioning. It is well-known that, for a specific pair X and X', the sets of all x and x' values that lead to this paradox form sets of probability 0. However, Kadane, Schervish and Seidenfeld (1986) illustrate how one can make the paradox occur with positive probability by considering more than countably many random variables at a time.

In contrast to the countably additive theory, (see Krauss, 1968 and Dubins, 1975) finitely additive conditional probabilities can be fully defined given each non-empty event in $\mathcal{B}$, while satisfying the following generalization of the product rule:

For all A, B, and C such that $B \cap C \neq \emptyset,\ P(A \cap C|B) = P(C|B)P(A|B \cap C)$.

However, the cost for these finitely additive conditional probabilities includes the penalty that they may fail to satisfy the integral property of regular conditional probabilities (clause 3 in Definition 1). In particular, there can exist a denumerable partition $\pi = \{A_i\}_{i=1}^{\infty}$ and an event B such that

$$P(B) \neq \sum_{i=1}^{\infty} P(B|A_i)P(A_i). \tag{1}$$

If (1) holds, we say that P *fails to be disintegrable* in the partition π. (See Definition 5 below for a precise definition.) Here is an elementary illustration, due to Dubins (1975) and discussed further by Kadane, Schervish and Seidenfeld (1996).

EXAMPLE 1. Let $\Omega = \{0, 1\} \times \{1, 2, \ldots\}$. Let P be a finitely additive probability such that satisfies the following:

- $P((1, i)) = 2^{-i-1}$ for $i = 1, 2, \ldots$,
- $P(B) = 1/2$, where $B = \{(0, 1), (0, 2), \ldots\}$,
- $P((0, i)) = 0$ for $i = 1, 2, \ldots$.

Define the partition $\pi = \{A_i\}_{i=1}^{\infty}$, where $A_i = \{(0, i), (1, i)\}$. Since $P(A_i) = 2^{-i-1} > 0$ for all i, we have

$$P(B|A_i) = \frac{P(B \cap A_i)}{P(A_i)} = 0,$$

for all i. Hence $\sum_{i=1}^{\infty} P(B|A_i)P(A_i) = 0 \neq P(B)$.

The concept of disintegrability is relevant to understanding some otherwise anomalous features of Bayesian statistical inference that arise when using so-called *improper* priors. These are instances of the so-called *marginalization paradoxes* of Dawid, Stone and Zidek (1973). As Kadane, Schervish and Seidenfeld (1996, Section 5) explain, an improper prior, e.g. Lebesgue measure over the whole real line, corresponds to a merely finitely additive prior probability on the real line. Each unit interval has equal probability, i.e. probability 0. Even when the formal posterior computed from the improper prior turns out to be countably additive, the joint (finitely additive) probability may fail to be disintegrable in the partition determined by the data.

In Example 1, we also see that the conditional probabilities $P(B|A_i)$ have the property that there exists $\epsilon > 0$ such that $P(B) > P(B|A_i) + \epsilon$ for every i. de Finetti (1930) says that such a probability *fails conglomerability* in the partition. (See Definition 5 for a more precise definition.) Schervish, Seidenfeld and Kadane (1984) show that each merely finitely additive probability fails conglomerability in some denumerable partition, which is not possible for countably additive probabilities. As we saw above, for nondenumerable partitions, the countably additive theory imposes disintegrability on the definition of conditional probability. Consequently the theory is not able to define $P(A|B)$ for arbitrary sets A and B, as we saw in the Borel Paradox.

Dubins (1975) established an important equivalence between conglomerability and disintegrability of finitely additive expectations, which de Finetti (1974) calls (coherent) previsions (see Definition 2 below). Dubins showed that, with respect to the class of bounded random variables, by replacing countably additive probability and conditional probability with the more general concepts of (finitely additive) expectations and conditional expectations, then a finitely additive expectation function is disintegrable in a partition if and only its conditional expectations are conglomerable in that partition. Evidently, when a finitely additive expectation fails conglomerability in a partition π, then it fails to be disintegrable in π. The converse inference is the heart of Dubins' result.

In this paper we extend Dubins' result to particular classes of unbounded random variables. To facilitate this, we extend the theory of finitely additive integrals to unbounded random variables using ideas from the theory of the Daniell integral. The finitely additive Daniell integral is equivalent to the integral as developed by Dunford and Schwartz (1958, Chapter III) for bounded random variables while extending the notion of coherent prevision to unbounded random variables. However, for unbounded random variables, not all real-valued coherent previsions admit an integral representation in the sense of Dunford and Schwartz. For discussion of the problem, see Berti, Regazzini, and Rigo (2001); Berti and Rigo (2000, 2002); Schervish, Seidenfeld, and Kadane (2008), and Seidenfeld, Schervish, and Kadane (2009).

In Section 2 we review de Finetti's concept of coherent previsions and conditional previsions, and show how to extend these to unbounded random variables.

In Section 3 we review Dubins' result and discuss how to extend conglomerability to unbounded random variables. In Section 4 we establish an equivalence between disintegrability and conglomerability for previsions of unbounded random variables, under the conditions that the domain of the prevision function includes the linear span of the random variables for which previsions are given and that previsions and conditional previsions are finite. In Section 5 we show that the equivalence is not valid if the domain of the prevision function includes merely the linear span of the bounded random variables. We offer a concluding discussion in Section 6.

2. Background. Let Ω be a fixed non-empty set, and define a *random variable* to be a real-valued function X defined on Ω. We require that $X(\omega)$ be finite for all $\omega \in \Omega$, but not necessarily bounded. All of the collections of random variables discussed in the definitions and results of this paper are allowed to include unbounded random variables unless explicitly stated otherwise.

The concept of coherent prevision on a collection of random variables was introduced by de Finetti (1974).

DEFINITION 2. Let $\mathcal{P}$ be a collection of random variables defined on Ω. A function $P : \mathcal{P} \to \mathbb{R}$ is called a *prevision*. We say that P is *incoherent* if there exists a finite subset $\{X_1, \ldots, X_n\}$ of $\mathcal{P}$ and scalars $\alpha_1, \ldots, \alpha_n$ and $\epsilon > 0$ such that, for all $\omega \in \Omega$,

$$\sum_{i=1}^{n} \alpha_i [X_i(\omega) - P(X_i)] < -\epsilon. \tag{2}$$

If P is not incoherent, we say that P is *coherent*.

An equivalent, and sometimes more convenient, way to define coherent prevision is to say that P is coherent if, for every finite subset $\{X_1, \ldots, X_n\}$ of $\mathcal{P}$ and all scalars $\alpha_1, \ldots, \alpha_n$,

$$\sup_{\omega \in \Omega} \sum_{i=1}^{n} \alpha_i [X_i(\omega) - P(X_i)] \geq 0. \tag{3}$$

It is not difficult to see that this is equivalent to Definition 2. As defined above, a coherent prevision P must assume only finite values, otherwise (3) would be impossible. There is a way to generalize the concept of coherent prevision to allow infinite values, but such a generalization will play no role in the results of this paper. (See Crisma, Gigante and Millossovich, 1997 for one such generalization.)

