

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

ON INDEPENDENCE FOR NON-ADDITIVE MEASURES,
WITH A FUBINI THEOREM

Paolo Ghirardato



SOCIAL SCIENCE WORKING PAPER 940

September 1995

ON INDEPENDENCE FOR NON-ADDITIVE MEASURES, with a Fubini Theorem

Paolo Ghirardato

Abstract

Recent models of decision making represent agents' beliefs by non-additive set-functions. An important technical question which arises in applications to diverse areas of economics is how to define independence of such set-functions. After arguing that the straightforward generalization of independence does not in general yield a unique product, in this work I show that, while Fubini's theorem is in general false if additivity is not granted, it is true when a certain type of function is being integrated. For these functions the iterated integrals coincide with the integral with respect to products which satisfy a certain property, strictly stronger than independence. I show that most of the assumptions made in these results are very close to being necessary. In general the mentioned property is still not strong enough to *uniquely* define a product. On the other hand I discuss some proposals which have been made in the literature, and I show that unicity can however be obtained when the product is assumed to be a belief function. Moreover I show that the unique product thus obtained has an intuitive justification when the marginals are distributions induced by random correspondences. Finally I use the results in the paper to discuss the question of randomization in decision models with non-additive beliefs.

ON INDEPENDENCE FOR NON-ADDITIVE MEASURES, with a Fubini Theorem*

Paolo Ghirardato

1 Introduction and Motivation

The last four decades have witnessed a steady increase in the usage of non-additive set-functions, rather than probabilities, to represent uncertainty. The mathematical theory of non-additive set-functions got its first important contribution with Gustave Choquet's *Theory of Capacities* [2] in 1953. Choquet's interest was applications to statistical mechanics and potential theory. On the other hand non-additive set-functions started to attract economists' attention after the seminal contribution of Lloyd Shapley [34] (also published in 1953) because of their applications to the study of cooperative games, but the connections with decision theory were not explicitly recognized at that time. It was only with the works of Arthur Dempster (e.g. [3]), later developed by Glenn Shafer [32], that applications to uncertainty and the representation of beliefs were considered. In fact Shafer baptized *belief functions* the particular set-functions he and Dempster discussed, emphasizing the knowledge representation aspect. Some years later the interest of statisticians for the subject was aroused by the understanding of its implications for robust Bayesian inference, explicitly presented by Huber [22] in 1973.

Decision theorists rediscovered non-additivity in 1982, when David Schmeidler first circulated an axiomatic model of choice with non-additive beliefs [31]. Spurred by subsequent contributions of Schmeidler and Itzhak Gilboa, like [17], the array of decision theories with non-additive beliefs has by now become substantial. It has been observed that non-additive uncertainty can arise in a lot of different environments and for many different reasons. For example, it can be due to the psychological attitude called "uncertainty aversion", made famous by Ellsberg's celebrated paradox [12], as in Schmeidler's

*This paper is a modified version of chapter 4 of my doctoral dissertation at UC Berkeley. I wish to thank my adviser Bob Anderson, Massimo Marinacci, Chris Shannon and especially Marco Scarsini for helpful comments and discussion. The usual disclaimer applies. Financial support from an Alfred P. Sloan Doctoral Dissertation Fellowship is gratefully acknowledged.

original work. It can be due to “complete ignorance” on subsets of the state space, as in Jaffray and Wakker [23], or to underspecification of the state space, as in Ghirardato [16], Hendon *et al.* [21] and Mukerji [29], or to the decision maker being able to take unobservable actions, as in Kelsey and Milne [24], and so on. All these works obtain some type of integral with respect to a non-additive set-function (the Lebesgue integral can only be defined for additive set-functions) as representation of the decision maker’s preferences, and most use the notion of integral presented by Choquet in [2].

