Decomposition-Integral: Unifying Choquet and the Concave Integrals*

Yaarit Even[†] and Ehud Lehrer[‡]

First draft

August 17, 2011

Abstract

This paper introduces a novel approach to integrals with respect to capacities. Any random variable is decomposed as a combination of indicators. A pre-specified vocabulary specifies which decompositions are allowed and which are not. Each allowable decomposition has a value determined by the capacity. The decomposition-integral of a random variable is defined as the highest of these values. Thus, different vocabularies induce different decomposition-integrals. It turns out that this decomposition approach unifies under one roof well known integrals, such as Choquet, the concave integral, Riemann and Shilkret, and other plausible integration schemes. For instance, Choquet integral is induced by a vocabulary that allows only increasing (with respect to inclusion) sequences of events. The concave integral, on the other hand, is induced by a vocabulary that allows all possible decompositions.

Various properties of the decomposition-integral, depending on the vocabulary used, are investigated. The main desirable properties of an integral scheme are concavity (risk-aversion), monotonicity with respect to stochastic dominance, and translation-invariance. The paper characterizes the vocabularies that induce decomposition-integrals that respect each of these properties. Finally, an advantage of the decomposition approach is pointed out. The decomposition approach enables one to extend the concave and Choquet integrals beyond the domain of classical capacities. This advantage is illustrated through bicapacities and fuzzy capacities.

*This research was supported in part by the Google Inter-university center for Electronic Markets and Auctions. Lehrer acknowledges the support of the Israel Science Foundation, Grant #538/11.

[†]The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel

[‡]The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel and INSEAD, Boulevard de Constance, 77305 Fontainebleau, France. e-mail: lehrer@post.tau.ac.il

1 Introduction

In economics, and particularly in decision theory under uncertainty, a rational decision maker is often described as an expected utility maximizer. The expected utility is calculated with respect to (w.r.t.) some prior probability over the state space. Although expected utility theory is useful and convenient to work with, different experiments, among which the Ellsberg's paradox [7], show that decision makers often violate this theory.

1.1 Non-additive integral

Schmeidler [25] proposed a theory of decision making, where the belief of the decision maker is represented by a non-additive probability (henceforth referred to as *capacity*). The representation of the belief by a capacity might reflect an incomplete or imprecise information that the decision maker has about the uncertain aspects of the decision problem under consideration. Schmeidler [25] proposed a model where the expected value of a random variable is calculated according to Choquet integral [3]. According to this model, among all alternatives the decision maker chooses the one that maximizes Choquet expected utility.

As an integration scheme, Choquet integral owns two essential properties and lacks one. On one hand, it is monotonic w.r.t. first order stochastic dominance and it is translation-invariant. That is, Choquet expected value of a portfolio with an added constant is equal to the expected value of the original portfolio plus the constant. On the other hand, Choquet integral does not respect risk aversion. In other words, the expected value of two portfolios mixed together is not necessarily greater than, or equal to, the mixture of the expected values of the two portfolios calculated separately.

Lehrer [17] introduced the concave integral with respect to capacities, which differs from Choquet integral. It hinges on the idea underlying the Lebesgue integral and thus respects risk aversion. The concave integral in based on decomposition of random variables to simple ingredients. A *decomposition* is a representation of a random variable as a positive linear combination of indicators.[1] When an indicator is replaced by the value of its corresponding event, the decomposition is transformed to a linear combination of numbers. In other words, a capacity assigns to each decomposition a value: the value corresponding to the linear combination of indicators. This value helps the decision

[1] An indicator of event A, denoted $\mathbb{I}_A$, is the random variable that attains the value 1 on A and the value 0, otherwise.

maker to evaluate any portfolio, even when the information available is incomplete or imprecise. The expected value of a random variable, according to the concave integral, is defined as the maximum value obtained among all its decompositions.

Not only the concave integral can be expressed in terms of decompositions, Choquet integral can also be described in these terms. While the concave integral does not impose any restriction on the decompositions allowed, Choquet integral does. A chain of events is a sequence of decreasing events w.r.t. inclusion. A Choquet decomposition is a decomposition that uses only chains. Like the concave integral, Choquet integral of a random variable is defined as the maximum value obtained among its decompositions, but in this case only among its Choquet decompositions.

Based on the decomposition method, this paper develops a new notion of integral w.r.t. capacities: the *decomposition-integral*. This integral scheme is determined by a vocabulary that dictates which decompositions are allowed and which are not. For instance, when all possible decompositions are allowed, the decomposition-integral coincides with the concave integral, and when only Choquet-decompositions are allowed, the decomposition-integral coincides with Choquet integral.

It turns out that the decomposition approach to integration unifies many other integral schemes. A decomposition of a random variable is partitional if any two of its indicators are disjoint (i.e., obtain the value 1 on disjoint events). Riemann integral coincides with the decomposition-integral when the vocabulary allows only partitional decompositions[2]. Another well known integral that can be expressed in terms of decompositions is Shilkret integral (see, Shilkret [27]). Suppose that a vocabulary allows to use only one indicator at a time. In this case the linear combination consist of merely one indicator. Obviously, there is no way to obtain any random variable as an indicator of an event multiplied by a positive scalar. This is the reason why the integral scheme uses also sub-decompositions. A sub-decomposition of a random variable is a linear combination of indicators, but unlike a decomposition, it does not necessarily coincide with the random variable (it may be smaller). Using the language of decomposition-integrals, Shilkret integral of a random variable is the maximum among all its sub-decompositions that employs only one indicator.

Few desirable properties are maintained by all decomposition-integrals, regardless of the particular vocabulary used. It is said that one random variable is greater than another if the former obtains a higher value than the latter in every possible state. It turns out that when one random variable is higher than another, its decomposition-integral is greater than that of the other. A similar property remains valid when comparing

[2]No reference to the Riemann integral w.r.t. capacities was found in the literature.

two capacities. A capacity is greater than another if it assigns every event a higher value than the other. Regardless of the vocabulary used, the decomposition-integral of the same random variable w.r.t two capacities maintains the order among the capacities. Furthermore, decomposition-integral is homogeneous[3] and is independent of irrelevant events [4]. However, there are desirable properties that are respected by some decomposition-integrals but not by other, depending on the vocabularies used.

One of the advantages of the decomposition method is that it clarifies the trade-off between different desirable properties. Once this trade-off is well formulated, it is left for the decision maker to choose the integration method that owns the properties she considers essential. These are the three desirable properties studied in depth: concavity (risk-aversion), monotonicity w.r.t. first order stochastic dominance, and translation-invariance. It turns out, for instance, that risk-aversion and monotonicity w.r.t. first order stochastic dominance cannot live together. Roughly speaking, the concave integral is the only plausible scheme that respects risk-aversion, while Choquet integral is the only plausible scheme that respects monotonicity w.r.t. first order stochastic dominance, as well as translation-invariance.

The paper also points to another advantage of the decomposition method. In various contexts Choquet integral is extended to domains that lie beyond classical capacities. For instance, Grabisch and Labreuche [9] introduced the notion of bicapacity which is consonant with the prospect theory of Kahneman and Tversky [15]. Bicapacities reflect different attitudes of decision makers toward gains and losses. Grabisch and Labreuche [10] define an integral that extends Choquet integral to the domain of bicapacities. As it turns out, the decomposition method provides a convenient manner to express this definition and to display its similarity with the classical definition.

Another non-classical domain is that of fuzzy capacities (see Lehrer [17]). It is shown that the decomposition approach allows for a natural way to expand Choquet integral to this terrain as well.

1.2 Other integral schemes and unifying approaches

Another well known concept for integration w.r.t. capacities is Sugeno integral [28], also known as the Fuzzy integral. When the capacity takes only the values zero and one

[3]The integral is homogeneous if for every random variable X, and for every positive number c, $\int cX dv = c \int X dv$.

[4]The integral is independent of irrelevant events if for every $A \subseteq N$, $\int \mathbb{I}_A dv = \int \mathbb{I}_A dv_A$, where v_A is defined over A, $v_A(T) = v(T)$ for every $T \subseteq A$.

(a simple game, in the terminology of cooperative games), Sugeno integral coincides with Choquet integral [22], but it does not coincide with the expected value when the capacity is additive. Sugeno integral is not generalized by the decomposition approach. That is, there is no vocabulary that induces a decomposition-integral which coincides with Sugeno integral.

Other unifying approaches were also proposed in the literature. One approach (see, de Campos et al. [4]) unifies Choquet and Sugeno integrals through four essential properties. Another approach (see Klement et al. [16]), which builds on Choquet, Sugeno and Shilkret integrals, defines a universal integral. Both methods use different binary operations instead of the regular addition and multiplication, and both do not generalize the concave integral. It is worth noting also that these unifying approaches do not necessarily coincide with the Lebesgue integral (i.e., the expectation) when the underlying capacity is a probability distribution.

1.3 Related literature

Schmeidler [25, 26] was the first to make the connection between Choquet integral and decisions under uncertainty. Schmeidler provides an axiomatization for Choquet expected utility maximization. Among the follow-ups on Schmeidler's work one can find Gilboa [8] who axiomatized Choquet expected utility maximization in Savage [24] framework, Wakker [31] and Nakamura [23] who examine a finite states space. Wakker [32] characterize optimistic and pessimistic risk attidudes in Schmeidler's model using the Choquet integral.

Dow and Werlang [6] and Lo [21] use Choquet expected utility maximization in a multi-agent models. They extend the notion of Nash equilibrium to cases where the beliefs of players about others' strategies are represented by capacities.