Given an arbitrary set Ω, it is always possible to find a coherent prevision P defined on the set $\mathcal{X}$ of all bounded random variables. However, this existence result, like many others in this domain, relies on the Axiom of Choice. We avail ourselves of the Axiom of Choice whenever it is needed in this paper. Of course, there are many coherent previsions on $\mathcal{X}$, but each of them will be a finitely additive probability when restricted to the collection of indicator functions of subsets of Ω. That is, using the standard notation of letting the name of an event stand for its indicator function, $P(\Omega) = 1$, $P(A \cup B) = P(A) + P(B)$ when $A \cap B = \emptyset$, and $P(A) \geq 0$ for all $A \subseteq \Omega$. By finite additivity and linearity of prevision, the prevision of a simple function (one that assumes only finitely many distinct values) $X = \sum_{i=1}^{n} \alpha_i A_i$ equals $\sum_{i=1}^{n} \alpha_i P(A_i)$. This resembles the formula for the integral of a simple function in the usual measure theoretic derivation. To carry the resemblance further, the values of P on $\mathcal{X}$ are uniquely determined from the finitely additive probability by means of the fact that

$$P(X) = \sup_{\text{simple } Y \leq X} P(Y) = \inf_{\text{simple } Y \geq X} P(Y). \tag{4}$$

The first equation in (4) is the same way that the Lebesgue integral of a nonnegative function X is defined in terms of the integrals of simple functions. Indeed P can be expressed as a finitely additive integral. For each $X \in \mathcal{X}$, we denote $P(X) = \int_{\Omega} X(\omega) P(d\omega)$. Generalizing from the definition of Daniell integral in Royden (1968, Chapter 13), we can call a prevision a finitely additive Daniell integral. Definition 3 applies to both bounded and unbounded functions.

DEFINITION 3. Let $\mathcal{L}$ be a linear space of functions that contains all constants, and let L be a linear functional defined on $\mathcal{L}$ that satisfies $L(X) \geq 0$ whenever $X(\omega) \geq 0$ for all ω. Then L is called a *nonnegative linear functional* or a *finitely additive Daniell integral* over $\mathcal{L}$. We can write $L(X) = \int_{\Omega} X(\omega) L(d\omega)$.

Although the space $\mathcal{L}$ in Definition 3 may contain unbounded functions, L must be finite since it is a linear functional. Extending a finite coherent prevision P on an arbitrary set of random variables $\mathcal{P}$ to a finitely additive Daniell integral is straightforward.

LEMMA 1. *P is a finite coherent prevision on a set $\mathcal{P}$ of random variables if and only if there exists a finitely additive Daniell integral L on a linear space $\mathcal{L}$ that contains $\mathcal{P}$ and all constants such that (i) $L(X) = P(X)$ for every $X \in \mathcal{P}$ and (ii) $L(1) = 1$.*

PROOF. For the "only if" direction, there is no loss of generality in assuming that $\mathcal{P}$ contains all constants and that $P(c) = c$ for each constant c. Next, let $\mathcal{L}$ be the linear span of $\mathcal{P}$, and for each $Y \in \mathcal{L}$, express Y as $\sum_{i=1}^{n} \alpha_i X_i$ with $X_1, \ldots, X_n \in \mathcal{P}$. Define $L(Y) = \sum_{i=1}^{n} \alpha_i P(X_i)$. To see that L is well-defined, assume that Y can be expressed two different ways as

$$Y = \sum_{i=1}^{n} \alpha_i X_i = \sum_{j=1}^{m} \beta_j Y_j.$$

Let $\ell_1 = \sum_{i=1}^{n} \alpha_i P(X_i)$ and $\ell_2 = \sum_{j=1}^{m} \beta_j P(Y_j)$. If $\ell_1 > \ell_2$, then for all ω

$$\sum_{i=1}^{n} \alpha_i [X_i(\omega) - P(X_i)] - \sum_{j=1}^{m} \beta_j [Y_j(\omega) - P(Y_j)] = \ell_2 - \ell_1 < 0.$$

This would make P incoherent, a contradiction. A similar contradiction arises if $\ell_1 < \ell_2$. So L is well-defined. It is straightforward to see that L is linear. To see that $X \geq 0$ implies $L(X) \geq 0$, suppose to the contrary that $L(X) = -\epsilon < 0$. Write $X = \sum_{i=1}^{n} \alpha_i X_i$. Then, for all ω,

$$\sum_{i=1}^{n} (-\alpha_i)[X_i - P(X_i)] = -X + L(X) \leq -\epsilon,$$

which is a contradiction to P being coherent.

For the "if" direction, suppose, to the contrary, that there exist $\epsilon > 0, X_1, \ldots, X_n \in \mathcal{L}$, and $\alpha_1, \ldots, \alpha_n$ such that, for all $\omega \in \Omega$,

$$\sum_{i=1}^{n} \alpha_i [X_i(\omega) - L(X_i)] < -\epsilon. \tag{5}$$

Let Y denote the random variable on the left side of (5), which is in $\mathcal{L}$. Since L is linear, it follows that $L(Y) = 0$ while $L(-\epsilon) = -\epsilon$. This contradicts that L is a nonnegative linear functional. $\square$

In view of Lemma 1, there is no loss of generality in assuming that every finite coherent prevision is defined on a linear space, which may include all bounded random variables, as indicated by the context. For the remainder of this paper, we make that assumption.

Here is an example of a finitely additive Daniell integral that is not countably additive. This example will be used later to illustrate our main result.

EXAMPLE 2. Let Ω be the set of ordered pairs of integers from 1 on up. Let $P(\{(x,y)\}) = 2^{-x-y}$ for all $x, y \geq 1$. Then P is countably additive as a probability on Ω. It follows easily that every bounded random variable X must have prevision $P(X) = \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} X(x,y)2^{-x-y}$. That is, P is countably additive on the set of bounded random variables. Let Y be the unbounded random variable $Y(x,y) = y$ for all x, y. We show now that a necessary condition for $P(Y) = p$ to be incoherent is $p < 2$, hence $P(Y)$ may be coherently chosen to be anything greater than or equal to 2. Suppose that $P(Y) = p$ is incoherent. Then there exists $\epsilon > 0$ an a combination of finitely many gambles with value strictly less that $-\epsilon$, as in (2). It is not difficult to show that the most general combination of finitely many gambles can be expressed, after collecting common terms, in the following form:

$$\beta(Y - p) + \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \alpha_{x,y}[I_{\{(x,y)\}} - 2^{-x-y}], \tag{6}$$

where $\{\alpha_{x,y} : x \geq 1, y \geq 1\}$ is an arbitrary bounded collection of real numbers. If we call the random variable in (6) Z, then $Z(x,y) = \beta(y-p) + \alpha_{x,y} - q$, where

$$q = \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \alpha_{x,y} 2^{-x-y}. \tag{7}$$

