Non-conglomerability for countably additive measures that are not κ-additive*

T.Seidenfeld, M.J.Schervish, and J.B.Kadane

Abstract

Let κ be a successor cardinal. Using the theory of coherent conditional probability associated with de Finetti (1974) and Dubins (1975), we show that each probability that is not κ-additive (but is λ-additive if λ < κ) has coherent conditional probabilities that fail to be conglomerable in a partition of cardinality κ. This generalizes our (1984) result, where we established that each finite but not countably additive probability has coherent conditional probabililties that fail to be conglomerable in some countable partition.

Key Words: κ-additive probability, non-conglomerability, coherent conditional probability, regular conditional probability distribution, descendingly incomplete ultrafilters.

1. Introduction. Consider a finitely, but not necessarily countably additive probability P(·) defined on a sigma-field of sets $\mathcal{B}$. Let B, C, D, E, F, G ∈ $\mathcal{B}$, with B ≠ ∅ and F ∩ G ≠ ∅.

Definition. A *coherent conditional probability* function P(· | B) satisfies the following three conditions:

(i) P(C ∪ D | B) = P(C | B) + P(D | B), whenever C ∩ D = ∅;

(ii) P(B | B) = 1

Moreover, following de Finetti (1974) and Dubins (1975), in order to regulate conditional probability given a non-empty event of unconditional or conditional probability 0, we require the following.

(iii) P(E ∩ F | G) = P(E | F ∩ G)P(F | G).

This account of coherent conditional probability is not the usual theory from contemporary Mathematical Probability. It differs from the received theory of Kolmogorovian *regular conditional distributions* in four ways:

1. The theory of regular conditional distributions requires that probabilities and conditional probabilities are countably additive. The de Finetti/Dubins theory of coherent conditional probability require only that probability is finitely additive. In this paper, we bypass this difference by exploring countably additive coherent conditional probabilities.

2. When P(B) = 0 and B is not empty, a regular conditional probability given B is relative also to a sub-sigma field $\mathcal{A} \subseteq \mathcal{B}$, where B ∈ $\mathcal{A}$. In the theory of coherent conditional probability, P(· | B), depends solely on the event B and not on any sub-field that embeds it. Example 2, which we present in Section 4 after *Lemma* 3, illustrates this difference.

3. Some countably additive probabilities do not admit regular conditional distributions relative to a particular sub-sigma field, even when both sigma-fields are countably generated. (See Corollary 1 in our [2001].) In contrast, Dubins (1975) establishes the existence of *full* coherent conditional probability functions: where, given a set **W** of arbitrary cardinality, a coherent conditional probability is defined with respect to each non-empty element of its powerset, i.e., $\mathcal{B}$ is the powerset of **W**. Hereafter, we require that each probability function includes its coherent conditional probabilities given each non-empty event $B \in \mathcal{B}$. However, we do not require that $\mathcal{B}$ is the powerset of the state-space for P.

4. Our focus in this paper is a fourth feature that distinguishes the de Finetti/Dubins theory of coherent conditional probability and the Kolmogorovian theory of regular conditional probability. This aspect of the difference involves *conglomerability* of conditional probability functions.

Let $E \in \mathcal{B}$, let $\boldsymbol{N}$ be an index set and let $\pi = \{h_\nu: \nu \in \boldsymbol{N}\}$ be a partition of the sure event where the conditional probabilities, $P(E \mid h_\nu)$, are well defined for each $\nu \in \boldsymbol{N}$.

Definition: The conditional probabilities $P(E \mid h_\nu)$ are *conglomerable* in π provided that, for each event $E \in \mathcal{B}$ and arbitrary real constants k_1 and k_2,

$$\text{if } k_1 \leq P(E \mid h_\nu) \leq k_2 \text{ for each } \nu \in \mathbf{N}, \text{ then } k_1 \leq P(E) \leq k_2.$$

In our (1984) we show that if P is merely finitely additive (i.e., if P is finitely but not countably additive) with coherent conditional probabilities, then P fails conglomerability in some countable partition. That is, for each merely finitely additive probability P there is an event E, an $\varepsilon > 0$, and a countable partition $\pi = \{h_n: n = 1, \ldots\}$, where $P(E) > P(E \mid h_n) + \varepsilon$ for each $h_n \in \pi$.

The following illustrates a failure of conglomerability for a merely finitely additive probability P in a countable partition $\pi = \{h_n: n \in \{1, 2, \ldots\}\}$, where each element of the partition is not null, i.e., $P(h_n) > 0$ for each $n \in \{1, 2, \ldots\}$. Then, by both theories of conditional probability, $P(E \mid h_n) = P(E \cap h_n)/P(h_n)$ is well defined. Thus, the failure of conglomerability in this example is due to the failure of countable additivity, rather than to a difference in how conditional probability is defined.

***Example* 1** (Dubins, 1975): Let the sure event $\mathbf{W} = \{(i, n): i \in \{1, 2\} \text{ and } n \in \{1, 2, \ldots\}\}$ and $\mathcal{B}$ be the powerset of **W**. Let $E = \{\{1, n\}: n \in \{1, 2, \ldots\}\}$ and $h_n = \{\{1,n\}, \{2, n\}\}$, and partition $\pi = \{h_n: n \in \{1, 2, \ldots\}\}$. Define $P(\{i, n\}) = 1/2^{n+1}$ if $i = 1$, $P(\{i, n\}) = 0$ if $i = 2$, and $P(E) = 0.5$. So P is merely finitely additive over E^c. Hence, $P(h_n) = 1/2^{n+1} > 0$ for each $n \in \{1, 2, \ldots\}$. Then P is not conglomerable in π as:

$P(E^c \mid h_n) = P(E^c \cap h_n)/P(h_n) = 0$, for each $n \in \{1, 2, \ldots\}$, whereas $P(E^c) = 0.5$. ◊ Example 1

In our [1996], we discuss this example in connection with the value of information.

In the appendix to our (1986) we show that for a continuous, countably additive probability defined on the continuum, and assuming coherent conditional probabilities rather than regular conditional distributions, then non-conglomerability results by considering continuum-many different partitions of the continuum. These alternative partitions are generated by sets of equivalent (non-linearly transformed) random variables. Conglomerability cannot be satisfied in all the partitions. Here we generalize that result to κ-additive probabilities.