The Choquet integral is also used for pricing insurance contracts and financial assets (see Chateauneuf [1], Waegenaere and Wakker [29], Wang, Young and Panjer [33] and others). Waegenaere, Kast and Lapied [30] show that the Choquet pricing is consistent with a general equilibrium.

Choquet integral is also used in multi criteria decision making and game theory (see Grabisch and Labreuche [11] for a summary on this subject). Marichal [20] uses the Choquet integral as a tool to aggregate interacting criteria. Chiang [5] uses Choquet integral in network implementation for decision analysis.

Lehrer [18] uses the concave integral in a model of decision making and in games with partially-specified probabilities. Lehrer and Teper [19] use the concave integral in a context of decision makers' growing awareness.

1.4 Organization

Section 2 introduces the notion of decompositions and the way they are used to define the decomposition-integral. It is shown that the decomposition-integral generalizes the concave, Choquet, Riemann and Shilkret integrals. This section demonstrates a way to extend Choquet integral to bicapacities and fuzzy capacities using the decomposition approach. This section ends with an elaboration on a few general properties of the decomposition-integral. Section 3 examines three essential properties. Concavity (the main property of Lehrer's concave integral) is discussed first, then monotonicity w.r.t. stochastic dominance and finally, the property of translation-invariance. The vocabularies that induce decomposition-integrals that respect each of these properties are fully characterized.

2 Capacity, decompositions and integrals

2.1 Capacity and a decomposition of a random variable

Let N be a finite set ($|N| = n$). A *capacity* v over N is a function $v : 2^N \to [0, \infty]$ satisfying:
(i) $v(\phi) = 0$; and (ii) $S \subseteq T \subseteq N$ implies $v(S) \leq v(T)$.

A *random variable* (r.v.) X over N is a function $X : N \to \mathbb{R}$. A subset of N will be called an *event*. For any event $A \subseteq N$, $\mathbb{I}_A$ denotes the characteristic function of A, which is the random variable that takes the value 1 over A and the value 0 otherwise.

With the exception of Section 2.5, this paper deals mostly with non-negative random variables and therefore, when we say a random variable, we refer to a non-negative one.

Definition 1 *Let X be a random variable.*

1. *A* sub-decomposition *of X is a finite summation $\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}$ such that*
   (i) $\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i} \leq X$; and
   (ii) $\alpha_i \geq 0$ and $A_i \subseteq N$ for every $i = 1, ..., k$.
2. *Let D be a set of subsets of N. That is, $D \subseteq 2^N$. $\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}$ is a $D-$*sub-decomposition *of X if $A_i \in D$ for every $i = 1, ..., k$.*

We say that $\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}$ is a *decomposition* of X if equality replaces inequality in (i). That is, $\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}$ is a decomposition of X if it is a sub-decomposition of X, and $\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i} = X$. A similar definition applies to $D-$decomposition of X.

Suppose, for instance, that $D = 2^N$ and $X = \mathbb{I}_N$. Then, $X = \sum_{i=1}^{n} \mathbb{I}_{\{i\}}$, and at the same time, $X = \mathbb{I}_N$. Both decompositions use subsets in D.

2.2 Decompositions and integrals

Using the terminology of D-decompositions we can reiterate the definition of the concave integral w.r.t. the capacity v (see [17]):

$$\int^{cav} X dv = \max \left\{ \sum_{i=1}^{k} \alpha_i v(A_i);\ \sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i} \text{is } 2^N\text{-sub-decomposition of } X \right\}. \tag{1}$$

Note that since v is monotonic w.r.t inclusion, in (1) sub-decomposition can be replaced by decomposition. That is,

$$\int^{cav} X dv = \max \left\{ \sum_{i=1}^{k} \alpha_i v(A_i);\ \sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i} \text{is } 2^N\text{-decomposition of } X \right\}.$$

In words, $\int^{cav} X dv$ is the maximum of the values $\sum_{i=1}^{k} \alpha_i v(A_i)$ over all possible decompositions of X. The concave integral imposes no restriction over the decompositions being used: all possible decompositions are taken into account when considering the maximum.

We show that Choquet integral can also be expressed in terms of decompositions. However, unlike the concave integral, Choquet integral does impose restrictions. We say that two subsets A and B of N are *nested* if either $A \subseteq B$ or $B \subseteq A$. A set $D \subseteq 2^N$ is called a *chain* if any two events $A, B \in D$ are nested. Denote by $\mathcal{F}^{Ch}$ the set of all chains. Note that $\mathcal{F}^{Ch}$ is a set of subsets of 2^N. In terms of decompositions, Choquet integral is defined as

$$\begin{aligned} \int^{Ch} X dv &= \max \left\{ \sum_{i=1}^{k} \alpha_i v(A_i);\ \sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i} \text{is } \mathcal{F}^{Ch}\text{-sub-decomposition of } X \right\} & (2) \\ &= \max \left\{ \sum_{i=1}^{k} \alpha_i v(A_i);\ \sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i} \text{is } \mathcal{F}^{Ch}\text{-decomposition of } X \right\}. & (3) \end{aligned}$$

Choquet integral allows to use only chains. Stated differently, Choquet integral is the maximum of $\sum_{i=1}^{k} \alpha_i v(A_i)$, over all decompositions in which every A_i and A_j are nested.

Since any chain is a subset of 2^N, it is evident from (1) and (3) that

$$\int^{Ch} \cdot\, dv \leq \int^{cav} \cdot\, dv.$$

Example 1 *Let* $N = \{1,2,3\}$, $v(N) = 1$, $v(12) = v(13) = 1/2$, $v(23) = 11/12$ *and* $v(1) = v(2) = v(3) = 1/3$. *Define* $X = (3,5,2)$ *to be a variable over* N. *The decomposition* $X = 3\mathbb{I}_{12} + 2\mathbb{I}_{23}$ *is the one at which the maximum of the right hand side of (1) is obtained. Therefore, the concave integral of* X *is*

$$\int^{cav} X dv = 3 \cdot (1/2) + 2 \cdot (11/12) = 3\frac{1}{3}.$$

On the other hand, Choquet integral of X *is obtained at the chain* $\{(2),(12),(123)\}$, *where the decomposition of* X *is* $2\mathbb{I}_2 + 1\mathbb{I}_{12} + 2\mathbb{I}_N$ *and*

$$\int^{Ch} X dv = 2 \cdot (1/3) + 1 \cdot (1/2) + 2 \cdot 1 = 3\frac{1}{6}.$$

2.3 Allowable decompositions and the decomposition-integral

In this part we show that the method of sub-decomposition enables us to unify many well-known and useful methods of integration under one general method. Suppose that $\mathcal{F}$ is a set of subsets of 2^N. Any member of $\mathcal{F}$ is thus a set of events. We refer to $\mathcal{F}$ as a *vocabulary.* A sub-decomposition of X is $\mathcal{F}$*-allowable* if it is $D-$sub-decomposition of X, with the restriction that $D \in \mathcal{F}$. In other words, it has the form $\sum_{A_i \in D} \alpha_i \mathbb{I}_{A_i}$, where $D \in \mathcal{F}$. Thus, in the sub-decomposition of X only events (i.e., $A_i \in D$) from the same D in the vocabulary $\mathcal{F}$ are allowed to be used. The key concept of this paper is introduced in the following definition.

Definition 2 *The* decomposition-integral w.r.t. $\mathcal{F}$ *is defined as follows.*

$$\int_{\mathcal{F}} X dv = \max\left\{ \sum_{A_i \in D} \alpha_i v(A_i);\ \sum_{A_i \in D} \alpha_i \mathbb{I}_{A_i}\ \textit{is}\ \mathcal{F}-\text{allowable sub}-\text{decomposition of X} \right\}. \tag{4}$$

The integral $\int_{\mathcal{F}} \cdot\, dv$ is the maximum over all sub-decompositions that use only A_i's from the same $D \in \mathcal{F}$. The sub-decomposition attaining the maximum in (4) is called the *v-optimal* sub-decomposition (or decomposition) of X w.r.t. $\mathcal{F}$. When no ambiguity arises, we just call it an optimal sub-decomposition (or decomposition) of X.

The following example illustrates the reason why in Definition 2 we allow for sub-decompositions and do not insist on decompositions.

Example 2 [Example 1 continued] *Consider $\mathcal{F}$ defined as follows.*

$$\mathcal{F} = \{\{(1),(23)\},\{(12)\},\{(2),(13)\}\}.$$

Here $\mathcal{F}$ consists of three subsets of 2^N. It turns out that a sub-decomposition, rather than a decomposition, attains the maximum in (4). The optimal sub-decomposition of X is $3\mathbb{I}_{(1)}+2\mathbb{I}_{(23)}$ obtained at $D=\{(1),(23)\}$. Thus, $\int_{\mathcal{F}} Xdv = 3\cdot(1/3)+2\cdot(11/12) = 2\frac{10}{12}$.

Denote by $\mathcal{F}^{cav}$ the vocabulary consisting of merely the set 2^N, then $\int_{\mathcal{F}^{cav}} \cdot\, dv = \int^{cav} \cdot\, dv$ and[5] $\int_{\mathcal{F}^{Ch}} \cdot\, dv = \int^{Ch} \cdot\, dv$. Hence, the concave and Choquet integral differ from each other in the decompositions that they allow for. The concave integral allows for all possible decompositions while Choquet integral allows for chain decompositions (or Choquet decompositions) only. Since the vocabulary $\mathcal{F}^{cav}$ allows for all decompositions, the following statement (stated without a proof) is obtained.

Proposition 1 *Suppose that $\mathcal{F}$ is a vocabulary. Then,*

$$\int_{\mathcal{F}} \cdot\, dv \leq \int_{\mathcal{F}^{cav}} \cdot\, dv$$

for every v.