We have $Z(x,y) < -\epsilon$ for all x, y if and only if $-\epsilon > \beta(y-p) + \alpha_{x,y} - q$ for all x, y. One necessary condition for these inequalities is $\beta < 0$, because otherwise $\beta(y-p)+\alpha_{x,y}-q$ takes arbitrarily large positive values. Without loss of generality, we can assume that $\beta = -1$. (Just divide all $\alpha_{x,y}$ by $-\beta$ and replace ϵ by $-\epsilon/\beta$.) Since $Z(x,y) < -\epsilon$ for all (x,y), it follows that $\sum_{x=1}^{\infty} \sum_{y=1}^{\infty} Z(x,y) r_{x,y} < -\epsilon$ for every countably additive probability $\{r_{x,y}\}$ over Ω. Let $r_{x,y} = 2^{-x-y}$, and see that

$$-\epsilon > \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} (-y + p + \alpha_{x,y} - q) 2^{-x-y} = -2 + p + q - q = p - 2.$$

That is, $p < 2 - \epsilon$ for some $\epsilon > 0$ is necessary for $P(Y) = p$ to be incoherent. Hence, every $p \geq 2$ is a coherent choice for $P(Y)$. In this example, we choose $P(Y) = 4$, which is coherent. Lemma 1 says that we can extend P to a nonnegative linear functional L on the span $\mathcal{L}$ of Y and the set of all bounded random variables as follows. First note that each element of $\mathcal{L}$ has a unique representation as $\alpha Y + W$ where α is real and W is a bounded random variable. For all real α

and all bounded W, define

$$L(\alpha Y + W) = 4\alpha + \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} W(x,y)2^{-x-y}.$$

This is a coherent prevision on $\mathcal{L}$ according to Lemma 1. This L is not countably additive on $\mathcal{L}$. For each positive integer n, let $A_n = \{(x,y) : 1 \leq x \leq n, 1 \leq y \leq n\}$, and define $X_n(x,y) = Y(x,y)I_{A_n}(x,y)$. Then $\{X_n\}_{n=1}^{\infty}$ is an increasing sequence of nonnegative random variables that converges pointwise to Y. If L were countably additive, then $\lim_{n\to\infty} L(X_n) = L(Y)$. But, $\lim_{n\to\infty} L(X_n) = 2$ while $L(Y) = 4$.

Coherent conditional prevision can be defined in a manner similar to coherent prevision.

DEFINITION 4. Let $\mathcal{Q}$ be a collection of pairs (X,h) where X is a random variable and h is a subset of Ω. a function $P(\cdot|\cdot) : \mathcal{Q} \to \mathbb{R}$ is called a *conditional prevision*. We say that P is *coherent* if, for every finite subset $\{(X_1,h_1),\ldots,(X_n,h_n)\}$ of $\mathcal{Q}$ and scalars $\alpha_1,\ldots,\alpha_n$ $\omega \in \Omega$,

$$\sup_{\omega\in\Omega} \sum_{i=1}^{n} \alpha_i h_i[X_i(\omega) - P(X_i|h_i)] \geq 0. \tag{8}$$

If P is not coherent, we say that P is *incoherent*.

By comparing Definitions 2 and 4, it is easy to see that $P(X)$ must be the same as $P(X|\Omega)$ for every random variable X. This makes coherent prevision the special case of coherent conditional prevision in which, for every pair $(X,h) \in \mathcal{Q}$, we have $h = \Omega$.

When $P(h) = 0$ coherence is insufficient to ensure that $P(X|h)$ has even the most basic intutive properties. For example, if $P(h) = 0$ then $P(h|h) \neq 1$ is coherent. To insure that conditional probabilities behave as much like probabilities as possible, Dubins (1975) introduces an additional assumption. We make a similar assumption that is slightly weaker than that of Dubins (1975) for bounded random variables, but we make our assumption for unbounded random variables also.

ASSUMPTION 1. Let h be a nonempty subset of Ω and let $\mathcal{L}$ be a linear space of random variables that includes h. Then

- if $X \in \mathcal{L}$, then $hX \in \mathcal{L}$,
- $P(\cdot|h)$ is a nonnegative linear functional on $\mathcal{L}$, and
- $P(Xh|h) = P(X|h)$ for all $X \in \mathcal{L}$.

The final clause of Assumption 1 is a formal way of saying that $P(\cdot|h)$ behaves as if h were the entire state space. Nothing that happens outside of h has any impact on $P(\cdot|h)$. It is straghtforward to show that $P(\cdot|h)$ satisfies Assumption 1 if $P(h) > 0$. Hence Example 2 satisfies Assumption 1 for every set h. Because Asssumption 1 places no restrictions on how $P(\cdot|h_1)$ and $P(\cdot|h_2)$ relate to each other when $h_1 \neq h_2$, one can construct $P(\cdot|h)$ separately for each h with $P(h) = 0$ and satisfy Assumption 1 simultaneously for all nonempty h.

3. Conglomerability and disintegrability. We turn now to precise definitions of conglomerability and disintegrability. Let π be a partition of Ω. That is, π is a collection $\{h : h \in \pi\}$ of mutually disjoint subsets of Ω such that their union is Ω. The conditional prevision of each random variable X given each element h of π is denoted $P(X|h)$. In order to make sense out of the loose phrase "the integral of the conditional prevision equals the marginal prevision," we need to be precise about what it means to integrate the conditional prevision.

Finitely additive Daniell integrals behave in many ways like the countably additive Lebesgue integral. One property that the two integrals share is the following transformation property that we use.

LEMMA 2. *Let $\mathcal{L}$ be a linear space of functions on Ω, and let $H : \Omega \to \Theta$ be a function. Let L be a finitely additive Daniell integral over $\mathcal{L}$. Let $\mathcal{W}$ be a linear space of functions on Θ that includes all constants and such that, $Z(H) \in \mathcal{L}$ for every $Z \in \mathcal{W}$. Define $L_H(Z) = L(X)$, where $X = Z(H)$. Then L_H is a finitely additive Daniell integral on $\mathcal{W}$ that we call the integral* induced from L by H.

PROOF. If $Z_1, Z_2 \in \mathcal{W}$ and $\alpha, \beta \in \mathbb{R}$, then

$$\begin{aligned} L_H(\alpha Z_1 + \beta Z_2) &= L[\alpha Z_1(H) + \beta Z_2(H)] = \alpha L[Z_1(H)] + \beta L[Z_2(H)] \\ &= \alpha L_H(Z_1) + \beta L_H(Z_2). \end{aligned}$$

If $Z(\theta) \geq 0$ for all $\theta \in \Theta$, then $X(\omega) = Z(H(\omega)) \geq 0$ for all $\omega \in \Omega$. Hence $L_H(Z) = L(X) \geq 0$. $\square$

In terms of integrals, Lemma 2 says that

$$\int_{\Theta} Z(\theta) L_H(d\theta) = \int_{\Omega} Z(H(\omega)) L(d\omega). \tag{9}$$

Let $H : \Omega \to \pi$ be the function defined by $H(\omega)$ equal to that unique $h \in \pi$ such that $\omega \in h$. Let P_H denote the finitely additive Daniell integral induced from P by H. For each function $Z : \pi \to \mathbb{R}$ defined on π, we use (9) to write

$$\int_{\pi} Z(h) P_H(dh) = \int_{\Omega} Z(H(\omega)) P(d\omega) = P[Z(H)]. \tag{10}$$

For a random variable X such that $P(X|h)$ is finite for all $h \in \pi$, let $P(X|h)$ be $Z(h)$ in (10). Then (10) becomes

$$\int_{\pi} P(X|h) P_H(dh) = \int_{\Omega} P(X|H(\omega)) P(d\omega) = P[P(X|H)]. \tag{11}$$

Then (11) is what we mean by the integral of the conditional prevision.