In the following definition, let α, β, and γ be ordinals and κ a cardinal.
Definition: A probability P is κ-additive if, for each increasing γ-sequence of measurable events $\{E_\alpha: \alpha < \gamma \leq \kappa\}$, where $E_\alpha \subseteq E_\beta$ whenever $\alpha < \beta < \gamma$, then

$$P(\cup_{\alpha<\gamma} E_\alpha) = \sup_{\alpha<\gamma} P(E_\alpha).$$

That is, with $\gamma \leq \kappa$, P is κ-additive provided that probability is continuous from below over γ-long sequences that approximate events from below. This agrees with the usual definition of countable additivity; let $\kappa = \aleph_0$.

Say that P is not κ-additive when, for some event E and increasing γ-sequence that approximates E from below, $P(\cup_{\alpha<\gamma} E_\alpha) > \sup_{\alpha<\gamma} P(E_\alpha)$.
If P is κ-additive for each cardinal κ, then call P *perfectly additive*.

Consider a countably additive probability P that is not κ-additive for some successor cardinal $\kappa = \lambda^+$. Here we show (in Section 4) the main *Proposition* of this paper:

- P fails to be conglomerable in some partition of cardinality κ.

Rather than thinking that non-conglomerability is an anomalous feature of finite but not countably additive probabilities, and arises solely with finitely but not countably additive probabilities in countable partitions, here we argue for a different conclusion: Let P be a coherent probability. Non-conglomerability of its coherent conditional probabilities $\{P(E \mid h_\nu): \nu \in \boldsymbol{N}\}$ occurs in a partition $\pi = \{h_\nu: \nu \in \boldsymbol{N}\}$ whose cardinality $|\pi| = \kappa$ matches the κ-non-additivity of P.

2. Other structural assumptions for the *Proposition*. Since the cardinals below a given cardinal form a well-ordered set, we consider the least cardinal κ for which P is not κ-additive. And since we assume that P is countably additive, then κ is some uncountable cardinal – unless P is perfectly additive. Thus, assume that for an uncountable cardinal κ, P is not κ-additive but is λ-additive for each cardinal λ < κ. Also, we assume that P includes its coherent conditional probability distributions and these, too, are λ-additive for each λ < κ.

Moreover, we take the measure completion of P, so that each subset of a P-null event is measurable. That is, if E is measurable with P(E) = 0, then each subset of E also is measurable. This assumption provides for a rich space of measurable events while stopping short of requiring P to be defined on a powerset, which otherwise would require κ to be greater than a weakly inaccessible cardinal, by Ulam's [1930] result.

Under these assumptions, let P be defined on a measurable space $<\mathbf{W}, \mathcal{B}>$, where $\mathcal{B}$ includes each of the points of the space, $\mathbf{W} = \{w_\alpha: \alpha < \kappa\}$, with α ranging over all ordinals less than κ. That is, without loss of generality, assume **W** has cardinality κ and where if a measurable event E is null, i.e., whenever P(E) = 0, then $\mathcal{B}$ includes each subset of E.

Since P is not perfectly additive, it follows that κ is a regular cardinal: it has cofinality κ. Otherwise, κ is singular with cofinality(κ) = $\lambda < \kappa$. Then, using this λ-sequence which is cofinal in κ, as P is λ-additive for each $\lambda < \kappa$, P would be κ-additive as well. In addition, for the proof of Lemma 4, below, we assume that κ is not inaccessible, i.e., we avoid the case that κ is a regular limit cardinal, whose existence is independent of ZFC. We make one additional structural assumption on $\mathcal{B}$ that depends upon a linear order $\uparrow$ over special sets of points (called *tiers*) that is defined in Section 3.

3. *Tiers* of points. The proof of the main *Proposition* is based on the structure of a linear order over equivalence classes (called *tiers*) defined by the following relation between pairs of points in **W**.

Definition: Consider the relation, ~, of relative-non-nullity on pairs of points in **W**. That is, for two points, $w_\alpha \neq w_\beta$, they bear the relation $w_\alpha \sim w_\beta$ provided that

$$0 < P(\{w_\alpha\} | \{w_\alpha, w_\beta\}) < 1.$$

We make ~ into an equivalence relation by stipulating that, for each point w, $w \sim w$.

Lemma 1: ~ is an equivalence relation.

Proof: Only transitivity requires verification. Assume $w_1 \sim w_2 \sim w_3$. That is, assume $0 < P(\{w_1\} | \{w_1, w_2\}), P(\{w_2\} | \{w_2, w_3\}) < 1$. Then by condition (iii) of coherent conditional probability:

$P(\{w_1\} | \{w_1, w_2, w_3\}) = P(\{w_1\} | \{w_1, w_2\})\, P(\{w_1, w_2\} | \{w_1, w_2, w_3\})$. Similarly,
$P(\{w_3\} | \{w_1, w_2, w_3\}) = P(\{w_3\} | \{w_2, w_3\})\, P(\{w_2, w_3\} | \{w_1, w_2, w_3\})$.

Now argue indirectly by cases.

- If $P(\{w_1\} | \{w_1, w_3\}) = 0$, then $P(\{w_1\} | \{w_1, w_2, w_3\}) = 0$ and $P(\{w_1, w_2\} | \{w_1, w_2, w_3\}) = 0$, since by assumption $P(\{w_1\} | \{w_1, w_2\}) > 0$. Then $P(\{w_2\} | \{w_1, w_2, w_3\}) = 0 = P(\{w_2\} | \{w_2, w_3\})$, which contradicts $w_2 \sim w_3$.
- If $P(\{w_1\} | \{w_1, w_3\}) = 1$, then $0 = P(\{w_3\} | \{w_1, w_3\}) = P(\{w_3\} | \{w_1, w_2, w_3\})$. Then $0 = P(\{w_2, w_3\} | \{w_1, w_2, w_3\})$, since $0 < P(\{w_3\} | \{w_2, w_3\})$. So, $0 = P(\{w_2\} | \{w_1, w_2, w_3\}) = P(\{w_2\} | \{w_1, w_2\})$, which contradicts $w_1 \sim w_2$.

Hence $0 < P(\{w_1\} | \{w_1, w_3\}) < 1$, as required. $\Diamond_{Lemma\ 1}$

Definition: The equivalence relation ~ partitions **W** into disjoint *tiers* τ of relative non-null pairs of points.
For each pair of points $\{w_1, w_2\}$ that belong to different tiers, $w_i \in \tau_i$ (i = 1, 2), $\tau_1 \neq \tau_2$, then $P(\{w_1\} \mid \{w_1, w_2\}) \in \{0,1\}$.