In other words, of all decomposition integrals, the concave is the highest.

2.4 Riemann integral, Shilkret integral and the minimum

It turns out that other integration schemes also conform to the decomposition method. A partition of N is a set $D=\{A_1,A_2,...,A_k\}$ consisting of pairwise disjoint events whose union is N itself. Denote by $\mathcal{F}^{part}$ the set of all partitions of N. The integral $\int_{\mathcal{F}^{part}} \cdot\, dv$ is Riemann integral.

Consider now the set $\mathcal{F}^{sing}=\{\{A\};\ A\subseteq N\}$. This $\mathcal{F}$ consists of all the singletons whose members are events. The maximum in (4) is obtained at the event that maximizes $\alpha v(A)$, subject to the constraint that $\alpha\mathbb{I}_A \leq X$. Formally,

$$\begin{aligned}\int_{\mathcal{F}^{sing}} Xdv &= \max\left\{\sum_i \alpha_i v(A_i);\ \sum_i \alpha_i \mathbb{I}_{A_i} \text{is } \mathcal{F}^{sing}\text{-allowable sub-decomposition of } X\right\} \\ &= \max\left\{\alpha v(A);\ \alpha\mathbb{I}_A \leq X,\ A\subseteq N,\ \alpha\geq 0\right\} = \max\left\{\alpha\cdot v(X\geq\alpha);\ \alpha\geq 0\right\}.\end{aligned}$$

[5]Coincidentally, the notation $\mathcal{F}^{Ch}$, derived from the word chain, resonates with the notation $\int^{Ch}$ that derives from Choquet.

The right hand side is the scheme known as Shilkret integral of X w.r.t. v.

Another natural vocabulary is the one consisting of a single member: an algebra of sets. We say that D is an *algebra* of sets if it is closed under unions and complement, That is, if $A, B \in D$ implies that $A \cup B$ and $N \setminus A$ are also in D. It might occur that a decision maker is forced or would like to rely only on events in an algebra D. This might happen, for instance, when the decision maker suspects that the information embedded in the capacity about events out of the algebra is unreliable. In this case, employing the integral $\int_{\{D\}} X dv$ to evaluate the random variable X seems to be a natural choice.

Finally, consider the vocabulary $\mathcal{F}$ that consists of $\{N\}$ alone. Then,

$$\int_{\mathcal{F}} X \ dv = \min X.$$

2.5 Decompositions as a means to extend existing integrals to general domains

In this section we demonstrate one advantage of the decomposition method. We have seen that Choquet integral can be defined by means of vocabulary consisting of all possible chains. We employ this method and show that Choquet integral can be naturally generalized to capacities defined over ordered vector spaces. This, in turn, enables us to define Choquet integral w.r.t. bicapacities and fuzzy capacities.

2.5.1 Generalized capacities defined on ordered vector spaces

Let U be a vector space endowed with a partial order $\geq_U$. A *generalized capacity* defined on U is a pair (v, A), such that (i) A is a subset of U containing 0; and (ii) v is a real-valued monotonic function w.r.t. $\geq_U$ and $v(0) = 0$.

Let D be a subset of U. Fix $X \in U$ (not necessarily non-negative). A $D-$sub-decomposition of X is a sum of the type $\sum_{a \in D} \alpha_a \cdot a$ such that (i) $\sum_{a \in D} \alpha_a \cdot a \leq X$; and (ii) $\alpha_a \geq 0$ for every $a \in D$. Let $\mathcal{F}$ be a set of subsets of 2^U. Similar to the definition of decomposition-integral above, we define $\mathcal{F}-$sub-decomposition of X as a $D-$sub-decomposition of X for some $D \in \mathcal{F}$.

We are ready to define the decomposition-integral.

$$\int_{\mathcal{F}} X dv = \sup \left\{ \sum_{a \in D} \alpha v(a); \ \sum_{a \in D} \alpha \cdot a \text{ is} \mathcal{F} - \text{sub-decompositionof } X \right\}. \tag{5}$$

Note that without any restriction on $\mathcal{F}$ or on X, the integral is not always defined. However, in the next two examples, that share the same space $U = \mathbb{R}^n$, but differ in A and in $\geq_U$, the integral is well defined.

2.5.2 First implication: Choquet integral w.r.t. bicapacity

Grabisch and Labreuche [9, 10] introduce bicapacities and define Choquet integral. We take the decomposition approach and deal with the same issue. A bicapacity v is defined over pairs (A, B) of disjoint subsets of N, satisfying the following conditions: (i) $v(\emptyset, \emptyset) = 0$; (ii) v is monotonically increasing in the first argument and monotonically decreasing in the second.

Equivalently, one may think of a bicapacity as a generalized capacity with A being the lattice $\{-1, 0, 1\}^N$, and $U = \mathbb{R}^n$ endowed with the order $\geq_U$, defined as follows: For any two vectors $Y = (y_1, ..., y_n)$ and $Z = (z_1, ..., z_n)$, we say that $Y \geq_U Z$ if $y_i \geq z_i \geq 0$ or $y_i \leq z_i \leq 0$, $i = 1, ..., n$. For instance, in case $n = 3$, one has $(2, 1, -2) \geq (1, 0, -1)$.

Let $\mathcal{F}^{Ch}$ be the collection of all chains (w.r.t. the partial order $\geq_U$) in A. Using only chains to decompose X (again, not necessarily non-negative), as in the classical case, we define Choquet integral to be,

$$\int^{Ch} X dv = \int_{\mathcal{F}^{Ch}} X dv. \tag{6}$$

This definition, it turns out, coincides with that of Grabisch and Labreuche [10] (see also Greco et al. [12]).

A comprehensive discussion on the concave integral w.r.t. bicapacities is deferred to another paper (see, Greco and Lehrer [13]).

2.5.3 Second implication: Choquet integral w.r.t. fuzzy capacities

A capacity might be interpreted as a way to encompass the information available regarding the odds of the various states. A conventional capacity, like an additive probability function, provides the decision maker with the probability of any event. The information about the underlying space, however, might contain only the probability of some, but not all, subsets of N, and the expectation of some random variables. While in the additive case it does not matter how information is given, whether through the probability of events or the expectation of random variables, in the non-additive case it makes a significant difference.

Lehrer [18] gave the example of a dynamic Ellsberg urn in which the decision maker does not know the probability of any color. The information available to the decision maker amounts to knowing the expected value of some, but not all, random variables. This motivated Lehrer [17] to introduce the notion of fuzzy capacities.

Consider the most familiar space of this kind, $\mathbb{R}^n$. For any two vectors $Y = (y_1, ..., y_n)$ and $Z = (z_1, ..., z_n)$, we say that $Y \geq_{\mathbb{R}^n} Z$ if $y_i \geq z_i$ for every $i = 1, ..., n$. A fuzzy capacity is a pair (v, A), where $A \subseteq \mathbb{R}^n$, $a \geq_{\mathbb{R}^n} 0$ for every $a \in A$, and $0 \in A$. Choquet integral uses chains in order to approximate the variable under consideration and is defined for non-negative random variables as in (6) (with the difference that $\mathcal{F}^{Ch}$ is the collection of all chains w.r.t. the partial order $\geq_{\mathbb{R}^n}$).

The following example illustrates the idea.

Example 3 *This example is inspired by Lehrer [18]. Suppose that a decision maker knows that there are three state of nature. Beyond the probability of the entire state space, the decision maker does not know the probability of any event. Instead, she knows that the expected value of the random variables* $(0, 1, 1)$ *and* $\left(1, 1, \frac{1}{2}\right)$ *are* 1 *and* $\frac{1}{2}$*, respectively. This situation is modeled as a fuzzy capacity. Define the fuzzy capacity* (v, A) *as follows:* $A = \left\{(0, 0, 0), (1, 1, 1), \left(1, 0, \frac{1}{2}\right)\right\}$, $v(0, 0, 0) = 0$, $v(1, 1, 1) = 1$ *and* $v\left(1, 0, \frac{1}{2}\right) = \frac{1}{2}$*. The optimal chain sub-decomposition of* $X = (3, 2, 3)$ *is* $2(1, 1, 1) + \left(1, 0, \frac{1}{2}\right)$ *and* $\int^{Ch} X dv = 2 + \frac{1}{2} = 2\frac{1}{2}$*. Thus, given the information available, the expected value of* X *is* $2\frac{1}{2}$*.*

2.6 Properties of the decomposition-integral

2.6.1 Positive homogeneity of degree one

The decomposition-integral is *positive homogeneous* for any vocabulary $\mathcal{F}$ one takes. Meaning, that for every $\lambda > 0$, $\int_{\mathcal{F}} \lambda X dv = \lambda \int_{\mathcal{F}} X dv$ for every X, v and $\mathcal{F}$.

2.6.2 The decomposition-integral and additive capacities

The integral w.r.t a general capacity is meant to generalize the notion of expectation in case the capacity is probability. Riemann, Choquet and the concave integrals indeed coincide with the expectation whenever v is probability, while Shilkret integral does not. The objective of this chapter is to find conditions on the vocabulary which guarantee that the decomposition-integral coincides with the expectation in case the capacity is a probability distribution. Denote by $\mathbb{E}_P(X)$ the expectation of X w.r.t. probability P.