Applications have not been slow in coming. The field of applications to Artificial Intelligence and automated reasoning is considerable, see for instance the reviews of Dubois and Prade [8], Shafer [33] or the recent book by Fagin *et al.* [14]. Wasserman [38] reviews applications to Statistics. In Economics there have been applications to finance and asset pricing (Dow and Werlang [6], Simonsen and Werlang [36] and Epstein and Wang [13]), agency theory (Ghirardato [15]) and game theory (Dow and Werlang [5], Klibanoff [25] and Eichberger and Kelsey [10]).

However the applications to game theory, finance and other fields opened some new theoretical problems. An important problem is defining the independent product of two non-additive set-functions, as a notion of independence is crucial for obtaining laws of large numbers and directly necessary in many applications. For instance, as Hendon *et al.* [20] pointed out, we need some notion of independent product of non-additive measures if we want to discuss games with more than two players in which players believe they are acting independently. A related question is the following: suppose that players have beliefs over random events judged to be somehow independent and that their payoff depends on the outcomes of both events. Are the iterated Choquet integrals — obtained by calculating the “expected payoff” of one random event assuming that the other gave a certain outcome first, and then calculating the integral of the results over all the outcomes of the other event — going to be the same, so that the order of integration does not matter? And, assuming some definition of independent product is used, when is the Choquet integral with respect to this product going to be equal to the iterated integrals? That is, does the way we model the players’ procedure for calculating the expected payoff matter or not?

More technically I am interested in are the following problems: 1) is there a sensible way of *uniquely* defining an independent product of non-additive set-functions? As we shall see presently, this is not a trivial problem, as requiring the usual multiplicative property is not enough for uniquely defining the product (unless the marginals are additive, that is). 2) When does Fubini’s theorem hold? In particular when are iterated Choquet integrals equal, and are they equal to the Choquet integral with respect to the proposed product?

The main result of this paper provides a partial answer to question 2, as it is only to possible to show that there is a fairly large class of functions (which I call *slice-comonotonic functions*) whose iterated integrals coincide for all monotonic and bounded set-functions. However I explain that there does not seem to be much better that we

can do on the way to extending this result to larger classes of functions. As for question 1, it turns out that while it is quite simple to characterize the independent products for which the integral of a slice-comonotonic function coincides with the iterated integrals, *in general* there will still be more than one set-function with this property (which I dub “Fubini property”). As slice-comonotonicity, the Fubini property is seen to be almost necessary to prove this part of the Fubini theorem. Unicity can be obtained when we require the product to have some additional structural properties, for instance that it is a *belief function* (see *infra*, subsection 2.1). Since the notion of integral to be used is, differently from Lebesgue’s, not linear, the validity of Fubini’s theorem turns out to play an important role in establishing whether the expected value of a sum of independent random (non-additive) variables is the sum of their expectations.

The problem of defining an independent product has been studied, to the best of my knowledge, by Hendon *et al.* [20] and, in passing, by Walley and Fine [37] and Gilboa and Schmeidler [17]. These authors have proposed specific notions of product which give rise to a unique product for cases in which marginals have some additional structural properties. Walley and Fine and Gilboa and Schmeidler use a definition which works only when the marginals are *convex* monotonic and bounded set-functions. I show that it has the Fubini property. Hendon *et al.* give a notion of independent product for belief functions, and also that is seen to satisfy the Fubini property. It plays a special role here because it is the object of the unicity result I mentioned above. Moreover I argue that it can be given a somewhat intuitive justification in the case in which the marginal belief functions are obtained as the distributions induced by *correspondences* on a probability space. The validity of Fubini’s theorem for non-additive set-functions has been studied also by Dyckerhoff [9]. Independently from this work, he found a version of the Fubini theorem presented here for the case in which the set-functions are continuous. This essentially leads him to prove a different theorem, with stronger measurability requirements and also some additional technical assumptions that are superfluous here.

Walley and Fine [37] and Dow and Werlang [7] discuss laws of large numbers for processes which are independent in some of the specific senses mentioned earlier. Marinacci [28] has proved that some convergence results can be obtained by independence alone, and that a strong version of Kolmogorov’s strong law of large numbers and a central limit theorem can be obtained if we use the notion of Walley, Fine, Gilboa and Schmeidler.