If $P(\{w_2\} \mid \{w_1, w_2\}) = P(\{w_3\} \mid \{w_2, w_3\}) = 1$, then $P(\{w_3\} \mid \{w_1, w_3\}) = 1$. Thus, the tiers are linearly ordered by the relation $\uparrow$, defined as:
Definition: $\tau_1 \uparrow \tau_2$ if for each pair $\{w_1, w_2\}$, $w_i \in \tau_i$ (i = 1, 2), $P(\{w_2\} \mid \{w_1, w_2\}) = 1$.
Since the reverse ordering also is linear, we express this as:
Definition: $\tau_2 \downarrow \tau_1$ if for each pair $\{w_1, w_2\}$, $w_i \in \tau_i$ (i = 1, 2), $P(\{w_2\} \mid \{w_1, w_2\}) = 1$, i.e., if and only if $\tau_1 \uparrow \tau_2$.

As a final structural assumption, we assume that each tier, τ, belongs to the algebra $\mathcal{B}$, and that the set of tiers below (or above) a tier in the linear order also belong to $\mathcal{B}$, i.e., the "intervals" $\{\tau' : \tau' \downarrow \tau\}$ and $\{\tau' : \tau' \uparrow \tau\}$ are measurable as well.

4. The Main Proposition.

Proposition: Let $<\mathbf{W}, \mathcal{B}, P>$ be a measure space satisfying the following six structural assumptions:

- $|\mathbf{W}| = \kappa$ and κ is an uncountable successor cardinal.
- Each point w in **W** belongs as a singleton to $\mathcal{B}$, $\{w\} \in \mathcal{B}$.
- *Tiers* of points, and their intervals under the linear order $\uparrow$ belong to $\mathcal{B}$.
- P is a complete measure, i.e., each subset of a P-null event belongs to $\mathcal{B}$.
- P admits coherent conditional probabilities given non-empty $B \in \mathcal{B}$.
- P is not κ-additive, but P and all its conditional probability functions are γ-additive for each $\gamma < \kappa$.

Then, there is a κ-sized measurable partition π and a measurable event E where P fails to be conglomerable, i.e., there exists an $\varepsilon > 0$ where

$$P(E) > P(E \mid h) + \varepsilon \text{ for each } h \in \pi. \ \blacklozenge_{\textit{Proposition}}$$

The proof of the *Proposition* proceeds through several lemmas, which occupy the rest of this section. The first lemma provides a sufficient condition that a probability P is not κ-additive.

Lemma 2: Consider a measurable λ-partition of an event E, $\pi_E = \{h_\alpha : \alpha < \lambda \leq \kappa\}$ – i.e., where $\cup_{\alpha<\lambda} h_\alpha = E$ and $h_\alpha \cap h_\beta = \varnothing$ whenever $\alpha \neq \beta$. If $P(E) > \Sigma_{\alpha<\lambda} P(h_\alpha)$, then P is not κ-additive.

Proof: Let $E_0 = h_0$, $E_{\alpha+1} = E_\alpha \cup h_{\alpha+1}$ for successor ordinals, and $E_\gamma = h_\gamma \cup_{\alpha<\gamma} E_\alpha$ for limit ordinals $\gamma < \lambda$. So, $E = \cup_{\alpha<\lambda} E_\alpha$. Clearly, $E_\alpha \subseteq E_\beta$ whenever $\alpha < \beta < \lambda$. By assumption, $P(E) > \Sigma_{\alpha<\lambda} P(h_\alpha)$. Let $\beta \leq \kappa$ be the least ordinal such that $P(\cup_{\alpha<\beta} h_\alpha) > \Sigma_{\alpha<\beta} P(h_\alpha)$. Then also $P(\cup_{\alpha<\beta} E_\alpha) > \sup_{\alpha<\beta} P(E_\alpha)$, and so P is not κ-additive. $\lozenge_{\textit{Lemma 2}}$

If P has discrete mass on some points, these form a top tier with cardinality less than or equal to $\aleph_0$. That is, let $\tau^* = \{w: P(\{w\}) > 0\}$. Evidently, by finite additivity, $|\tau^*| \leq \aleph_0$. Since P is countably additive, $P(\tau^*) = \sum_{w\in\tau^*} P(\{w\})$. By the assumption that P is not κ-additive, then $P(\tau^*) < 1$, i.e., P is not perfectly additive. If $\tau^* \neq \emptyset$ then for each other tier, $\tau \neq \tau^*$, $\tau \uparrow \tau^*$. The proof of the main *Proposition* proceeds by considering two cases, depending upon whether some tier τ ($\tau \neq \tau^*$) is non-null, $P(\tau) > 0$ (*Lemma* 3), or whether each tier τ ($\tau \neq \tau^*$) is null, $P(\tau) = 0$ (*Lemma* 4).

Lemma 3: If there exists some tier $\tau \neq \tau^*$ with $P(\tau) > 0$, then P is not conglomerable.
Proof: Since $P(\{w\}) = 0$ whenever $w \notin \tau^*$, because $P(\tau) > 0$ and P is λ-additive for each cardinal $\lambda < \kappa$, then $|\tau| = \kappa$ (*Lemma* 2). Partition τ into two disjoint sets, $T_0 \cap T_1 = \emptyset$ with $T_0 \cup T_1 = \tau$; each with cardinality κ, $|T_0| = |T_1| = \kappa$; and label them so that $P(T_0) \leq P(T_1) = d > 0$.

We identify a partition of cardinality κ, which we write as $\pi = \{h_\alpha: \alpha < \kappa\} \cup \{h'_\beta: \beta < \gamma \leq \kappa\}$, where $\{h_\alpha: \alpha < \kappa\} \cap \{h'_\beta: \beta < \gamma \leq \kappa\} = \emptyset$, and where $P(T_1 \mid h) < d/2$ for each $h \in \pi$. Possibly the second set, $\{h'_\beta: \beta < \gamma \leq \kappa\}$, is empty, as we explain below. Each element $h \in \pi$ is a finite set. Each element h_α contains exactly one point from T_1, and some positive finite number of points from T_0, selected to insure that $P(T_1 \mid h) < d/2$. If the second set, $\{h'_\beta: \beta < \gamma \leq \kappa\}$, is not empty, each $h'_\beta = \{w_\beta\}$ is a singleton with $w_\beta \in \mathbf{W} - T_1$. So, if $\{h'_\beta: \beta < \gamma \leq \kappa\}$ is not empty, then $P(T_1 \mid h'_\beta) = 0$ for each h'_β. Next we establish the existence of such a measurable partition π.