Proposition 2 *Let P be probability and $\mathcal{F}$ be a vocabulary. Then, $\mathbb{E}_P(X) = \int_{\mathcal{F}} XdP$ for every r.v. X, if and only if every X has D-decomposition with $D \in \mathcal{F}$.*

Proof Suppose first that $\mathbb{E}_P(X) = \int_{\mathcal{F}} XdP$ for every r.v. X. In order to attain the value $\mathbb{E}_P(X)$, $\mathcal{F}$-allowable sub-decomposition of X needs to be a decomposition of X. Thus, every X has D-decomposition with $D \in \mathcal{F}$. As for the inverse direction, suppose that every X has a D-decomposition which is $\mathcal{F}$-allowable. Since P is additive, any decomposition of X induces the same value, $\mathbb{E}_P(X)$. Thus, $\mathbb{E}_P(X) = \int_{\mathcal{F}} XdP$ for every X. ■

2.6.3 Monotonicity

The first observation regarding monotonicity refers to fixed vocabularies and capacity. Fix v and $\mathcal{F}$, and suppose that $X \leq Y$. Then, $\int_{\mathcal{F}} Xdv \leq \int_{\mathcal{F}} Ydv$.

The second observation refers to comparison between two capacities. Fix a vocabulary $\mathcal{F}$. If for every $D \in \mathcal{F}$ and every $A \in D$, $v(A) \geq u(A)$, then for every r.v. X, $\int_{\mathcal{F}} Xdu \leq \int_{\mathcal{F}} Xdv$.

The third observation refers to the comparison between two vocabularies. Any vocabulary $\mathcal{F}$ induces a decomposition-integral. The question arises as whether any two different vocabularies induce different integrals. The answer to this question is negative. The following proposition characterizes the circumstances in which the decomposition-integral w.r.t $\mathcal{F}$ is always smaller than, or equal to, that w.r.t $\mathcal{F}'$. For this purpose we need the following definition and lemma.

Definition 3 *Fix a set $C \subseteq 2^N$ of subsets of N. We say that C is* minimal *if the variables $\mathbb{I}_A$, $A \in C$, are algebraically independent.*

In other words, C is minimal if for every variable X there are no two different C-decompositions of X. The $C = \{(12), (1)\}$ is minimal, while $C = \{(12), (1), (2)\}$ is not because $\mathbb{I}_{(1)}, \mathbb{I}_{(2)}$ and $\mathbb{I}_{(12)}$ are linearly dependent. This is demonstrated also by the fact that $\mathbb{I}_{(1)} + \mathbb{I}_{(2)}$ and $\mathbb{I}_{(12)}$ are two different decompositions of the same variable, which employ indicators of events from C.

Lemma 1 *Fix v, $\mathcal{F}$ and X. Suppose that there exists an optimal D-sub-decomposition w.r.t. $\mathcal{F}$. Then, there is a minimal $C \subseteq D$ and an optimal C-sub-decomposition of X.*

The proof is postponed to the Appendix.

Proposition 3 *Suppose that $\mathcal{F}$ and $\mathcal{F}'$ are two vocabularies. Then, $\int_{\mathcal{F}} \cdot\, dv \leq \int_{\mathcal{F}'} \cdot\, dv$ for every v, if and only if for every $D \in \mathcal{F}$ and every minimal set $C \subseteq D$, there is $D' \in \mathcal{F}'$ such that $C \subseteq D'$.*

Proof Suppose that for every $D \in \mathcal{F}$ and minimal $C \subseteq D$, there is $D' \in \mathcal{F}'$ such that $C \subseteq D'$. Fix v and X. Suppose that the optimal sub-decomposition of X w.r.t. $\mathcal{F}$ is obtained at D. By Lemma 1, there is a minimal subset C of D and an optimal C-sub-decomposition of X. By assumption, there is $D' \in \mathcal{F}'$ that contains C as a subset. Thus, there is a D'-sub-decomposition of X that achieves at least the level attained by the D-sub-decomposition of X. Thus, $\int_{\mathcal{F}} X\ dv \leq \int_{\mathcal{F}'} X\ dv$ and since X is arbitrary, $\int_{\mathcal{F}} \cdot\, dv \leq \int_{\mathcal{F}'} \cdot\, dv$.

Now assume that $\int_{\mathcal{F}} \cdot\, dv \leq \int_{\mathcal{F}'} \cdot\, dv$ for every v. Suppose, to the contrary of the proposition, that there are C and $D \in \mathcal{F}$ such that C is minimal, $C \subseteq D$ and no $D' \in \mathcal{F}'$ that contains C as a subset. We construct v and X such that $\int_{\mathcal{F}} X\ dv > \int_{\mathcal{F}'} X\ dv$. Consider the smallest capacity such that $v(A) = \frac{|A|}{|N|}$ for every $A \in C$. That is, $v(B) = 0$, unless $A \subseteq B$ for some $A \in C$, in which case $v(B) = \frac{|A|}{|N|}$, where A is the largest set in C such that $A \subseteq B$. Define, $X = \sum_{A \in C} \mathbb{I}_A$. Thus, $\int_{\mathcal{F}} X\ dv = \sum_{A \in C} \frac{|A|}{|N|}$. Define P to be a uniform distribution – the probability that assigns each point in N a weight of $\frac{1}{|N|}$. Note that

$$\int_{\mathcal{F}} X dv = \mathbb{E}_P(X). \tag{7}$$

Suppose that the optimal sub-decomposition of X w.r.t. $\mathcal{F}'$ is obtained at D'. Denote this sub-decomposition as $\sum_{B \in D'} \beta_B \mathbb{I}_B$. Thus,

$$\int_{\mathcal{F}'} X\ dv = \sum_{B \in D'} \beta_B v(B). \tag{8}$$

We can assume that each $B \in D'$ whose β_B is strictly positive contains at least one $A \in C$ as a subset (since otherwise, $v(B) = 0$). Denote $A(B)$ the largest event in C that is a subset of B, $B \in D'$. By the definition of v,

$$\sum_{B \in D'} \beta_B v(B) = \sum_{B \in D'} \beta_B v(A(B)) = \mathbb{E}_P \left(\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)} \right) \leq \mathbb{E}_P(X). \tag{9}$$

The reason of the last inequality is $\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)} \leq \sum_{B \in D'} \beta_B \mathbb{I}_B \leq X$. Since $\sum_{B \in D'} \beta_B v(B) \leq \mathbb{E}_P(X)$ and due to (7) and (8), we obtain

$$\int_{\mathcal{F}'} X\ dv \leq \int_{\mathcal{F}} X\ dv. \tag{10}$$

We show now that this inequality is strict.

There exist two cases. the first case is when every $A \in C$ has $B \in D'$ such that $A = A(B)$. Since C is not a subset of D', there is $A \in C$ such that $A \notin D'$, implying that $A \subsetneqq B$. This, in turn, implies that $\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)} \neq \sum_{B \in D'} \beta_B \mathbb{I}_B$. Thus,

$$\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)} \neq X. \tag{11}$$

Since P assigns every point in N positive probability, Eqs. (9) and (11) imply

$$\mathbb{E}_P \left(\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)} \right) < \mathbb{E}_P(X).$$

Thus, in light of (8) and (9) inequality (10) is strict.

The second case is where not every $A \in C$ has $B \in D'$ such that $A = A(B)$. It means that not every A in C appears in $\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)}$. Since C is minimal, there are no two C-decompositions of X. This means that $\sum_{B \in D'} \beta_B \mathbb{I}_{A(B))} \neq X$. As in the previous case, this implies that $\mathbb{E}_P \left(\sum_{B \in D'} \beta_B \mathbb{I}_{A(B)} \right) < \mathbb{E}_P(X)$. Thus, again, inequality (10) is strict. ∎

2.6.4 Additivity

A well known property of Choquet integral is comonotone additivity. Two variables X and Y are comonotone if for every $i, j \in N$, $X(i) \geq X(j)$ iff $Y(i) \geq Y(j)$. It turns out that this property can be expressed in terms of vocabularies and optimal decompositions. Consider the vocabulary $\mathcal{F}^{Ch}$ (recall, it consists of all chains). Then, X and Y are comonotone iff the optimal decompositions of X and Y use the same D in $\mathcal{F}^{Ch}$. Comonotone additivity means that if X and Y use the same D for their optimal decomposition, then $\int_{\mathcal{F}^{Ch}} X dv + \int_{\mathcal{F}^{Ch}} Y dv = \int_{\mathcal{F}^{Ch}} (X + Y) \, dv$. A natural question arises as whether this is a general property of the decomposition-integral. That is, whether for any vocabulary $\mathcal{F}$, if X and Y use the same $D \in \mathcal{F}$ for their optimal sub-decomposition, then $\int_{\mathcal{F}} X dv + \int_{\mathcal{F}} Y dv = \int_{\mathcal{F}} (X + Y) \, dv$.

The answer to this question proves to be negative. Indeed, consider the vocabulary $\mathcal{F}^{part}$ (recall, the one consisting of all partitions of N), and a capacity v, defined on $N = \{1, 2\}$ as follows: $v(1) = v(2) = 1/3$ and $v(12) = 1$. Define $X = (\varepsilon, 1), Y = (1, \varepsilon)$, $\varepsilon > 0$. Assume that ε is small enough, so that the optimal decomposition of both X and Y use $D = \{(1), (2)\}$. In this case, $\int_{\mathcal{F}} X dv = \int_{\mathcal{F}} Y dv = (1/3)(1 + \varepsilon)$. As for the sum $X + Y$, taking $D' = \{(12)\}$ yields $\int_{\mathcal{F}} (X + Y) \, dv = 1 + \varepsilon$, which is strictly greater than $\int_{\mathcal{F}} X dv + \int_{\mathcal{F}} Y dv = (2/3)(1 + \varepsilon)$.

The following proposition refers to additivity in case two integrands use the same $D \in \mathcal{F}$ for their optimal decomposition w.r.t to $\mathcal{F}$ and a specific v.