The organization of the paper is as follows. Section 2 presents the definitions and results that I need for the later analysis. In particular subsection 2.2 explains the problem of uniquely defining an independent product and presents the relevant notions of independence. The main result is presented and proved in subsection 3.1. In subsection 3.2 I give a brief discussion of the possibility of extending the results. There I show that slice-comonotonicity is almost a necessary condition to obtain the equality of iterated integrals, even if the class of set-functions to which the result is applied is limited in some natural way. I also explain my previous claim that the family of slice-comonotonic functions is quite large. Finally I show that the Fubini property is almost necessary in the same way. Section 4 considers the special case in which the set-functions satisfy the

stronger property of convexity. The notion of independence of Gilboa and Schmeidler is presented and discussed there. Section 5 deals with belief functions. In subsection 5.1 I present the notion of product of Hendon *et al.*, discuss its relation to the one presented in section 4, and prove that it is the only independent product belief function with the Fubini property. Then, in subsection 5.2, I give the additional motivation using random correspondences. In subsection 5.3 I show that when one of the marginal set-functions is a probability then there is only one *independent* product belief function. Finally, as an illustration of the possible applications of the results of the paper, in subsection 5.4 I discuss a decision-theoretic problem: Whether it is generally true that uncertainty averse decision makers always prefer to (additively) randomize, or equivalently whether one-stage and two-stage lotteries are ever equivalent with non-additive beliefs.

2 Some Definitions and Preliminary Results

2.1 Capacities and Choquet Integrals

For every space Ω and algebra \mathcal{A} of subsets of Ω a set-function $\sigma : \mathcal{A} \rightarrow \mathbf{R}$ is called a (normalized) *capacity* if it satisfies the following:

- (i) $\sigma(\emptyset) = 0, \sigma(\Omega) = 1,$
- (ii) $\forall A, B \in \mathcal{A} : A \subseteq B \Rightarrow \sigma(A) \leq \sigma(B).$

Note that (i) and (ii) imply that the range of σ is contained in $[0, 1]$. A capacity is called *convex* (or *supermodular*, or *2-monotonic*) if in addition to (i)-(ii) it satisfies the additional property

- (iii) $\forall A, B \in \mathcal{A} : \sigma(A \cup B) \geq \sigma(A) + \sigma(B) - \sigma(A \cap B).$

A special type of convex capacities are the *belief functions* presented and discussed by Dempster [3] and Shafer [32]. These satisfy the following stronger version of (iii) (called *total monotonicity*):

- (iii') For every $n > 0$ and every collection $A_1, \dots, A_n \in \mathcal{A}$

$$\sigma\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{\substack{I \subseteq \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \sigma\left(\bigcap_{i \in I} A_i\right)$$

where $|I|$ is the cardinality of set I .

A capacity is called a *probability* if (iii) holds everywhere with equality (that is, it is *additive*). It is a well-known fact that in such a case one can use induction to show that (iii') will also hold everywhere with equality (the so-called *inclusion-exclusion formulas*). Obviously every probability is a belief function and a convex capacity, but not *vice versa*.

Suppose that Ω is finite, and that \mathcal{A} is the set of all its subsets. Let \mathcal{A}' the set of its non-empty subsets. For $A \in \mathcal{A}'$ let u_A be the capacity (actually a belief function)¹ defined as follows: for every $B \in \mathcal{A}$

$$u_A(B) = \begin{cases} 1 & B \supseteq A, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The following result is well known.

Proposition 1 *The set $\{u_A\}_{A \in \mathcal{A}'}$ forms a linear basis for the set of all capacities on Ω . The unique coefficients $\{\varphi_\sigma(A)\}_{A \in \mathcal{A}'}$ satisfying*

$$\sigma = \sum_{A \in \mathcal{A}'} \varphi_\sigma(A) u_A \quad (2)$$

are given by

$$\varphi_\sigma(A) = \sigma(A) - \sum_{\substack{I \subseteq \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \sigma(\cap_{i \in I} A_i) \quad (3)$$

where $A_i = A \setminus \{\omega_i\}$ and $A = \{\omega_1, \dots, \omega_n\}$.