The conditions stated in the following assumption will be assumed for each random variable and each partition.

ASSUMPTION 2. Let $X : \Omega \to \mathbb{R}$ and let π be a partition of Ω.

1. $P(X)$ is finite,
2. for each $h \in \pi$, $P(X|h)$ is finite, and
3. $P[P(X|H)]$ is finite.

The set $\mathcal{Z}$ of all X that satisfy Assumption 2 in a specific partition π is a linear space and it contains all bounded random variables. Also, if $X \in \mathcal{Z}$ then $P(X|H) \in \mathcal{Z}$. There exist partitions such that all of the random variables in Example 2 satisfy Assumption 2 in those partitions, as we will illustrate in Example 4. There are cases in which the first two bullets of Assumption 2 hold, but the third fails, and our results will not apply to these cases.

DEFINITION 5. Let π be a partition of Ω. Let $\mathcal{W}$ be a collection of random variables defined on Ω such that Assumption 2 holds for π and all $X \in \mathcal{W}$. P is *conglomerable* in π with respect to $\mathcal{W}$ if, for each $X \in \mathcal{W}$

$$\inf_{h \in \pi} P(X|h) \leq P(X) \leq \sup_{h \in \pi} P(X|h). \tag{12}$$

P is *disintegrable* in π with respect to $\mathcal{W}$ if, for each $X \in \mathcal{W}$,

$$\int_{\pi} P(X|h)P_H(dh) = \int_{\Omega} X(\omega)P(d\omega). \tag{13}$$

In view of (11), there is an alternative way to express that P is disintegrable in a partition.

PROPOSITION 1. *P is disintegrable in π with respect to $\mathcal{W}$ if and only if, for each $X \in \mathcal{W}$, $P(X) = P[P(X|H)]$.*

Readers of Dubins (1975) will note that the definition of conglomerable in Definition 5 looks different from the corresponding definition of Dubins. Specifically, Dubins' definition is that P is conglomerable in π with respect to the collection $\mathcal{W}$ if

$$\text{for all } X \in \mathcal{W}, P(X|h) \geq 0 \text{ for all } h \in \pi \text{ implies } P(X) \geq 0. \tag{14}$$

Definition 5 is a straightforward generalization of the definition that de Finetti (1974, p. 143) gives for indicators of events. Definition 5 and (14) are equivalent when $\mathcal{W} = \mathcal{X}$, the collection of all bounded random variables. The proof relies on the fact that $\mathcal{X}$ is a linear space and contains all constants. The two definitions are not necessarily equivalent for every collection that is not a linear space and/or does not contain all constants.

EXAMPLE 3. Consider the same situation as Example 1. Let $\mathcal{W}$ be the collection of all nonnegative bounded random variables. Because each $X \in \mathcal{W}$ is nonnegative, it follows that $P(X|A_i) \geq 0$ for all i and $P(X) \geq 0$. Hence, (14) holds. On the other hand, let $X((j,i)) = j$ for all $j = 0, 1$ and $i = 1, 2, \ldots$. Then $P(X|A_i) = 1$ for all i while $P(X) = 1/2$ and P is not conglomerable by Definition 5.

In order to maintain the spirit of Dubins' definition when $\mathcal{W}$ is not a linear space or does not contain all constants, we need to strengthen (14).

DEFINITION 6. Let $\mathcal{W}$ be a collection of random variables. Let π be a partition, and let P be a coherent prevision on $\mathcal{W}$. We say that P is *D-conglomerable in π with respect to $\mathcal{W}$* if the following is true. For all $X \in \mathcal{W}$ and all real c,

- $P(X|h) \leq c$ for all $h \in \pi$ implies $P(X) \leq c$, and
- $P(X|h) \geq c$ for all $h \in \pi$ implies $P(X) \geq c$.

We now show that Definition 6 is equivalent to conglomerability from Definition 5 when Assumption 2 holds.

LEMMA 3. *Let π be a partition, and let $\mathcal{W}$ be a collection of random variables such that Assumption 2 holds for π and all $X \in \mathcal{W}$. Let P be a coherent prevision on $\mathcal{W}$. Then P is conglomerable in π with respect to $\mathcal{W}$ if and only if P is D-conglomerable in π with respect to $\mathcal{W}$.*

PROOF. For the "if" direction, suppose that P is D-conglomerable in π with respect to $\mathcal{W}$. Let $X \in \mathcal{W}$, and let $c_1 = \inf_{h\in\pi} P(X|h)$ and $c_2 = \sup_{h\in\pi} P(X|h)$. If c_1 is finite, then $P(X|h) \geq c_1$ for all $h \in \pi$ and Definition 6 says that $P(X) \geq c_1$. If c_2 is finite, then $P(X|h) \leq c_2$ for all $h \in \pi$ so that $P(X) \leq c_2$. If $c_1 = -\infty$, then $c_1 \leq P(X)$ is obvious. Similarly, if $c_2 = \infty$, then $P(X) \leq c_2$ is obvious. The cases when $c_1 = \infty$ or $= c_2 = -\infty$ violate Assumption 2.

For the "only if" direction, suppose that P is conglomerable in π with respect to $\mathcal{W}$. Let $X \in \mathcal{W}$. Then Definition 5 implies that

$$\inf_{h\in\pi} P(X|h) \leq P(X) \leq \sup_{h\in\pi} P(X|h).$$

Let $c \in \mathbb{R}$. If $P(X|h) \geq c$ for all $h \in \pi$, then $c \leq \inf_{h\in\pi} P(X|h) \leq P(X)$. Similarly, if $P(X|h) \leq c$ for all $h \in \pi$, then $c \geq \sup_{h\in\pi} P(X|h) \geq P(X)$. □

The following result contains some facts that will be useful in the proof of our main result.

LEMMA 4. *Let $\mathcal{L}$ be a linear space and let P be a coherent prevision on $\mathcal{L}$. Let $h \subseteq \Omega$, and assume that Assumption 1 holds for $\mathcal{L}$ and h. For each $X \in \mathcal{L}$,*

$$\inf_{\omega\in h} X(\omega) \leq P(X|h) \leq \sup_{\omega\in h} X(\omega). \tag{15}$$

Let c be a real constant. If $X(\omega) = c$ for all $\omega \in h$, then

$$P(X|h) = c. \tag{16}$$

PROOF. To prove (15), we prove only the first inequality since the second follows by applying the first equality to $-X$. If $\inf_{\omega\in h} X(\omega) = -\infty$, the first inequality is trivial. If $\inf_{\omega\in h} X(\omega) = \infty$, then X would not be real valued, a contradiction. Suppose, then, that $\inf_{\omega\in h} X(\omega) = c$, a finite real number. Then $X(\omega) - ch(\omega) \geq 0$ for all $\omega \in h$. By Assumption 1, $P(X - ch|h) \geq 0$, from which it follows that

$$P(X|h) \geq P(ch|h) = cP(h|h) = c. \tag{17}$$

Finally, for (16), note that $Xh = ch \in \mathcal{L}$. Then apply the final clause of Assumption 1 together with the last two equalities in (17). □

4. Extending the equivalence of conglomerability and disintegrability to unbounded variables. Lemma 5 shows that, under Assumption 2, disintegrability implies conglomerability for arbitrary collections.