Fix a vocabulary $\mathcal{F}$ and a capacity v. We say that the variable Y is *leaner than* the variable X if there exist (i) an optimal decomposition of Y: $\sum_{A\in C'} \beta_A \mathbb{I}_A$ with $\alpha_A > 0$, $A \in C'$; and (ii) an optimal decomposition of X : $\sum_{A\in C} \alpha_A \mathbb{I}_A$ with $\alpha_A > 0, A \in C$, such that $C' \subset C$. In words, Y is leaner than X, if there are optimal decompositions in which X employs every indicator that Y employs.

Proposition 4 **[Co-decomposition additivity]** *Fix a vocabulary $\mathcal{F}$ such that every X has an optimal decomposition w.r.t. $\mathcal{F}$ for every capacity. Suppose that for every $D, D' \in \mathcal{F}$, whenever there are two different decompositions of the same variable, $\sum_{A\in D} \delta_A \mathbb{I}_A = \sum_{B\in D'} \gamma_B \mathbb{I}_B$, there is $D'' \in \mathcal{F}$ that contains all the A's with $\delta_A > 0$ and all the B's with $\gamma_B > 0$. Then, for every v and every two variables X and Y where Y is leaner than X,*

$$\int_{\mathcal{F}} X dv + \int_{\mathcal{F}} Y dv = \int_{\mathcal{F}} (X+Y)\, dv. \tag{12}$$

Note that the condition of the proposition is readily satisfied by $\mathcal{F}^{Ch}$. The reason is that every random variable essentially (ignoring indicators whose coefficients are zero) has a unique decomposition. This proposition implies the comonotone additivity of Choquet integral. Indeed, considering $\mathcal{F}^{Ch}$, if Y is leaner than X, then Y and X are comonotone. Proposition 4 thus implies Eq. (12), which is precisely comonotone additivity.

This proposition implies that whenever $\mathcal{F}$ consists of only one D, like $\mathcal{F}^{cav}$, its decomposition-integral respects the additivity property (12). It implies that if each of X and Y are equivalently lean (i.e., the same indicators possess positive coefficients in their optimal decompositions), then Eq. (12) holds true. In particular, the concave integral is linear over those variables that use the same indicators in their optimal decompositions.

The additivity of Choquet integral, as expressed in Eq. (12), does not depend on the underlying capacity. Two random variables are comonotone regardless of the capacity v, and for such variables, Eq. (12) would be always true. On the other hand, whether or not Eq. (12) applies to the variables X and Y and the concave integral, does depend on v. The reason for this difference between the integrals is that in Choquet integral the optimal decomposition does not depend on v (it is always the same chain for every v), while it does depend on v when the concave integral is concerned.

Proof of Proposition 4 Fix v and suppose that $X = \sum_{A\in C} \alpha_A \mathbb{I}_A$ with $\alpha_A > 0, A \in C$ is an optimal decomposition of X and $\sum_{A\in C} \beta_A \mathbb{I}_A$ with $\beta_A \geq 0, A \in C$ is an optimal decomposition of Y. We show Eq. (12).

Let $\sum_{B\in D'} \gamma_B \mathbb{I}_B$ be v-optimal decomposition of $X+Y$. If this decomposition equals $\sum_{A\in C}(\alpha_A+\beta_A)\mathbb{I}_A$, then Eq. (12) is true. Otherwise, $\sum_{A\in C}(\alpha_A+\beta_A)\mathbb{I}_A$ and $\sum_{B\in D'} \gamma_B \mathbb{I}_B$

are both different decompositions of $X+Y$. This implies that $\int_{\mathcal{F}} X dv + \int_{\mathcal{F}} Y dv \leq \int_{\mathcal{F}} (X+Y)\, dv$. By assumption, there is D'' that contains all the A's with $\alpha_A + \beta_A > 0$ and all the B's with $\gamma_B > 0$. Thus, X, Y and $X+Y$ all have D'' optimal decompositions (i.e., optimal decompositions of X, Y and $X+Y$ that use only members in D'').

Suppose, to the contrary of the proposition, that $\int_{\mathcal{F}} X dv + \int_{\mathcal{F}} Y dv < \int_{\mathcal{F}} (X+Y)\, dv$. Recall that in the optimal decomposition of X, $\sum_{A\in C} \alpha_A \mathbb{I}_A$, all the coefficients α_A are strictly positive. Thus, for $\varepsilon > 0$ sufficiently small, $\sum_{A\in C} \alpha_A \mathbb{I}_A - \varepsilon \sum_{A\in C} (\alpha_A + \beta_A)\mathbb{I}_A + \varepsilon \sum_{B\in D'} \gamma_B \mathbb{I}_B$ is a D''-decomposition of X (that is, all the coefficients are non-negative). Thus,

$$\begin{aligned} \int_{\mathcal{F}} X dv \;\; \geq \;\; & \sum_{A\in C} \alpha_A v(A) - \varepsilon \sum_{A\in C} (\alpha_A + \beta_A) v(A) + \varepsilon \sum_{B\in D'} \gamma_B v(B) \;\; > \\ & \int_{\mathcal{F}} X dv - \varepsilon \int_{\mathcal{F}} X dv - \varepsilon \int_{\mathcal{F}} Y dv + \varepsilon \left(\int_{\mathcal{F}} X dv + \int_{\mathcal{F}} Y \right) = \int_{\mathcal{F}} X dv. \end{aligned}$$

Since this is a contradiction, Eq. (12) is proven. ■

Definition 4 *Two vocabularies $\mathcal{F}$ and $\mathcal{F}'$ are* equivalent *if they induce the same integral. That is, for every v, $\int_{\mathcal{F}} \cdot dv = \int_{\mathcal{F}'} \cdot dv$.*

The following lemma also refers to a vocabulary, similar to $\mathcal{F}^{cav}$, that consists of only one D.

Lemma 2 *A vocabulary $\mathcal{F}$ is equivalent to a singleton vocabulary $\mathcal{F}'$ iff for every minimal set $C \subseteq \cup\mathcal{F}$ [6] there exists $D \in \mathcal{F}$ such that $C \subseteq D$.*

The proof is postponed to the Appendix.

3 Three essential properties

In this section we state and prove three theorems that deal with the essential properties: concavity, monotonicity w.r.t. stochastic dominance and translation-invariance. We characterize the vocabularies corresponding to decomposition-integrals that maintain each of these properties. Amongst the known integrals we discussed, Choquet integral maintains monotonicity w.r.t. stochastic dominance and translation-invariance, but does not maintain concavity. The concave integral, on the other hand, maintains concavity, but does not maintain monotonicity w.r.t. stochastic dominance and translation-invariance. As we can see, there is a trade-off between the different properties, meaning

[6] $\cup\mathcal{F}$ is a set that contains all $D \in \mathcal{F}$. That is, $\cup\mathcal{F} = \{A \mid A \in D \in \mathcal{F}\}$

that if we want the integral to maintain concavity, we have to give up on monotonicity w.r.t. stochastic dominance, for instance, and vice verse.

3.1 Concavity

In this section we characterize the vocabularies $\mathcal{F}$ for which $\int_{\mathcal{F}} \cdot\, dv$ is concave. We say that $\int_{\mathcal{F}} \cdot\, dv$ is concave if for every two variables X and Y, and $\gamma \in [0,1]$ the following inequality holds true,

$$\int_{\mathcal{F}} (\gamma X + (1-\gamma)Y)\, dv \geq \gamma \int_{\mathcal{F}} X dv + (1-\gamma) \int_{\mathcal{F}} Y dv.$$

3.1.1 Decomposition-integrals that are concave

Theorem 1 *The decomposition-integral $\int_{\mathcal{F}} \cdot\, dv$ is concave for every v, if and only if there exists a vocabulary $\mathcal{F}'$ containing only one D ($D \subseteq 2^N$) such that $\int_{\mathcal{F}} \cdot\, dv = \int_{\mathcal{F}'} \cdot\, dv$.*

Obviously, the concave integral maintains the condition of this theorem, since the vocabulary inducing it is a singleton – it includes only the power set of N.

Proof Suppose there exists a vocabulary $\mathcal{F}'$ containing only one D' ($D' \subseteq 2^N$) such that $\int_{\mathcal{F}} \cdot\, dv = \int_{\mathcal{F}'} \cdot\, dv$. Fix two variables X and Y and $\gamma \in [0,1]$. Consider $\mathcal{F}'$ and denote the optimal sub-decompositions of X and Y by, $\sum_{A\in D'} \alpha_A \mathbb{I}_A$ and $\sum_{A\in D'} \beta_A \mathbb{I}_A$, respectively. The combination $\gamma \sum_{A\in D'} \alpha_A \mathbb{I}_A + (1-\gamma)\sum_{A\in D'} \beta_A \mathbb{I}_A$ is a sub-decompositions of $X+Y$, and its value is $\gamma \int_{\mathcal{F}'} X dv + (1-\gamma)\int_{\mathcal{F}'} Y dv$. Thus, $\int_{\mathcal{F}'}(\gamma X + (1-\gamma)Y) dv$ is greater than, or equal to, this figure. Since, $\int_{\mathcal{F}} \cdot\, dv = \int_{\mathcal{F}'} \cdot\, dv$, we obtain that $\int_{\mathcal{F}} \cdot\, dv$ is concave.