By the Axiom of Choice, consider a κ-long well ordering of T_1, $\{w_1, w_2, \ldots, w_\beta, \ldots\}$ with ordinal indices $0 < \beta < \kappa$. We define π by induction. As each of T_0, T_1 is a subset of the same tier τ, consider the countable partition of T_0 into sets

$$\rho_{1,n} = \{w \in T_0: (n-1)/n \leq P(\{w_1\} \mid \{w_1, w\}) < n/(n+1)\}, \text{ for } n = 1, 2 \ldots .$$

Observe that $\cup_n \rho_{1,n} = T_0$. Since $|T_0| = \kappa \geq \aleph_1$, by the pigeon-hole principle, consider the least n* such that ρ_{1,n^*} is infinite. Let $U_1 = \{w_{1,1}, \ldots, w_{1,m}\}$ be m-many points chosen from ρ_{1,n^*}. Note that $P(\{w_1\} \mid U_1 \cup \{w_1\}) \leq n^*/(m+n^*)$. Choose m sufficiently large so that $n^*/(m+n^*) < d/2$. Let $h_1 = U_1 \cup \{w_1\}$.

For ordinals $1 < \beta < \kappa$, define h_β, by induction, as follows. Denoting $T_{0,1} = T_0$, let $T_{0,\beta} = T_0 - (\cup_{0<\alpha<\beta} h_\alpha)$. Since, for each α, $0 < \alpha < \beta$, by hypothesis of induction h_α is a finite set, then $|\cup_{0<\alpha<\beta} h_\alpha| < \kappa$. So, $|T_{0,\beta}| = \kappa$. Since $T_{0,\beta}$ is a subset of τ, just as above, consider the countable partition of $T_{0,\beta}$ into sets

$$\rho_{\beta,n} = \{w \in T_{0,\beta}: (n-1)/n \leq P(\{w_\beta\} \mid \{w_\beta, w\}) < n/(n+1)\}, \text{ for } n = 1, 2, \ldots .$$

Again, by the pigeon-hole principle, consider the least integer n* such that ρ_{β,n^*} is infinite. Let $U_\beta = \{w_{\beta,1}, \ldots, w_{\beta,m}\}$ be m-many points chosen from ρ_{β,n^*}. Note that

$P(\{w_\beta\} \mid U_\beta \cup \{w_\beta\}) \leq n^*/(m+n^*)$. Choose m sufficiently large that $n^*/(m+n^*) < d/2$. Let $h_\beta = U_\beta \cup \{w_\beta\})$.

Observe that $T_1 \subset \cup_{0<\beta<\kappa} h_\beta$ and that for each $0 < \beta < \kappa$, $P(T_1 \mid h_\beta) < d/2$. In order to complete the partition π, consider a catch-all set with all the remaining points $w_\beta \in \mathbf{W} - \cup_{0<\beta<\kappa} h_\beta$. Note that each such w_β is not a member of T_1, if any such points exist. Add each such point $\{w_\beta\} = h'_\beta$ as a separate partition element of π. Thus, if there are any such points, $P(T_1 \mid h'_\beta) = 0 < d/2$. Hence, P is not conglomerable in π as $P(T_1) = d > 0$, yet for each $h \in \pi$, $P(T_1 \mid h) < d/2$.$\Diamond$*Lemma* 3

Next, we illustrate *Lemma* 3 and also a difference between the de Finetti/Dubins theory of coherent conditional probability used in this paper and the theory of regular conditional distributions from the received (Kolmogorovian) theory of Mathematical Probability.

Example 2: Let <**W**, $\mathcal{B}$> be the measurable space of Lebesgue measurable subsets of the half-open unit interval of real numbers: **W** = [0,1) and $\mathcal{B}$ is its algebra of Lebesgue measurable subsets. Let P be the uniform, countably additive probability with constant density function $f(w) = 1$ for each real number $0 \leq w < 1$, and $f(w) = 0$ otherwise. So $P(\{w\}) = 0$ for each $w \in \mathbf{W}$. Evidently P is not $\kappa = 2^{\aleph_0}$ additive, because **W** is the union of $2^{\aleph_0}$-many null sets: apply *Lemma* 2 to a well order on **W**.

As an illustration of *Lemma* 3, use the uniform density function f to identify coherent conditional probability given finite sets as uniform over those finite sets, as well. That is, when $F = \{w_1, \ldots, w_k\}$ is a finite subset of **W** with k-many points, let $P(\cdot \mid F)$ be the perfectly additive probability that is uniform on these k-many points. These conditional probabilities create a single tier $\tau = \mathbf{W}$, as $P(\{w_1\} \mid \{w_1, w_2\}) = 0.5$ for each pair of points in **W**.

Consider the two events $E = \{w: 0 \leq w < 0.9\}$ and its complement with respect to **W**, $E^c = \{w: 0.9 \leq w < 1\}$, where $P(E) = 0.9$. Let g be the 1-1 (continuous) map between E and E^c defined by $g(w) = 0.9 + w/9$, for $w \in E$. Consider the κ-size partition of **W** by pair-sets, $\pi = \{\{w, g(w)\}: w \in E\}$. By assumption, $P(\{w\} \mid \{w, g(w)\}) = 1/2$ for each pair in π. But then P is not conglomerable in π.

The usual theory of regular conditional distributions treats the example differently. We continue the example from that point of view. Consider the measure space <**W**, $\mathcal{B}$, P> as above. Let the random variable $X(w) = w$, so that $X \sim U[0,1)$, X has the uniform distribution on **W**. In order to consider conditional probability given the pair of points $\{w, g(w)\}$, let

$$g(X) = (X/9) + 0.9 \quad \text{if } 0 \leq X < 0.9$$
$$= 9(X - 0.9) \quad \text{if } 0.9 \leq X < 1.$$

Define the random variable $Y(w) = X(w) + g(X(w)) - 0.9$.
Observe that $Y \sim U[0, 1.0)$. Also, note that Y is 2-to-1 between W and [0.0, 1.0). That is Y = y is entails that either $w = 0.9y$ or $w = 0.1(y + 9)$.