LEMMA 5. *Let π be a partition, and let $\mathcal{W}$ be a collection of random variables such that Assumption 2 holds for π and all $X \in \mathcal{W}$. Let P be a coherent prevision on $\mathcal{W}$ such that P is disintegrable in π with respect to $\mathcal{W}$. Then P is conglomerable in π with respect to $\mathcal{W}$.*

PROOF. Let $H : \Omega \to \pi$ be defined by $H(\omega)$ equal to the unique $h \in \pi$ such that $\omega \in h$. Let $X \in \mathcal{W}$, and define X' by $X'(\omega) = P(X|H(\omega))$. Then we have defined $P[P(X|H)]$ to be $P(X')$. By disintegrability, $P(X') = P(X)$. Because P is coherent,

$$\inf_{\omega\in\Omega} X'(\omega) \le P(X') \le \sup_{\omega\in\Omega} X'(\omega). \tag{18}$$

From the definition of X', we see that

$$\inf_{\omega\in\Omega} X'(\omega) = \inf_{h\in\pi} P(X|h), \text{ and } \sup_{\omega\in\Omega} X'(\omega) = \sup_{h\in\pi} P(X|h). \tag{19}$$

Combining (18) and (19) we get that (12) holds. Since the above argument applies to all $X \in \mathcal{W}$, P is conglomerable in π with respect to $\mathcal{W}$. $\Box$

In light of Lemma 5, for each partition π and each coherent prevision P, every set $\mathcal{W}$ of random variables satisfying Assumption 2 falls into one of three classes.

DEFINITION 7. Let P be a coherent prevision, and let π be a partition. Let $\mathcal{W}$ be a collection of random variables that have previsions under P. We say that

- *$\mathcal{W}$ is of Class 0 relative to P and π* if P is neither conglomerable nor disintegrable in π with respect to $\mathcal{W}$.
- *$\mathcal{W}$ is of Class 1 relative to P and π* if P is conglomerable in π with respect to $\mathcal{W}$ but P is not disintegrable in π with respect to $\mathcal{W}$.
- *$\mathcal{W}$ is of Class 2 relative to P and π* if P is both conglomerable and disintegrable in π with respect to $\mathcal{W}$.

Theorem 1 of Dubins (1975) can be reexpressed as saying that, for each partition π and each coherent prevision P, the class $\mathcal{X}$ of bounded random variables is either of Class 0 or of Class 2 but never of Class 1 relative to P and π. In Section 5, we give an example of a coherent prevision P, a partition π, and a collection $\mathcal{Y}$ of random variables such that $\mathcal{X} \subset \mathcal{Y}$ and $\mathcal{Y}$ is of Class 1 relative to P and π. The following result is a straightforward consequence of the class definitions.

PROPOSITION 2. *Let P be a coherent prevision that satisfies Assumption 1, and let π be a partition. Let $\mathcal{W}$ be a collection of random variables that satisfy Assumption 2. If $\mathcal{W}$ is of Class 0 relative to P and π, then every superset of $\mathcal{W}$ is also of Class 0. If $\mathcal{W}$ is of Class 2 relative to P and π, then every subset of $\mathcal{W}$ is also of Class 2.*

Our extension of Dubins' theorem gives a sufficient condition for a collection $\mathcal{W}$ of random variables to not be of Class 1.

THEOREM 1. *Let P be a coherent prevision that satisfies Assumption 1. Let π be a partition of Ω, and let $H : \Omega \to \pi$ be defined by $H(\omega)$ equal to that unique $h \in \pi$ such that $\omega \in h$. Let $\mathcal{W}$ be a set of real-valued random variables defined on Ω that satisfy Assumption 2. Finally, assume that $\mathcal{W}$ satisfies the following condition:*

$$\text{for every } X \in \mathcal{W},\ X - P(X|H) \in \mathcal{W}. \tag{20}$$

Then, with respect to the collection $\mathcal{W}$, P is conglomerable in π if and only if P is disintegrable in π.

PROOF. Let P be a coherent prevision over the collection $\mathcal{W}$. We show first that P is both conglomerable and disintegrable in the finest partition Ω with respect to $\mathcal{W}$. Let $\pi = \Omega$. To see that P is conglomerable in Ω, let $X \in \mathcal{W}$. Assumption 1 implies that $P(X|\omega) = X(\omega)$ and, by coherence of the unconditional prevision P,

$$\inf_{\omega\in\Omega} X(\omega) \leq P(X) \leq \sup_{\omega\in\Omega} X(\omega).$$

To see that P is disintegrable in Ω, note that $H(\omega) = \{\omega\}$ for all ω, and $P(X|H) = X$ for all $X \in \mathcal{W}$. Hence, we have $P[P(X|H)] = P(X)$, and P is both conglomerable and disintegrable in Ω with respect to $\mathcal{W}$.

Next, let π be a general partition. By (16) in Lemma 4, $P[P(X|H)|h] = P(X|h)$ for all $h \in \pi$ and all $X \in \mathcal{W}$. By linearity of conditional prevision, it follows that, for each $h \in \pi$ and $X \in \mathcal{W}$,

$$P[X - P(X|H)|h] = 0.$$

Hence

$$\inf_{h\in\pi} P[X - P(X|H)|h] = 0 = \sup_{h\in\pi} P[X - P(X|H)|h].$$

We have assumed that $X - P(X|H) \in \mathcal{W}$. If P is conglomerable in π, then $P[X - P(X|H)] = 0$, from which it follows that $P(X) = P[P(X|H)]$, so that P is disintegrable in π.

If P is disintegrable in π then Lemma 5 shows that P is conglomerable in π. $\square$

It is easy to see that the collection $\mathcal{Z}$ of all random variables that satisfy Assumption 2 satisfies the conditions of Theorem 1, and hence is not of Class 1. The key assumption in Theorem 1 is (20). For an arbitrary collection $\mathcal{W}$ that satisfies Assumption 2, define

$$\begin{aligned} \mathcal{W}_- &= \{X - P(X|H) : X \in \mathcal{W}\}, \\ \mathcal{W}_+ &= \mathcal{W} \cup \mathcal{W}_-. \end{aligned}$$

The following results (the second of which is trivial) help to distinguish some collections of random variables by their class.