As for the inverse direction, assume that $\int_{\mathcal{F}} \cdot\, dv$ is concave and, in a way of contradiction, that for every $\mathcal{F}'$ that includes only one D, from Lemma 2, there is a minimal $C \subseteq \cup\mathcal{F}$, with no $D \in \mathcal{F}$, such that $C \subseteq D$. This ensures the existence of two disjoint subsets of C, say C_1 and C_2, each contained in a different $D \in \mathcal{F}$ (i.e., $C_i \subseteq D_i \in \mathcal{F}, i = 1,2$) and that no other $D \in \mathcal{F}$ contains both. Since C is minimal, so are C_1 and C_2.

We construct two variables X, Y, and a capacity v, and find $0 < \gamma < 1$ such that $\int_{\mathcal{F}} \gamma X dv + \int_{\mathcal{F}} (1-\gamma) Y dv > \int_{\mathcal{F}} (\gamma X + (1-\gamma)Y) dv$. Define $X = \sum\limits_{A\in C_1} \mathbb{I}_A$ and $Y = \sum\limits_{A\in C_2} \mathbb{I}_A$. Consider the smallest capacity such that $v(A) = \frac{|A|}{|N|}$ for every $A \in C_1 \cup C_2$. That is, $v(B) = 0$ unless $A \subseteq B$ for some $A \in C_1 \cup C_2$, in which case $v(B) = \frac{|A|}{|N|}$, where A is the largest set in $C_1 \cup C_2$, such that $A \subseteq B$. From the definition of v we have obtained that $\int_{\mathcal{F}} X dv = \sum_{A \in C_1} \frac{|A|}{|N|}$ and $\int_{\mathcal{F}} Y dv = \sum_{A \in C_2} \frac{|A|}{|N|}$.

Fix $0 < \gamma < 1$ and suppose that the optimal sub-decomposition of $\gamma X + (1-\gamma) Y$ is $\sum_{E \in D} \beta_E \mathbb{I}_E$. Thus, $\int_{\mathcal{F}} (\gamma X + (1-\gamma) Y) dv = \sum_{E \in D} \beta_E v(E)$, where $D \in \mathcal{F}$. We can assume that each $E \in D$, whose β_E is strictly positive, contains at least one $A \in C_1 \cup C_2$ as a subset (since otherwise, $v(E) = 0$). Denote $A(E)$ the largest set in $C_1 \cup C_2$ that is a subset of E, $E \in D$. Thus, $\sum_{E \in D} \beta_E v(E) = \sum_{E \in D} \beta_E v(A(E))$. There exist two cases. The first case is when every $A \in C_1 \cup C_2$ has $E \in D$ such that $A = A(E)$. Since D does not contain $C_1 \cup C_2$, there is at least one E with $\beta_E > 0$ such that $A(E) \subsetneq E$. Thus, $\sum_{E \in D} \beta_E \mathbb{I}_{A(E)}$ is not a decomposition of $\gamma X + (1-\gamma) Y$ but rather a sub-decomposition of it, implying that

$$\begin{aligned} \int_{\mathcal{F}} (\gamma X dv + (1-\gamma) Y) dv &= \sum_{E \in D} \beta_E v(A(E)) < \\ \gamma \sum_{A \in C_1} v(A) + (1-\gamma) \sum_{A \in C_2} v(A) &= \gamma \int_{\mathcal{F}} X dv + (1-\gamma) \int_{\mathcal{F}} Y dv, \end{aligned}$$

which contradicts concavity. The second case is where not every $A \in C_1 \cup C_2$ has $E \in D$ such that $A = A(E)$. Since $C_1 \cup C_2$ is minimal, there are no two $C_1 \cup C_2$-decompositions of X. This means that $\sum_{E \in D} \beta_E \mathbb{I}_{A(E)}$ is not a decomposition of $\gamma X + (1-\gamma) Y$. As in the previous case, this implies that $\int_{\mathcal{F}} (\gamma X dv + (1-\gamma) Y) dv < \gamma \int_{\mathcal{F}} X dv + (1-\gamma) \int_{\mathcal{F}} Y dv$.

■

3.1.2 An alternative characterization of the concave integral

Another contribution of the decomposition approach is that it provides a new characterization to the concave integrals. The following characterization is corollary of Theorem 1. The first condition enforces that for every event A there is a $D \in \mathcal{F}$ such that $A \in D$, and combined with the second condition we get the concave integral.

Corollary 1 *A decomposition-integral $\int_{\mathcal{F}} \cdot dv$ satisfies (i) $\int_{\mathcal{F}} \mathbb{I}_A dv \geq v(A)$ for every event A and capacity v; and (ii) $\int_{\mathcal{F}} \cdot dv$ is concave, if and only if $\int_{\mathcal{F}} \cdot\, dv = \int^{cav} \cdot\, dv$.*

3.2 Monotonicity w.r.t. stochastic dominance

In this section we characterize $\mathcal{F}$ for which $\int_{\mathcal{F}} \cdot\, dv$ is monotonic w.r.t. stochastic dominance.

Definition 5 *(i) Let v be a capacity defined over N, and X, Y be two variables over N. We say that X* stochastically dominates *Y w.r.t. v (denoted $X \succeq^v Y$), if for every number $t \in \mathbb{R}$, $v(X \geq t) \geq v(Y \geq t)$.*
(ii) In case $X \succeq^v Y$ and $Y \succeq^v X$, we say that X and Y are stochastically equivalent and denote it $X \sim^v Y$.
(iii) We say that $\int_{\mathcal{F}} \cdot dv$ is monotonic w.r.t. first order stochastic dominance *(or simply,* monotonic w.r.t. stochastic dominance*) if $X \succeq^v Y$ implies $\int_{\mathcal{F}} X dv \geq \int_{\mathcal{F}} Y dv$.*

The following definition is important only for the proof and bears no conceptual significance.

Definition 6 *We say that two chains of size k are* similar *if there is a size-preserving one-to-one map between them. Formally, the chains D and G are similar if there is one-to-one map $\phi : D \to G$, such that for every $A \in D, |\phi(A)| = |A|$.*

The following example demonstrates Definitions 5 and 6 and an idea that appears in the proof of Theorem 2.

Example 4 *Let $N = \{1,2,3,4\}$ and $D = \{(1234),(124)\}$. Consider $X = \sum_{T \in D} \mathbb{I}_T = (2,2,1,2)$. We complete D to a chain of size 4: $D' = \{(1),(12),(124),(1234)\}$. Define $F' = \{(3),(34),(234),(1234)\}$. Notice that F' is the complementary chain of D' in the sense that for every $A \in D'$ which is not N, the event $N \setminus A$ belongs to F'.*

Let F be the sub-chain of F' that is similar to D. Thus, $F = \{(234),(1234)\}$. Consider $Y = \sum_{B \in F} \mathbb{I}_B = (1,2,2,2)$. In order to define v, we start with the events in G' or in F'. For every $A \in G' \cup F'$, define $v(A) = \frac{|A|}{4}$. This definition makes X and Y stochastically equivalent ($X \sim^v Y$). On every $B \notin G' \cup F'$ define v to be the minimum possible, while maintaining monotonicity w.r.t. inclusion. That is, $v(B) = \max\{v(A); A \in G' \cup F'$ and $A \subseteq B\}$.

Now suppose that the vocabulary $\mathcal{F}$ includes only D (i.e., $\mathcal{F} = \{D\}$). Then, the optimal sub-decomposition of X is in fact a decomposition: $X = \sum_{T \in D} \mathbb{I}_T$. Thus, $\int_{\mathcal{F}} X dv = \sum_{T \in D} v(T) = 3/4 + 1 = 1\frac{3}{4}$. On the other hand, the optimal sub-decomposition of Y is $\mathbb{I}_N$. Therefore, $\int_{\mathcal{F}} Y dv = v(N) = 1$. We obtain that although $X \sim^v Y$, $\int_{\mathcal{F}} X dv > \int_{\mathcal{F}} Y dv$ and hence, $\int_{\mathcal{F}} \cdot dv$ is not monotonic w.r.t. first order stochastic dominance.

Notice that $\mathcal{F}$ consists only of chains. Theorem 2 states that in order for $\int_{\mathcal{F}} \cdot dv$ to be monotonic w.r.t. first order stochastic dominance, $\mathcal{F}$ must include only chains, and if it includes one chain, it must include all the other chains that have the same size. Thus, the reason for the lack of monotonicity w.r.t. first order stochastic dominance in this example is that $\mathcal{F}$ does not include all chains whose size is the same as D.

3.2.1 The decomposition-integrals that are monotonic w.r.t. stochastic dominance

Theorem 2 *The decomposition-integral $\int_{\mathcal{F}} \cdot\, dv$ is monotonic w.r.t. stochastic dominance, if and only if $\mathcal{F}$ is the collection of all chains of the same size k ($k \in \mathbb{N}$).*

Proof It is easy to show that if $\mathcal{F}$ is the collection of all chains of the same size k ($k \in \mathbb{N}$), then it is monotonic w.r.t. stochastic dominance.

We first show that $\mathcal{F}$ consists only of chains. Assume to the contrary that $\mathcal{F}$ includes at least one D_0 that is not a chain. We construct two variables X, Y, and a capacity v, such that $X \sim^v Y$, but $\int_{\mathcal{F}} X dv > \int_{\mathcal{F}} Y dv$. Since D_0 is not a chain, there are at least two events A, B that are not nested. There are two possible cases. First, $A \cap B = \emptyset$. Define $X = \mathbb{I}_N$ and $Y = \mathbb{I}_A$. Consider the smallest capacity such that $v(A) = v(B) = v(N) = 1$. Obviously, $X \sim^v Y$, but $\int_{\mathcal{F}} X dv = v(A) + v(B) = 2$ and $\int_{\mathcal{F}} Y dv = v(A) = 1$.