Let the sub-sigma field $\mathcal{A}$ be generated by the random variable Y. The regular conditional distribution relative to this sub-sigma field, $P(\mathcal{B} \mid \mathcal{A})(w)$, is a real-valued function defined on **W** that is $\mathcal{A}$-measurable and satisfies the integral equation

$$\int_A P(B \mid \mathcal{A})(w)\, dP(w) = P(A \cap B)$$

whenever $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

In our case, then $P[B \mid \mathcal{A}](w)$ almost surely satisfies:

$$P(X = 0.9Y \mid Y)(w) = 0.9$$

and

$$P(X = 0.1(Y + 9.0) \mid Y)(w) = 0.1.$$

Thus, relative to the random variable Y, this regular conditional distribution assigns conditional probabilities as if $P(\{w\} \mid \{w, g(w)\}) = 0.9$ for almost all pairs $\{w, g(w)\}$ with $0 \leq w < 0.9$. However, just as in the Borel "paradox" (Kolmogorov, 1933), for a particular pair $\{w, g(w)\}$, the evaluation of $P(\{w\} \mid \{w, g(w)\})$ is not determinate and is defined only relative to which sub-sigma field $\mathcal{A}$ embeds it.

For an illustration of this last feature of the received theory of regular conditional distributions, consider a different pair of complementary events with respect to **W**. Let $F = \{w: 0 \leq w < 0.5\}$ and $F^c = \{w: 0.5 \leq w < 1\}$. So, $P(F) = 0.5$.

Let

$$f(X) = 1.0 - X \quad \text{if } 0 < X < 1.$$
$$= 0 \quad \text{if } X = 0.$$

Analogous to the construction above, let $Z(w) = |X(w) - f(X(w))|$. So Z is uniformly distributed, $Z \sim U[0, 1)$, and is 2-to-1 from **W** onto [0, 1). Consider the sub-sigma field $\mathcal{A}'$ generated by the random variable Z. Then the regular conditional distribution $P(\mathcal{B} \mid \mathcal{A}')(w)$, almost surely satisfies:

$$P(X = 0.5 - Z/2 \mid Z \neq 0)(w) = 0.5$$

and

$$P(X = 0.5 + Z/2 \mid Z \neq 0)(w) = 0.5$$

and for convenience, $P(X = 0 \mid Z = 0) = P(X = 0.5 \mid Z = 0) = 0.5$.

However, $g(.09) = .91 = f(.09)$ and $g(.91) = .09 = f(.91)$. That is, Y = 0.1 if and only if Z = 0.82. So in the received theory, it is permissible to have $P(w = .09 \mid Y = 0.1) = 0.9$ as evaluated with respect to the sub-sigma field generated by Y, and also to have $P(w = .09 \mid Z = 0.82\}) = 0.5$ as evaluated with respect to the sub-sigma field generated by Z, even though the conditioning events are the same event. ◊ Example 2

We resume the proof of the Proposition by turning to the second main case, *Lemma* 4, where each tier τ (other than perhaps τ^*) is a P-null event. The proof of *Lemma* 4, in particular the argument for subcase 2, is indirect. It involves considering a sequence of partitions where, if P is conglomerable in each partition in the sequence, that establishes that P is remote (with extreme values 0 or 1 only) on sets of tiers, i.e., then P is a non-principal ultrafilter distribution on the measurable space of the sets of tiers. Then, using a result of Chang/Kunen-Prikry, P fails to be λ-additive for some $\lambda < \kappa$, establishing the contradiction needed for the indirect argument.

Lemma 4: If for each $\tau \neq \tau^*$, $P(\tau) = 0$, then P is not conglomerable.
Proo*f*: Assume for each $\tau \neq \tau^*$, $P(\tau) = 0$. Let $T = \{\cup\tau: \tau \neq \tau^*\}$. We have assumed that P is not κ additive. So, $P(T) > 0$. And then the cardinality of the set of tiers is $\kappa = |\{\tau\}|$, as P is λ-additive for each cardinal $\lambda < \kappa$.

Consider the linear orders $\uparrow$ and $\downarrow$ over the set of tiers, as defined above. By a familiar result in set theory, either $\uparrow$ or (exclusively) $\downarrow$ is a well order of the set of tiers, or (exclusively) there are two countable subsets $L_\downarrow = \{\tau'_1, \ldots, \tau'_n, \ldots\}$ and $M_\uparrow = \{\tau_1, \ldots, \tau_n, \ldots\}$ of the set of tiers well ordered respectively as the natural numbers, ($\boldsymbol{N}$, <). That is, then elements of $L_\downarrow$ satisfy: $\tau'_m \downarrow \tau'_n$ and elements of $M_\uparrow$ satisfy $\tau_m \uparrow \tau_n$ whenever $n > m$.

We complete the proof of Lemma 4 reasoning by these three sub-cases.

Sub-case 1: Suppose $\uparrow$ is a well order, which we index with an initial segment of the ordinals beginning with 1. Let β be the least ordinal in this well order such that $P(\cup_{\alpha<\beta} \tau_\alpha) > 0$ and let R be this set of tiers, $R = \{\tau_\alpha: \alpha < \beta\}$. Then β is a limit ordinal with $|\beta| = \kappa$, since $P(\tau_\alpha) = 0$ for each tier, and P is λ-additive for each cardinal $\lambda < \kappa$. Note that there is no greatest (last) element of R under $\uparrow$. Partition the tiers in R into those with successor-ordinal indices (S) and those with limit-ordinal indices (L): Since $|\beta| = \kappa$, a regular cardinal, $|S| = |L| = \kappa$.

Because each of S and L has cardinality κ and is cofinal in R, it is an elementary fact that there exist a pair of injective functions f:$\cup S \Rightarrow \cup L$ and g:$\cup L \Rightarrow \cup S$ where $P(\{w\} \mid \{w, f(w)\}) = 0$ and $P(\{w\} \mid \{w, g(w)\}) = 0$, whenever w is in the domain, respectively, of the function f or g, i.e., whenever $w \in \cup S$ or $w \in \cup L$, respectively. That is, each of f and g maps an element of its domain into a distinct element of its range belonging to a higher tier in the well order $\uparrow$. In other words, f pairs points in $\cup S$ with points in $\cup L$ having a higher tier under $\uparrow$. Likewise, g pairs points in $\cup L$ with points in $\cup S$ having a higher tier under $\uparrow$.