LEMMA 6. *Let P be a coherent prevision and π a partition. Let H be as defined in Theorem 1. Let $\mathcal{W}$ be a collection of random variables that satisfy Assumption 2. Then*

1. *$\mathcal{W}_+$ satisfies (20),*
2. *$\mathcal{W}$ is of Class 2 relative to P and π if and only if $\mathcal{W}_+$ is also of Class 2, and*
3. *if $\mathcal{W}$ is not of Class 2 relative to P and π, then $\mathcal{W}_+$ is of Class 0.*

PROOF. For part (1), let $X \in \mathcal{W}$ so that $X - P(X|H) \in \mathcal{W}_+$. Also $P[X - P(X|H)|H]$ is identically 0, hence

$$X - P(X|H) - P[X - P(X|H)] \in \mathcal{W}_+.$$

For part (2), the "if" direction is immediate from Proposition 2. For the "only if" direction, note that for every $Y \in \mathcal{W}_-$, $P(Y|H)$ is identically 0 and $P(Y) = 0$ if $\mathcal{W}$ is of Class 2. For part (3), Theorem 1 says that $\mathcal{W}_+$ is either of Class 0 or Class 2. If $\mathcal{W}$ is not of Class 2, then no superset of it, such as $\mathcal{W}_+$, can be of Class 2. Hence $\mathcal{W}_+$ must be of Class 0. $\square$

PROPOSITION 3. *If P is a countably additive prevision on a class $\mathcal{W}$ of random variables and every element of π has positive probability, then $\mathcal{W}$ is of Class 2 relative to P and π.*

One subtle point concerning Proposition 3 is that P can be a countably additive prevision on the collection of all bounded random variables but fail to be countably additive on a collection that includes unbounded random variables. The following example (and the more complicated example in Section 5) illustrates this circumstance. Example 4 also contains examples of both Class 0 and Class 2 in the same distribution but two different partitions.

EXAMPLE 4. Return to the situation described in Example 2. Consider the following two partitions:

$$\begin{aligned} \pi_1 &= \{h_x : x \geq 1\}, \\ \pi_2 &= \{g_y : y \geq 1\}, \end{aligned}$$

where $h_x = \{(x, y) : y \geq 1\}$ for each $x \geq 1$, and $g_y = \{(x, y) : x \geq 1\}$ for each $y \geq 1$. It is straightforward to show that $P(h_x) = 2^{-x}$ for all $x \geq 1$, and $P(g_y) = 2^{-y}$ for all $y \geq 1$. Hence, for all x and y, $P(g_y|h_x) = 2^{-y}$ and $P(h_x|g_y) = 2^{-x}$. For each bounded random variable X

$$\begin{aligned} P(X|h_x) &= \sum_{y=1}^{\infty} X(x, y)2^{-y}, \text{ for all } x, \\ P(X|g_y) &= \sum_{x=1}^{\infty} X(x, y)2^{-x}, \text{ for all } y. \end{aligned}$$

It follows easily that both (12) and (13) hold for each bounded X and for both π_1 and π_2. Whether P is conglomerable and/or disintegrable in π_1 or π_2 depends on which of (12) and/or (13) holds for $X = Y$.

Consider first π_1. For each x, Yh_x is unbounded. We have not yet assigned a prevision to Yh_x. An argument similar to the one given in Example 2 shows that $P(Yh_x)$ must be assigned a value greater than or equal to 2^{-x+1}, hence $P(Y|h_x) \geq 2$ is required for coherence. To see that $P(Y|h_x) = 2$ (equivalently $P(Yh_x) = 2^{-x+1}$) for all x is a coherent assignment, we argue as in Example 2. The most general gamble available can be rewritten as follows.

$$\beta(Y - 4) + \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} \alpha_{x,y}[I_{\{(x,y)\}} - 2^{-x-y}] + \sum_{i=1}^{n} \gamma_i \left(\left[\sum_{y=1}^{\infty} yI_{\{x_i,y\}} \right] - 2^{-x_i+1} \right), \tag{21}$$

where $x_i, \ldots, x_n \geq 1$ are distinct integers. As before $\beta < 0$ is necessary for (21) to be uniformly negative. We will let $\beta = -1$. As before, let q be as in (7). The value of (21) when $\omega = (x, y)$ is

$$Z(x, y) = -y + 4 + \alpha_{x,y} - q + \sum_{i=1}^{n} \gamma_i[yI_{\{x_i\}}(x) - 2^{-x_i+1}].$$

A necessary condition for (21) to be uniformly negative is that, for some $\epsilon > 0$,

$$-\epsilon > \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} Z(x,y)2^{-x-y} = 2 + q - q + \sum_{i=1}^{n} 2\gamma_i[2^{-x_i} - 2^{-x_i}] = 2,$$

which is impossible. Hence, we can coherently assign $P(Y|h_x) = 2$ for all x. If we make these assignments, then (12) fails in π_1 because

$$\inf_{h\in\pi_1} P(Y|h) = 2 = \sup_{h\in\pi_1} P(Y|h),$$

while $P(Y) = 4$. Also, (13) fails because

$$\int_{\pi_1} P(Y|h)P_H(dh) = 2 \neq \int_{\Omega} Y(\omega)P(d\omega) = 4.$$

We see that P is neither disintegrable nor conglomerable in π_1.

Next, consider π_2. For each y, $Yg_y = yg_y$. Hence, $P(Y|g_y) = y$. It follows that (13) holds with $X = Y$ because both sides equal $P(Y) = 4$. Also, (12) holds because

$$\inf_{h\in\pi_2} P(Y|h) = 1 < 4 = P(Y) < \infty = \sup_{h\in\pi_2} P(Y|h).$$

We see that P is both disintegrable and conglomerable in π_2.

5. A conglomerable example that is not disintegrable. The following construction is a modification of an example given in Dubins (1975, Theorem 2). Let $\Omega = \{(i,k) : i = 1,2,\ldots, k = i, i+1,\ldots\}$ be the set of ordered pairs of positive integers in which the first coordinate is greater than or equal to the second. We use the algebra $\mathcal{A} = 2^{\Omega}$. Define a countably additive probability P on Ω by $P(\{(i,k)\}) = 2^{-k}$ for each $(i,k) \in \Omega$. It follows that $\emptyset$ is the only subset of Ω with 0 probability. Since P is countably additive over $\mathcal{A}$, it is also countably additive on the set $\mathcal{X}$ of bounded random variables defined on Ω. By Proposition 3, $\mathcal{X}$ is of Class 2 relative to P and every partition π that we choose to consider.

Next, we extend P to a collection $\mathcal{Y}$ of random variables that are bounded below in such a way that $P(|Y|) < \infty$ for each $Y \in \mathcal{Y}$. Part of the extension relies on the expectation operator

$$E(Y) = \sum_{(i,k)\in\Omega} P(\{(i,k)\})Y((i,k)),$$

which is well-defined for all Y that are bounded below. Also, $E(X) = P(X)$ for all $X \in \mathcal{X}$. As de Finetti (1974, Section 3.12) showed, if Y is unbounded above, it is possible that $P(Y) > E(Y)$. Indeed, unless there exists at least one unbounded Y whose prevision differs from its expectation, P will be countably additive on the collection of all random variables with finite prevision and hence will be both conglomerable and disintegrable in every partition. We call the difference $\beta(Y) = P(Y) - E(Y)$ the *boost function*. (See Seidenfeld, Schervish and Kadane, 2009 for a more detail about the boost function.) We have that $\beta(X) = 0$ for every bounded X, and β is a linear functional that is nonnegative for every random variable bounded below.