The second case is where $A \cap B \neq \emptyset$. Define $X = \mathbb{I}_A + \mathbb{I}_B$ and $Y = \mathbb{I}_N$. With the v before, $X \sim^v Y$, but $\int_{\mathcal{F}} X dv = v(A) + v(B) = 2$ and $\int_{\mathcal{F}} Y dv = v(N) = 1$.

Next we prove that if $\mathcal{F}$ includes one chain of size k, it must include *all* chains of the same size. Assume the opposite. Suppose the longest chain (i.e., one that includes the maximal number of events) in $\mathcal{F}$ is of the size k. First, suppose that D and G are similar chains of size k. If $D \in \mathcal{F}$ and $G \notin \mathcal{F}$, define $X = \sum\limits_{A \in D} \mathbb{I}_A$ and $Y = \sum\limits_{C \in G} \mathbb{I}_C$, and v as the uniform additive probability. We obtain, $X \sim^v Y$ and therefore, $\mathbb{E}_v(Y) = \mathbb{E}_v(X)$. Since we proved that the vocabulary $\mathcal{F}$ consists of chains only, the optimal sub-decomposition of Y is a chain. The variable Y (like any other variable) has exactly one chain decomposition. However, G that used in this decomposition is not in $\mathcal{F}$. Thus, Y has only an optimal sub-decomposition, which is not a decomposition. Suppose that this sub-decomposition is $\sum_{B \in G'} \beta_B \mathbb{I}_B$, where $G' \in \mathcal{F}$. Thus, $\sum_{B \in G'} \beta_B \mathbb{I}_B \leq Y$ and $\sum_{B \in G'} \beta_B \mathbb{I}_B \neq Y$. Since v is the uniform distribution, $\int_{\mathcal{F}} Y dv = \sum_{B \in G'} \beta_B v(B) < \mathbb{E}_p(Y) = \mathbb{E}_p(X) = \int_{\mathcal{F}} X dv$. Therefore, $\int_{\mathcal{F}} X dv > \int_{\mathcal{F}} Y dv$, which contradicts monotonicity w.r.t. first order stochastic dominance. We therefore conclude that similar chains of size k are either all in $\mathcal{F}$, or all out of $\mathcal{F}$.

We show now that all chains of size k are in $\mathcal{F}$. Suppose that $D \in \mathcal{F}$. We complete D to a chain of size n (in an arbitrary way), say $D_1 = \sum_{1 \leq j \leq n} B_j$. Thus, D_1 is a chain of size n that contains D. Define G_1 to be the chain that includes $E_n = N$ and $E_j = N \setminus B_{n-j}$, $j = 1, ..., n-1$. In a sense, G_1 is the complementary chain of D_1.

Let G be a sub-chain of G_1 of size k, such that (i) $G \notin \mathcal{F}$, and (ii) D and G are not similar. As before, define $X = \sum_{A_i \in D} \mathbb{I}_{A_i}$ and $Y = \sum_{C_i \in G} \mathbb{I}_{C_i}$. Consider the smallest capacity such that $v(A_i) = v(C_i) = \frac{i}{n}$ for every $1 \leq i \leq k$. By construction, every $A_i \in D$ and $C_i \in G$ are not nested unless $A_i = N$ or $C_i = N$. Thus, the definition of v does not violate monotonicity w.r.t. inclusion and is therefore well defined. Furthermore, $X \sim^v Y$.

Since $G \notin \mathcal{F}$, the integral of Y is attained at a chain that is a sub-decomposition, say G'. From the definition of v we obtain $\sum_{C \in G} v(C) > \sum_{T \in G'} v(T)$, and thus, $\int_{\mathcal{F}} X dv > \int_{\mathcal{F}} Y dv$, which also contradicts monotonicity w.r.t. first order stochastic dominance. Thus, G must be in $\mathcal{F}$. We conclude that any sub-chain of G_1 whose size is k belongs to $\mathcal{F}$.

We conclude the proof by noting that any chain of size k is similar to a sub-chain of G_1, and as such must be also in $\mathcal{F}$. ■

3.2.2 A new characterization of Choquet integral

Using the notion of decomposition-integral, Theorem 2 provides a new characterization of Choquet integral, one that does not use comonotone additivity. Alongside with the requirement that every variable X has a decomposition, which implies that $k = n$ in Theorem 2, (or alternatively, by Proposition 2, that $\mathbb{E}_P(X) = \int_{\mathcal{F}} X dP$ for every variable X and P additive), we get the following corollary.

Corollary 2 *A decomposition-integral $\int_{\mathcal{F}} \cdot \, dv$ satisfies (i) $\int_{\mathcal{F}} \cdot \, dP = \mathbb{E}_P(\cdot)$ for every probability P; and (ii) it is monotonic w.r.t. stochastic dominance for every v, if and only if $\int_{\mathcal{F}} \cdot \, dv = \int^{Ch} \cdot \, dv$.*

3.3 Translation-invariance

This section provides a characterization of those vocabularies $\mathcal{F}$ that induce a decomposition-integral which is translation-invariant for every v: for every $c > 0$, $\int_{\mathcal{F}}(X + c)dv =$

$\int_{\mathcal{F}} X dv + c$, when $v(N) = 1$. The following illustrates an example where the integral is not translation-invariant.

Example 5 *Let $N = \{1, 2, 3\}$. Consider $\mathcal{F}$ defined as follows. $\mathcal{F} = \{(12), (23), (123)\}$. Define $X = (2, 4, 1)$ and $c = 1$. $v(N) = 1$, $v(12) = v(23) = 2/3$. Then $\int_{\mathcal{F}} X dv = 2 \cdot (2/3) + 1 \cdot (2/3) = 2$ and $\int_{\mathcal{F}} (X + 1) dv = 3 \cdot (2/3) + 2 \cdot (2/3) = 3\frac{1}{3} > 2 + 1$.*

3.3.1 The decomposition-integrals that respect translation-invariance

The following theorem characterizes the vocabularies that always induce an integral which is translation-invariant, regardless of v.

Theorem 3 *The integral $\int_{\mathcal{F}} \cdot\, dv$ is translation-invariant for every v, if and only if the vocabulary $\mathcal{F}$ is (i) composed of chains; and (ii) any $D \in \mathcal{F}$ is contained in $D' \in \mathcal{F}$ that includes $\mathbb{I}_N$.*

Proof Suppose first that the vocabulary $\mathcal{F}$ is (i) composed of chains; and (ii) Any $D \in \mathcal{F}$ is contained in $D' \in \mathcal{F}$ that includes $\mathbb{I}_N$. Fix X. W.l.o.g. one can assume that the optimal sub-decomposition of X is obtained in $D' \in \mathcal{F}$ that includes $\mathbb{I}_N$. Thus, if $\sum_{E \in D'} \alpha_E \mathbb{I}_E$ is an optimal sub-decomposition of X, then $\sum_{E \in D'} \alpha_E \mathbb{I}_E + c\mathbb{I}_N$ is an optimal sub-decomposition of $X + c$, for every $c > 0$. This implies translation-invariance of $\int_{\mathcal{F}} \cdot dv$ for every v.

As for the inverse direction, assume that $\int_{\mathcal{F}} \cdot\, dv$ is translation-invariant for every v. We show first that every $D \in \mathcal{F}$ must be a chain. Else, there is D which contains two non-nested events, say A and B. Let v be the smallest capacity such that $v(A) = v(B) = 2/3$ and $v(N) = 1$. We divide the proof into two cases. The first case is when $A \cap B = \emptyset$. In this case $\mathbb{I}_A + \mathbb{I}_B$ is a sub-decomposition of $X = \mathbb{I}_N$. Thus, $\int_{\mathcal{F}} X\ dv \geq 4/3$. But then, $\int_{\mathcal{F}} (X + 1)\ dv = \int_{\mathcal{F}} 2X\ dv = 2 \int_{\mathcal{F}} X\ dv \geq 8/3$, which implies that $\int_{\mathcal{F}} (X + 1)\ dv \neq \int_{\mathcal{F}} X\ dv + 1$. This constitutes a violation of translation-invariance. It remains to show that in the second case, where $A \cap B \neq \emptyset$, there is also a violation of translation invariance. Consider the variable X defined (similar to Example 5),

$$X(s) = \begin{cases} 2, & \text{if } s \in A \setminus B \\ 4, & \text{if } s \in A \cap B \\ 1, & \text{if } s \in B \setminus A \end{cases}$$

The sum $2\mathbb{I}_A + \mathbb{I}_B$ is an optimal sub-decomposition of X, and therefore $\int_{\mathcal{F}} X\ dv = 2$. However, $3\mathbb{I}_A + 2\mathbb{I}_B$ is a decomposition of $X + 1$ and therefore, $\int_{\mathcal{F}} (X + 1)\ dv \geq$

$5(2/3) > \int_{\mathcal{F}} X \ dv + 1$. Thus, $\int_{\mathcal{F}} \cdot \ dv$ is not translation-invariant. We therefore conclude that Fis composed of chains. We now show that any $D \in \mathcal{F}$ is contained in $D' \in \mathcal{F}$ that includes $\mathbb{I}_N$. Suppose, in a way of contradiction, there exists $D \in \mathcal{F}$ and no $D' \in \mathcal{F}$ that includes $\mathbb{I}_N$ and contains D. Define $X = \sum_{T \in D} \mathbb{I}_T$ and v as the uniform additive probability. Thus, $\int_{\mathcal{F}} X \ dv = \mathbb{E}_v(X)$. The variable $X{+}1$ has exactly one chain decomposition, which is $D \cup N$. By our assumption, $G = D \cup N \notin \mathcal{F}$, thus, $X{+}1$ has only an optimal sub-decomposition, which is not a decomposition, say G'. $\sum_{B \in G'} \beta_B \mathbb{I}_B \leq X + 1$ and $\sum_{B \in G'} \beta_B \mathbb{I}_B \neq X + 1$. Since v is the uniform distribution, $\int_{\mathcal{F}}(X + 1)dv = \sum_{B \in G'} \beta_B v(B) < \mathbb{E}_v(X + 1) = \mathbb{E}_v(X) + 1$. This contradicts translation-invariance. ■

3.3.2 Choquet integral as a decomposition-integral that satisfy translation-invariance

Together with Proposition 2, Theorem 3 provides another characterization of Choquet integral.