For example, write each of S and L as a κ-union of disjoint, cofinal sets: $S = \cup_{\alpha<\kappa} S_\alpha$ and $L = \cup_{\alpha<\kappa} L_\alpha$, where $S_\alpha \cap S_\beta = L_\alpha \cap L_\beta = \emptyset$ if $\alpha \neq \beta$, and each set S_α (respectively, L_α) is cofinal in S (respectively, L). Let f map each tier $\tau_\sigma \in S$ whose index σ is a successor ordinal into elements of the set of tiers L_σ whose indices are limit ordinals greater than σ. Let g map each tier $\tau_\lambda \in$ L whose index λ is a limit ordinal into elements of the set of tiers S_λ whose indices are successor ordinals greater than λ.

Use the functions *f* and *g* to create two κ-size partitions, π_f and π_g, similar in kind to the partition of the *Lemma* 3, as defined below. Without loss of generality, when considering *f* (respectively, *g*), index its domain – for *f* that is the set of points $w \in \cup S$ (respectively for *g*, that is the set of points $w \in \cup L$) – using an initial segment of ordinals beginning with 1 running through κ. That is, when considering *f*, write $\cup S = \{w_1, w_2, \ldots, w_\alpha, \ldots\}$ with $0 < \alpha < \kappa$, and similarly for *g*.

For each ordinal α, $0 < \alpha < \kappa$, define the partition element h_α of π_f to be the pair-set $h_\alpha = \{w_\alpha, f(w_\alpha)\}$. As before define the catch-all set: **W** – $[\cup S \cup \text{Range}(f)]$. And as before, if this set is non-empty add its elements as singleton sets h'_β to create the partition $\pi_f = \{h_1, \ldots, h_\alpha, \ldots\} \cup \{h'_\beta\}$. Then for each $h \in \pi_f$, $P(S \mid h) = 0$. In parallel fashion, with respect to function *g*, define π_g so that for each $h \in \pi_g$, $P(L \mid h) = 0$. Since at least one of S or L is not a P-null set, that is since $max\{P(S), P(L)\} > 0$, P is not conglomerable in at least one of these two partitions, π_f and π_g. ◊ *sub-case 1*

For each of the remaining two subcases within *Lemma* 4, we make use of the following elementary result, which we label *Lemma* 5.

Lemma 5: Let each of U and V be a union of two disjoint sets of tiers, with $P(V) > 0$, $|U| = \kappa$, and with U entirely above V in the linear ordering of ⇂ tiers. That is, for each pair $\tau_U \subset U$ and $\tau_v \subset V$, $\tau_U ⇂ \tau_V$. Then P is not conglomerable.

Proof: This is an easy cardinality argument. Because $\tau_U ⇂ \tau_V$, for each two points $w_U \in \tau_U \subset U$ and $w_v \in \tau_V \subset V$, $P(\{w_V\} \mid \{w_U, w_V\}) = 0$. Since U and V have the same cardinality, κ, consider a 1-1 function to pair them. Let these pair-sets be the partition elements, h_α (for $0 < \alpha < \kappa$) of a κ-size partition, π, augmented by one additional partition element, $h_0 = \mathbf{W} - (U \cup V)$, if h_0 is not empty. Then, for each $h \in \pi$, $P(V|h) = 0$, But $P(V) > 0$. ◊ *Lemma* 5

The following example alerts the reader that sub-cases 1 and 2, where respectively ↾ and ⇂ well order the set of tiers, are not sufficiently parallel to allow using the proof of sub-case 1 for sub-case 2.

Example 3. Consider the case where **W** is countable. Then there cannot be a countably additive probability P as in sub-case 1. That is, if $\mathbf{W} = \{w_1, w_2, \ldots, w_n, \ldots\}$ and each atom constitutes its own tier, $P(\{w_m\} \mid \{w_m, w_n\}) = 0$ whenever $m < n$, then $P(\{w_i\}) = 0$, i = 1, 2, …, contradicting the additivity of P. However, if as in sub-case 2, $P(\{w_m\} \mid \{w_m, w_n\}) = 1$ whenever $m < n$ then this well ordering of the tiers corresponds to a perfectly additive probability P where $P(\{w_1\}) = 1$, and for each nonempty subset $\emptyset \neq S \subseteq \mathbf{W}$, P(E | S) = 1 if and only if E includes the minimal element of S. Note that this probability, P, is remote as are all its conditional probabilities. That is, P(S) and P(E | S) $\in$ {0,1}. ◊ Example 3

As we show below, it is no coincidence that in sub-case 2 of Lemma 4, P is a remote distribution on tiers.

Sub-case 2: Suppose ↓ is a well order of the set of tiers, each of which is P-null. The reasoning begins similarly as for sub-case 1 but relies on results (Chang, 1967; Kunen and Prikry, 1971) concerning descendingly incomplete ultrafilters.

We index the well order ↓ with an initial segment of the ordinals greater than 0. Let β be the least ordinal in this well order such that $P(\cup_{0<\alpha<\beta} \tau_\alpha) > 0$ and let R be this interval of tiers, $R = \{\tau_\alpha: 0 < \alpha < \beta\}$. Then β is a limit ordinal with $|\beta| = \kappa$, since $P(\tau) = 0$ for each tier in R, and P is λ-additive for each cardinal $\lambda < \kappa$. Moreover, by *Lemma* 5, in order for P to be conglomerable, we may assume that $\beta = \kappa$. Note that there is no last (least) element of R under ↓.

By the hypothesis of *Lemma* 4, and in the light of *Lemma* 5, in order for P to be conglomerable, we may also assume that given an ordinal γ, $0 < \gamma < \kappa$, then $|\cup_{0<\alpha<\gamma} \tau_\alpha| < \kappa$. Therefore, $P(\cup_{0<\alpha<\gamma} \tau_\alpha) = 0$. So, for each ordinal α, $0 < \alpha < \kappa$, $P(R) = P(\cup_{\alpha<\gamma<\kappa} \tau_\gamma)$.

In addition, unless P is non-conglomerable, P is remote on sets of tiers in R, i.e. for each (measurable) subset Q of tiers of R, P(Q) $\in$ {0, 1}. This is established by an indirect argument, as follows.