We begin by extending P to a particular unbounded random variable W, and then to a linear space $\mathcal{L}$ including W and all of $\mathcal{X}$. We then let $\mathcal{Y}$ be the subcollection of $\mathcal{L}$ consisting of the random variables that are bounded below. The starting random variable is $W(i, k) = k$. We set $P(W) = 14 = E(W) + 10$, so that $\beta(W) = 10$. In order to find a partition in which P is conglomerable but not disintegrable, we need to extend P beyond the linear span of $\mathcal{X}$ and $\{W\}$. We do this by defining β on a larger collection of random variables.

DEFINITION 8. Let Ω be a non-empty set. A collection p of subsets of Ω is called an *ultrafilter* if the following conditions hold:

- For every subset $A \subseteq \Omega$, either $A \in p$ or $A^C \in p$, but not both.
- If $A, B \in p$, then $A \cap B \in p$.
- If $A \in p$ and $A \subseteq B$, then $B \in p$.

The simplest ultrafilters are the *principal* ultrafilters that consist of all subsets that contain a specific element of Ω. All other ultrafilters are called *non-principal*. A proof of the existence of non-principal ultrafilters can be found in Comfort and Negrepontis (1974). The following fact about non-principal ultrafilters is useful.

LEMMA 7. *Let Ω be an infinite set, and let p be a non-principal ultrafilter. Then every element of p has infinitely many elements. In particular, the complement of every finite set is in p.*

PROOF. Let $A \in p$. Suppose, to the contrary, that A has finitely many elements. Split A into two nonempty subsets $A_1 \cup A_2 = A$. Then either A_1 or $A_1^C \in p$. If $A_1^C \in p$, then $A_2 = A_1^C \cap A \in p$. So either A_1 or A_2 is in p. Repeat this exercise with the set that is in p until one arrives at a set in p with exactly one element. This would make p a principal ultrafilter. To see that the complement of every finite set A is in p, suppose to the contrary that $A^C \notin p$. Then $A \in p$, which contradicts what we have already proved. □

Let p and q be non-principal ultrafilters on the positive integers, and construct the ultrafilter product $q \cdot p$ on Ω as follows. For each $S \subseteq \Omega$, define $S_i = \{k : (i,k) \in S\}$. We say that $S \in q \cdot p$ if $\{i : S_i \in p\} \in q$. The following result follows easily from elementary properties of non-principal ultrafilters. (See Comfort and Negrepontis, 1974, p. 157.)

LEMMA 8. *A necessary (but not sufficient) condition for $S \in q \cdot p$ is that there exist infinitely many i such that $(i,k) \in S$ for infinitely many k. A sufficient (but not necessary) condition for $S \in q \cdot p$ is that for all but finitely many i $(i,k) \in S$ for all but finitely many k. In particular, if the supremum over k of the cardinality of $\{i : (i,k) \in S\}$ is finite, then $S \notin q \cdot p$.*

PROOF. For the necessary condition, assume that $S \in q \cdot p$, so that $T = \{i : S_i \in p\} \in q$. It follows from Lemma 7 that T has infinitely many elements. Also, for each $i \in T$, S_i has infinitely many elements according to Lemma 7 because $S_i \in p$. For the sufficient condition, assume that for all but finitely many i $(i,k) \in S$ for all but finitely many k. Let $T = \{i : S_i \text{ is the complement of a finite set}\}$. Lemma 7 says that $S_i \in p$ for each $i \in T$. Also, $T \in p$ because T is the complement of a finite set. This implies $S \in q \cdot p$. For the final claim, assume to the contrary that $S \in q \cdot p$. Let n be the supremum over k of the cardinality of $T_k = \{i : (i,k) \in S\}$. Because $S_1, \ldots, S_{n+1} \in p$, $T = \cap_{i=1}^{n+1} S_i \in p$. Let $k \in T$. Then $k \in S_i$ for $i = 1, \ldots, n+1$. But then the cardinality of T_k is at least $n+1$, a contradiction. □

We use Lemma 8 to determine which of the following three sets is in the ultrafilter $q \cdot p$, as these sets are key to the main example of this section:

$$L = \{(i,k) : k > 2i\},\ U = \{(i,k) : k < 2i\},\ \text{and } Q = \{(i,k) : k = 2i\}. \quad (22)$$

Then $L \in q \cdot p$ according to the sufficient condition in Lemma 8, while $U \notin q \cdot p$ according to the necessary condition. Also, $Q \notin q \cdot p$, according to the last result in Lemma 8.

We now make use of the ultrafilter $q \cdot p$ to extend β to the collection of all random variables of the form WS where S is the indicator function of an arbitrary subset $S \subseteq \Omega$. Set $\beta(WS) = 10$ if $S \in q \cdot p$ and $\beta(WS) = 0$ otherwise. The prevision P now extends easily to the linear span $\mathcal{L}$ of all of the random variables for which β has been defined. The set $\mathcal{L}$ consists of all random variables of the form

$$Y = X + \alpha W S, \tag{23}$$

where X is bounded, α is real, and S is an indicator of a subset of Ω. Define $\mathcal{Y}$ to be the set of all $Y \in \mathcal{L}$ for which $\alpha \geq 0$ in (23). Because every nonempty event has positive probability, if $Y \in \mathcal{Y}$, then $P(Y|h)$ is finite for every $h \subseteq \Omega$. The following result follows easily from the linearity of β.

PROPOSITION 4. *If $Z \in \mathcal{Y}$ and $\{Z \leq cW\} \in q \cdot p$, then $\beta(Z) \leq 10c$. Hence, if for all $c > 0$ $\{Z > cW\} \notin q \cdot p$, then $\beta(Z) = 0$.*

Next, partition Ω by the values of W. That is, $\pi_W = \{h_k\}_{k=2}^{\infty}$ where $h_k = \{W = k\}$. Each h_k consists of exactly $k-1$ points, i.e., $h_k = \{(k,1), \ldots, (k, k-1)\}$ for $k = 2, 3, \ldots$. The conditional distribution given h_k is uniform over those $k-1$ points.

THEOREM 2. *With respect to the collection $\mathcal{Y}$, previsions are conglomerable in π_W but they are not disintegrable in π_W.*

PROOF. First we show that P is not disintegrable in π_W. Let L, U, and Q be as defined in (22). For each k, $U \cap h_k$ and $L \cap h_k$ have the same number of elements and all elements of h_k have the same probability. Hence,

$$P(U) = P(L) = \frac{1}{2}[1 - P(Q)] > 0.$$

Because $L \in q \cdot p$ and $U \notin q \cdot p$, $P(LW) = P(UW) + 10$. Because each $h \in \pi_W$ is a finite set, $P(Z|h) = E(Z|h)$ for each $h \in \pi_W$. Hence, for $k = 2, 3, \ldots$,

$$P(UW|h_k) = P(LW|h_k) = \begin{cases} \frac{k}{2} & \text{if } k \text{ is odd,} \\ \frac{k(k-2)}{2(k-1)} & \text{if } k \text{ is even.} \end{cases}$$

Hence,

$$P[P(UW|H)] = P[P(LW|H)],$$

but $P(UW) \neq P(LW)$, and P is not disintegrable in π_W.