Clearly σ is a belief function if and only if $\varphi_\sigma \geq 0$, and a probability if additionally φ_σ is non-zero only on singleton subsets. Equation (2) can be rewritten more clearly as follows: for every $A \in \mathcal{A}'$,

$$\sigma(A) = \sum_{B \subseteq A} \varphi_\sigma(B). \quad (4)$$

A function φ_σ satisfying this property is called the *Möbius transform* of σ . So proposition 1 can be restated as: Every capacity has a Möbius transform. For the finite case it was first proved by Shapley [34] and later refined by Dempster [3] and Shafer [32]. Extensions to infinite Ω (with a suitably chosen algebra \mathcal{A}) have been obtained by Gilboa and Schmeidler [18] and Marinacci [27].

Convention. Let me establish here the convention that for every capacity σ , $\varphi_\sigma(\{\emptyset\}) = 0$. This will allow me to extend the summation in (2) and related equations to all of \mathcal{A} (thus avoiding the notationally cumbersome primes) and it brings no loss in generality.

Since we allow the possibility that σ is not additive, we cannot use the integral in the Lebesgue sense to integrate with respect to σ . The notion of integral we will use is

¹These are called *unanimity games* in cooperative game theory, hence the choice of mnemonic.

due to originally to Choquet [2] and it was independently rediscovered and extended by Schmeidler [31]. If $f : \Omega \rightarrow \mathbf{R}$ is a bounded \mathcal{A} -measurable function and σ is any capacity on Ω we define the *Choquet integral* of f with respect to σ to be the number

$$\int_{\Omega} f(\omega) d\sigma(\omega) = \int_0^{\infty} \sigma(\{\omega \in \Omega : f(\omega) \geq \alpha\}) d\alpha + \int_{-\infty}^0 [\sigma(\{\omega \in \Omega : f(\omega) \geq \alpha\}) - 1] d\alpha \quad (5)$$

where the integrals are taken in the sense of Riemann. In particular, if Ω is finite (that is, $\Omega = \{\omega_1, \dots, \omega_n\}$) and $f(\omega_1) \geq f(\omega_2) \geq \dots \geq f(\omega_n)$, then

$$\int_{\Omega} f(\omega) d\sigma(\omega) = \sum_{i=1}^{n-1} (f(\omega_i) - f(\omega_{i+1}))\sigma(\{\omega_1, \dots, \omega_i\}) + f(\omega_n). \quad (6)$$

Notice that since the integrands are monotone, the Choquet integral always exists, and if σ is a probability it reduces to a standard Lebesgue integral. Henceforth all the integrals will be taken in the sense of Choquet, except where otherwise noted.

One of the characteristic traits of the Choquet integral is that it is not in general additive. It is, as we shall presently see, only on special classes of functions.

Definition 1 Let $f : \Omega \rightarrow \mathbf{R}$ and $g : \Omega \rightarrow \mathbf{R}$ be two bounded \mathcal{A} -measurable functions. We say that f and g are *comonotonic* (short for “commonly monotonic”) if for every $\omega, \omega' \in \Omega$,

$$(f(\omega) - f(\omega'))(g(\omega) - g(\omega')) \geq 0. \quad (7)$$

A class of functions \mathcal{F} is said to be *comonotonic* if for every $f, g \in \mathcal{F}$, f and g are comonotonic.

The following proposition summarizes some facts about properties of the Choquet integral which we shall make use of. It is proved in Denneberg [4, Propositions 5.1 and 5.2], which contains an extensive discussion of Choquet integrals and capacities. In the proposition 1_A denotes the characteristic function of $A \in \mathcal{A}$, and the domain of integration is Ω throughout.