Corollary 3 *A decomposition-integral $\int_{\mathcal{F}} \cdot dv$ satisfies (i) $\int_{\mathcal{F}} \cdot dP = \mathbb{E}_P(\cdot)$ for every probability P; and (ii) it is translation-invariant for every v, if and only if $\mathcal{F} = \mathcal{F}^{Ch}$.*

References

[1] Chateauneuf, A., Kast. R and Lapied, A.: Choquet pricing for financial markets with frictions, Mathematical Finance 6, 323-330 (1996)

[2] Chateauneuf, A.: Comonotonicity axioms and rank-dependent expected utility theory for arbitrary consequence. Journal of Mathematical Economics 32, 21–45 (1999)

[3] Choquet, G.: Theory of capacities. Ann Inst Fourier 5, 131-295 (1955)

[4] de Campos, L. M., Lamata, M. T., Moral, S.: A unified approach to define fuzzy integrals. Fuzzy Sets and Systems 39, 75-90 (1991)

[5] Chiang, J.H.: Choquet fuzzy integral-based hierarchical networks for decision analysis. Fuzzy Systems 7, 63-71 (1999)

[6] Dow, J., Werlang, S. R. C.: Nash Equilibrium under Knightian Uncertainty: Breaking down backward induction. Journal of Economic Theory 64, 305-324 (1994)

[7] Ellsberg, D.: Risk, Ambiguity, and the Savage Axioms. The Quarterly Journal of Economics 75, 643-669 (1961)

[8] Gilboa, I.: Expected utility with purely subjective non-additive probabilities. J Math Econ 16, 65-88 (1987)

[9] Grabisch, M., Labreuche, C.: Bi-capacities - Part I: definition, Möbius transform and interaction. Fuzzy Sets and Systems 151, 211-236 (2005)

[10] Grabisch, M., Labreuche, C.: Bi-capacities - Part II: Choquet integral. Fuzzy Sets and Systems 151, 237-259 (2005)

[11] Grabisch, M., Labreuche, C.: A decade of application of Choquet and Sugeno integrals in multi-criteria decision aid. http://ces.univ-paris1.fr/membre/Grabisch/articles/4or07.pdf (2007)

[12] Greco, S., Matarazzo, B. and Giove, S.: Choquet integral with respect to a level dependent capacity. Fuzzy Sets and Systems 175, 1-35 (2011)

[13] Greco, S., and Lehrer, E.: Concave integral for Bi-capacity, mimeo (2011)

[14] Gilboa, I. and Schmeidler, D.: Additive representations of non-additive measures and Choquet integral. Annals of Operations Research 51, 43-65 (1994)

[15] Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk, Econometrica, 263-291 (1979)

[16] Klement, E. P., Mesiar, R., Pap, E.: A universal integral as common frame for Choquet and Sugeno integral. IEEE Transactions on Fuzzy Systems 18, 178-187 (2010)

[17] Lehrer, E.: A new integral for capacities. Economic Theory 39, 157-176 (2009)

[18] Lehrer, E.: Partially specified probabilities: decisions and games, to appear in American Economic Journal: Microeconomics (2007)

[19] Lehrer, E. and Teper, R.: Extension rules or what would the sage do? mimeo (2011)

[20] Marichal, J.L.: An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. Fuzzy Systems 8, 800-807 (2000)

[21] Lo, K.C.: Equilibrium in beliefs under uncertainty, Journal of Economic Theory, 71, 443-483 (1996)

[22] Murofushi, T., Sugeno, M.: Some quantities represented by Choquet integral. Fuzzy Sets and Systems 56, 229-235 (1993)

[23] Nakamura, Y.: Subjective Expected Utility with Non-additive Probabilities on Finite State Spaces. Journal of Economic Theory 51, 346-366 (1990)

[24] Savage, L.J.: The foundations of statistics (Wiley, New York) (1954)

[25] Schmeidler, D.: Integral representation without additivity. Proc Am Math Soc 97, 255-261 (1986)

[26] Schmeidler, D.: Subjective probability and expected utility without additivity. Econometrica 57, 571-587 (1989)

[27] Shilkret, N.: Maxitive measure and integration. Indag Math 33, 109-116 (1971)

[28] Sugeno, M.: Theory of fuzzy integrals and its applications, PhD Dissertation, Tokyo Institute of Technology, Tokyo, Japan (1974)

[29] de Waegenaere, A. and Walker,P.P.: Nonmonotonic Choquet integrals, Journal of mathematical Economics 36, 45-60 (2001)

[30] de Waegenaere, A. , Kast, R. and Lapied, A.: Choquet pricing and equilibrium. Insurance: Mathematics and Economics 32, 359-370 (2003)

[31] Wakker, P.P.: Continuous subjective expected utility with nonadditive probabilities. Journal of Mathematical Economics 18, 1-27 (1989)

[32] Wakker, P.P.: Characterizing optimism and pessimism directly through comonotinicity. Journal of Economic Theory 52, 453-463 (1990)

[33] Wang, S.S., Young, V.R. and Panjer, H.H.: Axiomatic characterization of insurance prices. Insurance: Mathematics and Economics 21, 173-183 (1997)

4 Appendix

Proof of Lemma 1: Suppose that there exists an optimal sub-decomposition of X w.r.t. $\mathcal{F}$, $\sum\limits_{A\in D}\alpha_A\mathbb{I}_A$, $D\in\mathcal{F}$. Define the set $D_X=\{A\in D\mid \alpha_A>0\}$. We may choose a sub-decomposition such that $|D_X|$ is minimal. If D_X is a minimal set, the proof is complete. Otherwise, the variables $\mathbb{I}_A, A \in D_X$, are algebraically dependent. Meaning, that there is a linear combination $\sum_{A \in D_X} \delta_A \mathbb{I}_A = 0$ where at least one $\delta_A \neq 0$. Let $A_0 = \text{argmin}_{A \in D_X, \delta_A \neq 0} \frac{\alpha_A}{|\delta_A|}$ and $\varepsilon = \frac{\alpha_{A_0}}{\delta_{A_0}}$. Since all the coefficients $\alpha_A - \varepsilon \cdot \delta_A$ are greater than or equal to 0, $\sum\limits_{A \in D} \alpha_A \mathbb{I}_A - \varepsilon \sum_{A \in D_X} \delta_A \mathbb{I}_A$ is an optimal sub-decomposition of X. Moreover, for at least one $A \in D_X$, the coefficient $\alpha_A - \varepsilon \cdot \delta_A = 0$. Thus, X has an optimal sub-decomposition that involves a smaller number of indicators than does D_X. This leads a contradiction to the choice of D_X, implying that D_X is indeed minimal. ■

Proof of Lemma 2: Suppose that for every minimal set $C \subseteq \cup\mathcal{F}$ there exists $D \in \mathcal{F}$ such that $C \subseteq D$. Define the singleton vocabulary $\mathcal{F}' = \cup\mathcal{F}$. By assumption, for every $D' \in \mathcal{F}'$ and every minimal set $C \subseteq D'$, there is $D \in \mathcal{F}$ such that $C \subseteq D$. Thus, from Proposition 3, $\int_{\mathcal{F}'} \cdot \, dv \leq \int_{\mathcal{F}} \cdot \, dv$. On the other hand, from the definition of $\mathcal{F}'$, for every $D \in \mathcal{F}$ and every minimal set $C \subseteq D$, there is $D' \in \mathcal{F}'$ such that $C \subseteq D'$. Thus, again, due to Proposition 3, $\int_{\mathcal{F}} \cdot \, dv \leq \int_{\mathcal{F}'} \cdot \, dv$, which leads us to conclude that $\int_{\mathcal{F}} \cdot \, dv = \int_{\mathcal{F}'} \cdot \, dv$.

As for the inverse direction, suppose $\int_{\mathcal{F}'} \cdot \, dv = \int_{\mathcal{F}} \cdot \, dv$, and $\mathcal{F}' = \{D'\}$ (i.e., $\mathcal{F}'$ is a singleton). We show that $\cup\mathcal{F} \subseteq D'$. Assume to the contrary that $\cup\mathcal{F} \nsubseteq D'$. Then, there exists $D_1 \in \mathcal{F}$ such that $D_1 \nsubseteq D'$. By assumption, $\int_{\mathcal{F}'} \cdot \, dv \geq \int_{\mathcal{F}} \cdot \, dv$, and from Proposition 3 we infer that for every minimal set $C \subseteq D_1$, there is $D \in \mathcal{F}'$ such that $C \subseteq D$. Any event in D_1 is a minimal set, thus D' must include any event in D_1, and thus must contain D_1 itself (i.e., $D_1 \subseteq D'$).

Finally, since $\int_{\mathcal{F}'} \cdot \, dv \geq \int_{\mathcal{F}} \cdot \, dv$, we obtain from Proposition 3 that for every minimal set $C \subseteq D'$, there exists $D \in \mathcal{F}$ such that $C \subseteq D$, which completes the proof. ■