Next, we show that with respect to variables in $\mathcal{Y}$, P is conglomerable in π_W. Let $Z \in \mathcal{Y}$. Recall that $P(Z) = E(Z) + \beta(Z)$, and $\beta(Z) \geq 0$. We know that

$$\inf_{h \in \pi_W} P(Z|h) = \inf_k E(Z|h_k) \leq E(Z) \leq P(Z).$$

What remains is to show that $P(Z) \leq \sup_{h \in \pi_W} P(Z|h)$.

If $\sup_{h \in \pi_W} P(Z|h) = \infty$, the proof is complete. So, assume that

$$\sup_{h \in \pi_W} P(Z|h) = r < \infty.$$

Hence $\sup_k E(Z|h_k) = r$ and $E(Z) \leq r$. Since $Z \in \mathcal{Y}$, there exists $b > -\infty$ such that $b \leq Z(i,k)$ for all i, k. Let $Z' = Z - b$ so that $Z' \geq 0$ and $\beta(Z') = \beta(Z)$. The conditional previsions $\{P(Z'|h_k)\}_{k=2}^{\infty}$ are bounded above by $r - b$ and below by 0. Let $d > 0$. The Markov inequality says that

$$P(Z' > dW|h_k) = P(Z' > dk|h_k) \leq \frac{r-b}{dk}.$$

Recall that the conditional distribution $P(\cdot|h_k)$ is uniform over the $k-1$ points in h_k. Hence, for each $d > 0$ and all k, at most $(r-b)/d$ out of the $k-1$ points in h_k may satisfy $Z'(i,k) > dW(i,k)$. That is, the event $\{Z' > dW\} \cap h_k$ contains of at most $(r-b)/d$ points for each k. By the last result in Lemma 8, we have $\{Z' > dW\} \notin q \cdot p$. Proposition 4 now says that $\beta(Z') = 0$, hence $\beta(Z) = 0$. So $P(Z) = E(Z) \leq r$, as required by conglomerability. $\square$

6. Discussion. Conglomerability and disintegrability are familiar concepts in the countably additive theory of probability, although the names may not be as familiar as the concepts. The law of total probability or "tower property" of conditional expectations is essentially disintegrability, namely that the mean of a conditional mean is the marginal mean. With disintegrability taken for granted, conglomerability is simply an instance of the property of countably additive expectations that the mean of a random variable lies in the closed convex hull of its range. Of course, the countably additive theory guarantees disintegrability by allowing the conditional probabilities of events to change with the partition on which one conditions. The well-known Borel paradox is a classic example of how this happens. In the countably additive theory Kadane, Schervish and Seidenfeld (1996) illustrates how pervasive the Borel paradox is. If one insists on $P(X|h)$ having a meaning for every random variable X and every nonempty event h, then not even the countably additive theory can guarantee disintegrability in every partition.

The finitely additive theory of probability avoids the Borel paradox, but at the price of having its conditional probabilities fail conglomerability. Dubins (1975) shows that for coherent previsions over a linear space of bounded random variables, conglomerability in a partition is equivalent to disintegrability in that same partition. In this paper we extend the concept of a coherent prevision to a class of unbounded random variables with finite previsions and finite conditional previsions. We show that a finitely additive version of the Daniell integral gives an extension of disintegration to this class of unbounded variables. When coherent previsions are defined for a linear space of such random variables, conglomerability and disintegrability of previsions are equivalent conditions.

As a final note, it is important to keep in mind that the concepts of conglomerability and disintegrability are defined with respect to a collection of random variables. The larger the collection of random variables, the more conditions of the form (12) and (13) that each concept requires. That is, in order for P to be conglomerable in π with respect to a collection $\mathcal{W}$, (12) must hold for every $X \in \mathcal{W}$. Similarly, for P to be disintegrable in π with respect to $\mathcal{W}$, (13) must hold for every $X \in \mathcal{W}$. Consider the three collections $\mathcal{X} \subset \mathcal{Y} \subset \mathcal{Z}$ that figure in the results of this paper. That is, $\mathcal{X}$ is the collection of all bounded random variables, $\mathcal{Z}$ is the collection of all random variables that satisfy Assumption 2, and $\mathcal{Y}$ is an intermediate collection such as the collection in Section 5. If P is conglomerable in π with respect to $\mathcal{Z}$, then $\mathcal{Z}$ is of Class 2 relative to P and π and so are $\mathcal{Y}$ and $\mathcal{X}$. Similarly, if P is disintegrable in $\mathcal{Z}$ with respect to π, then all three collections are of Class 2. However, the equivalence of conglomerability and disintegrability does not carry over from larger collections to smaller collections. The reason is that $\mathcal{Z}$ might be of Class 0 while $\mathcal{Y}$ is of Class 1 and $\mathcal{X}$ is of Class 2. Indeed, this is precisely what occurs in the example of Section 5.

REFERENCES

COMFORT, W. W. and NEGREPONTIS, S. (1974). *The Theory of Ultrafilters*, Springer-Verlag, Berlin.

CRISMA, L., GIGANTE, P. and MILLOSSOVICH, P. (1997). A notion of coherent prevision for arbitrary random quantities. *Journal of the Italian Statistical Society* **6** 233–243.

DAWID, A., STONE, M. and ZIDEK, J. (1973). Marginalization paradoxes in bayesian and structural inference. *Journal of the Royal Statistical Society B* **35** 189–233.

DE FINETTI, B. (1930). Sulla preprieta conglomerativa della probabilita subordinate. *Rend. R. Inst. Lombardo (Milano)* **63** 414–418.

DE FINETTI, B. (1974). *Theory of Probability (2 vols.)*, Wiley, New York.

DUBINS, L. (1975). Finitely additive conditional probabilites, conglomerability and disintegrations. *Annals of Probability* **3** 89–99.

KADANE, J., SCHERVISH, M. and SEIDENFELD, T. (1986). Statistical implications of finitely additive probability. In *Bayesian Inference and Decision Techniques* (P. Goel and A. Zellner, eds.) 59–76. North-Holland, Amsterdam.

KADANE, J. B., SCHERVISH, M. J. and SEIDENFELD, T. (1996). Reasoning to a foregone conclusion. *Journal of the American Statistical Association* 91(435) 1228–1235.

KOLMOGOROV, A. (1956). *Foundations of the Theory of Probability*, Chelsea, New York.

KRAUSS, P. H. (1968). Representation of conditional probability measures on boolean algebras. *Acta Math. Sci. Hungar* **19** 229–241.

ROYDEN, H. L. (1968). *Real Analysis*, Macmillan, London.

SCHERVISH, J. J., SEIDENFELD, T. and KADANE, J. B. (1984). The extent of non-conglomerability of finitely additive probabilities. *Z. f. Warscheinlichkeitstheorie* **66** 205–226.

SEIDENFELD, T., SCHERVISH, M. J. and KADANE, J. B. (2009). Preference for equivalent random variables: A price for unbounded utilities. *Journal of Mathematical Economics* **45** 329–340.

MARK J. SCHERVISH
JOSEPH B. KADANE
DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
E-MAIL: mark@stat.cmu.edu,
kadane@stat.cmu.edu

TEDDY SEIDENFELD
DEPARTMENT OF PHILOSOPHY AND
DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
E-MAIL: teddy@stat.cmu.edu