Let Q be a P-non-remote subset of tiers, i.e., 0 < P(Q) < 1. Then, also $0 < P(Q^c) < 1$. By the analysis in the previous paragraph, $|Q| = |Q^c| = \kappa$ and each set is cofinal in the well order ↓ on R. As subsets of the well order ↓ we index each of Q and Q^c by the positive ordinals less than κ. That is, write $Q = \{\tau_\alpha^Q: 0 < \alpha < \kappa\}$ and $Q^c = \{\tau_\alpha^{Q^c}: 0 < \alpha < \kappa\}$ where, for each $0 < \alpha < \kappa$, there exist ordinals $\alpha \leq \beta$, $\alpha \leq \delta$ (with at least one inequality strict) where $\tau_\alpha^Q = \tau_\beta$ and $\tau_\alpha^{Q^c} = \tau_\delta$. We use the convenience of this common indexing of Q and Q^c by ordinals less than κ in order to pair elements of Q and Q^c, as follows.

Let $V = \{\tau_\alpha^Q \in Q: \tau_\alpha^Q \downarrow \tau_\alpha^{Q^c}\}$. That is, when $\tau_\alpha^Q \in V$ then $\tau_\alpha^Q = \tau_\beta$ and $\tau_\alpha^{Q^c} = \tau_\delta$ and $\beta < \delta$. Likewise, when $\tau_\alpha^Q \in Q–V$ then $\tau_\alpha^Q = \tau_\beta$ and $\tau_\alpha^{Q^c} = \tau_\delta$ and $\delta < \beta$.

Observe that $P(Q \cap V^c \mid \{\tau_\alpha^Q, \tau_\alpha^{Q^c}\}) = 0$ whenever $\tau_\alpha^Q \in V$. And as $P(Q \mid \{\tau_\alpha^Q, \tau_\alpha^{Q^c}\}) = 0$ for $\tau_\alpha^Q \in Q–V$, also we have $P(Q \cap V^c \mid \{\tau_\alpha^Q, \tau_\alpha^{Q^c}\}) = 0$ if $\tau_\alpha^Q \in Q\text{-}V$. Hence, if P is conglomerable in the partition by pairs $\pi = \{\{\tau_\alpha^Q, \tau_\alpha^{Q^c}\}: 0 < \alpha < \kappa\}$ then $P(Q \cap V^c) = 0$. Since $P(Q) = P(Q \cap V) + P(Q \cap V^c)$, we conclude that $P(Q) = P(V)$ and so $P(Q^c) = P(V^c)$.

For convenience, index V by the positive ordinals less than κ, $V = \{\tau_\alpha^V : 0 < \alpha < \kappa\}$. Let τ'_α denote that element of Q^c with the same ordinal index as τ_α^V has in the well order of Q. That is, if $\tau_\alpha^V = \tau_\beta^Q$ then $\tau'_\alpha = \tau_\beta^{Q^c}$.

Let $h^*_0 = \{\tau: \tau \downarrow \tau_1^V\}$, which is the interval of tiers preceding tier τ_1^V. By the previous analysis, for each ordinal α, $0 < \alpha < \kappa$, $P(R) = P(\cup_{\alpha<\gamma<\kappa} \tau_\gamma)$. Thus, $P(h^*_0) = 0$, so $P(Q) = P(V) = P(V - h^*_0)$ and likewise, $P(V^c) = P(Q^c) = P(V^c - h^*_0)$.

For $0 < \alpha < \kappa$, let $h^*_\alpha = \{\tau_\alpha^V, \tau'_\alpha\} \cup \{\tau \in Q: \tau_\alpha^V \downarrow \tau \downarrow \tau_{\alpha+1}^V\} \cup \{\tau' \in Q^c: \tau'_\alpha \downarrow \tau' \downarrow \tau'_{\alpha+1}\}$. Observe that if $\tau \in h^*_\alpha$ and $\tau \neq \tau_\alpha^V$ then $\tau_\alpha^V \downarrow \tau$. That is, τ_α^V is the lead element of h^*_α under the well order ↓. Hence, for each α, $0 < \alpha < \kappa$, $P(\{\tau_\alpha^V\} \mid h^*_\alpha) = 1$. So, for each $\alpha < \kappa$, $P([Q^c - h^*_0] \mid h^*_\alpha) = 0$. As $\pi^* = \{h^*_\alpha: \alpha < \kappa)$ partitions the set R, if P is conglomerable in π^*, then $P(Q^c) = 0$. This contradicts the supposition that $0 < P(Q) < 1$. Therefore, if P is conglomerable in π^*, P is remote on all sets of tiers in R. That is, P is a non-principal ultrafilter probability on the algebra of tiers in T.
Next, for each $\alpha < \kappa$, consider the interval I_α within R of tiers below τ_α in the ordering ↓. $I_\alpha = \{\tau \in R: \tau_\alpha \downarrow \tau\}$. These form a κ-long sequence of downward nested intervals, $I_\alpha \supset I_\beta$ whenever $\alpha < \beta < \kappa$, each of which satisfies $P(I_\alpha) = P(R)$. But $P(\cap I_\alpha) = 0$. So P is κ-descendingly incomplete. By a result of Chang (1967) (strengthed by Kunnen and Prikry, 1971 and reported here in an appendix), since κ is a regular successor cardinal, then P admits a λ-descendingly incomplete sequence, for some $\lambda < \kappa$. This contradicts the assumption that P is λ-additive for each $\lambda < \kappa$. Hence, P is not conglomerable in some partition previously identified. ◊ *sub-case 2*

Sub-case 3: There are two countable sets of tiers $L_\downarrow = \{\tau'_1, \ldots, \tau'_n, \ldots\}$ and $M_\uparrow = \{\tau_1, \ldots, \tau_n, \ldots\}$ well ordered respectively as the natural numbers, (***N***, <). That is, the elements of $L_\downarrow$ satisfy: $\tau'_m \downarrow \tau'_n$ and elements of $M_\uparrow$ satisfy $\tau_m \uparrow \tau_n$ whenever $n > m$. Combine these two sequences to form a single countable set ordered (either by ↑ or by ↓) as the combined negative and positive integers under their natural order. That is, form a linearly ordered set of tiers with integer indices, τ_i, for i = …. –n, -(n-1), …., -1, 0, 1, 2, …., (n-1), n, … . By assumption for *Lemma* 4, each of these tiers is a P-null set, and so is their union, which is countable.

Use this null-set of tiers to define countably many intervals of tiers, $I_i = \{\tau \in R: \tau_i \downarrow \tau \downarrow \tau_{i+1}\}$, for I = 0, +/- 1, +/- 2, … . Form a partition of R by adding the two extreme intervals, $I_\infty = \{\tau \in R: \tau_i \downarrow \tau$, for i = 1, 2, …}, and $I_{-\infty} = \{\tau \in R: \tau \downarrow \tau_i$, for i = -1, -2, … }. By the *Lemma* 5, if P is conglomerable, then only one of these intervals is not null. Call it the interval I^*_0. That is, $P(R) = P(I^*_0)$. Thus P is remote on these countably many intervals.

The linear order of tiers within the interval I^*_0 is again one of the three types, corresponding to subcases 1, 2, or 3. If I^*_0 produces a linear order that is a well order, corresponding to either subcase 1 or 2, complete the reasoning for subcase 3 by duplicating that for the respective subcase 1 or 2 applied to the interval I^*_0. If the linear ordering within I^*_0 is also an instance of subcase 3, then repeat the reasoning to produce a subinterval, $I^*_1 \subset I^*_0$ where $P(R) = P(I^*_1)$. Continue in this fashion (letting $I^*_\lambda = \cap I^*_\beta$ for $\beta < \lambda$ at limit ordinals λ) until either subcase 1 or subcase 2 occurs, else there will be a γ-long sequence of nested subintervals $I^*_0 \supset I^*_1 \supset I^*_2 \supset \ldots \supset I^*_\alpha \supset \ldots$, where $P(I^*_\alpha) = P(R)$ for each $\alpha < \gamma$, where $|\gamma| = \kappa$. This will form a κ-descendingly incomplete sequence as $\cap I^*_\alpha = \varnothing$. By appeal to the same result of Chang/Kunnen-Prikry, there is a λ-descendingly incomplete sequence, with $\lambda < \kappa$. This contradicts the assumption that P is λ-additive for each $\lambda < \kappa$. ◊ Subcase 3 and Lemma 4.

The *Proposition* is immediate from Lemmas 3 and 4. ◊ Proposition

5. Conclusion. Given a probability P that satisfies the six structural assumptions of the *Proposition*, we show that non-conglomerability of its coherent conditional probabilities is linked to the index of non-additivity of P. Specifically, as P is not κ-additive then there is a κ-size partition $\pi = \{h_\nu: \nu < \kappa\}$ where the coherent conditional probabilities $\{P(\cdot \mid h_\nu)\}$ are not conglomerable. Namely, there exists an event E and a real number $\varepsilon > 0$ where, for each $h_\nu \in \pi$, $P(E) > P(E \mid h_\nu) + \varepsilon$.

This permits us to conclude that the anomalous phenomenon of non-conglomerability is a result of adopting the de Finetti/Dubins theory of coherent conditional probability instead of the rival Kolmogorovian theory of regular conditional distributions, and not a result of the associated debate over whether probability is allowed to be merely finitely additive rather than satisfying countable additivity. Restated, our conclusion is that even when P is γ-additive for each $\gamma < \kappa$, if P is not κ-additive and has coherent conditional probabilities, then P will experience non-conglomerability in a κ-sized partition. The received theory of regular conditional distributions sidesteps non-conglomerabilithy by allowing conditional probability to depend upon a sub-sigma field, rather than being defined given an event.

*__Acknowledgments:__ In writing this paper we have benefitted by helpful comments from Jeremy Avigad, Jessi Cisewski, Paul Pedersen, and Wilfried Sieg.

References

De Fineitti, B. (1974) *Theory of Probability*. Wiley, New York.

Dubins, L. (1975) Finitely Additive conditional probabilities, conglomerability and disintegrations. Annals of Probability 3: 89-99.

Chang, C.C. (1967) Descendingly Incomplete Ultrafilters. Trans. Amer. Math. Soc. 126: 108-118.

Kadane, J.B., Schervish, M.J., and Seidenfeld, T. (1986) *Statistical Implications of Finite Additivity*. In P.K.Goel and A.Zellner, eds. *Bayesian Inference and Decision Techniques*. Elsevier, Amsterdam: pp. 59-76.

Kadane, J.B., Schervish, M.J., and Seidenfeld, T. (1996) Reasoning to a Foregone Conclusion. *J. American Statistical Association* 91: 1228-1235.

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

Kunen, K. and Prikry, K. (1971) On Descendingly Incomplete Ultrafilters. *J. Symbolic Logic* 36: 650-652.

Schervish, M.J., Seidenfeld, T., and Kadane, J.B. (1984) The Extent of Non-Conglomerability of Finitely Additive Probabilities. *Z. Wahrscheinlichkeits-theorie* 66: 205-226.

Seidenfeld, T., Schervish, M.J. and Kadane, J.B. (2001) Improper Regular Conditional Distributions. *Annals of Probability* 29:1612-1624.

Ulam, S., (1930) Zur Masstheorie in der allgemeinen Mengenlehre, *Fund. Math*. 16: 140-150.

__Appendix__.

Let α, β, λ, and κ be infinite cardinals, δ and γ ordinals, and D an ultrafilter on a set I.

Defn. D is α-descendingly incomplete if there are sets $X_\delta \in D$ (where δ ranges over all ordinals less than α) such that both

(i) for each pair of ordinals δ, γ with $\delta < \gamma < \alpha$, $X_\delta \supseteq X_\gamma$,

and (ii) $\bigcap_{\delta<\alpha} X_\delta = \varnothing$.

GCH abbreviates the Generalized Continuum Hypothesis: $2^\lambda = \lambda^+$

Theorem (using *GCH*, Chang, 1967; without *GCH*, Kunen and Prikry, 1971)

(a) If λ is a regular cardinal and D is λ^+-descendingly incomplete, then D is λ-descendingly incomplete.

(b) If κ = cofinality(λ) < λ and ultrafilter D is λ^+-descendingly incomplete, then either

(i) D is κ-descendingly incomplete, or

(ii) There is an $\alpha < \lambda$ such that D is β-descendingly incomplete for all regular β such that $\alpha < \beta < \lambda$.