Proposition 2 If σ and λ are capacities on (Ω, \mathcal{A}) , and $f, g : \Omega \rightarrow \mathbf{R}$ are bounded \mathcal{A} -measurable functions then:

- (i) $\int 1_A d\sigma = \sigma(A)$, for all $A \in \mathcal{A}$;
- (ii) (positive homogeneity) $\int cf d\sigma = c \int f d\sigma$ for $c \geq 0$;
- (iii) (monotonicity) $f \geq g$ implies $\int f d\sigma \geq \int g d\sigma$;
- (iv) $\int (f + c) d\sigma = \int f d\sigma + c$, for $c \in \mathbf{R}$;

(v) (*comonotonic additivity*) If f, g are comonotonic then

$$\int (f + g) d\sigma = \int f d\sigma + \int g d\sigma;$$

(vi) (*monotonicity in the capacity*) If $\sigma \geq \lambda$ then $\int f d\sigma \geq \int f d\lambda$.

The principal reason for the comonotonic additivity of the Choquet integral is that when integrating comonotonic functions with respect to some capacity σ , their Choquet integrals turn out to be equal to standard integrals with respect to the *same* probability measure.

Lemma 1 *Suppose that \mathcal{F} is a comonotonic class of bounded and \mathcal{A} -measurable functions from Ω into \mathbf{R} and σ is a capacity on (Ω, \mathcal{A}) . Then one can find a probability measure P on (Ω, \mathcal{A}) such that for every $f \in \mathcal{F}$,*

$$\int_{\Omega} f d\sigma = \int_{\Omega} f dP. \quad (8)$$

Proof. For every function $f \in \mathcal{F}$, let \mathcal{U}_f be the set of all upper “intervals” in Ω of the form $\{\omega \in \Omega : f(\omega) \geq \alpha\}$ or $\{\omega \in \Omega : f(\omega) > \alpha\}$ for some $\alpha \in \mathbf{R}$. Clearly \mathcal{U}_f is a chain (a family completely ordered by inclusion) of sets. Comonotonicity of \mathcal{F} can be seen to be equivalent (see Denneberg [4, Proposition 4.5]) to the class

$$\mathcal{U}_{\mathcal{F}} \equiv \bigcup_{f \in \mathcal{F}} \mathcal{U}_f$$

being a chain itself. Consider the restriction of σ to $\mathcal{U}_{\mathcal{F}}$. By proposition 2.10 of Denneberg [4] there exist a unique additive extension P of σ to the algebra $\mathcal{A}_{\mathcal{F}}$ generated by $\mathcal{U}_{\mathcal{F}}$. Since $\mathcal{A}_{\mathcal{F}} \subseteq \mathcal{A}$ by the measurability assumption, we can further extend P to \mathcal{A} so as to preserve additivity (obviously the latter extension is not unique). From formula (5) it is clear that

$$\int_{\Omega} f(\omega) d\sigma(\omega) = \int_{\Omega} f(\omega) dP(\omega)$$

and the same P applies to any f by construction.

Remark 1 As I pointed out in the proof the additive extension of P from $\mathcal{A}_{\mathcal{F}}$ to \mathcal{A} is far from being unique. But this is irrelevant for the result, as I only want to find *one* P which satisfies (8) and any such extension will do. The extensions differ only on sets which do not matter for the calculation of the integral.

For the case of finite² Ω , there is a characterization of the Choquet integral on σ as a standard integral on 2^{Ω} with respect to its Möbius transform which is interesting and useful.

²Again, the result can be generalized to infinite Ω , as shown by Gilboa and Schmeidler in [18].

Proposition 3 (Gilboa and Schmeidler, [19]) For every capacity σ on $(\Omega, 2^\Omega)$ and every function $f : \Omega \rightarrow \mathbf{R}$,

$$\int_{\Omega} f d\sigma = \sum_{A \in \mathcal{A}'} \varphi_{\sigma}(A) \left[\min_{\omega \in A} f(\omega) \right]. \quad (9)$$

2.2 Product Capacities

Let X and Y be two sets and let $Z = X \times Y$. While I will focus on the simple case of the product of two sets only, it will be clear that all results presented here can be immediately generalized to any finite product of sets. Let $\mathcal{A}_X, \mathcal{A}_Y$ denote algebras³ of subsets of X and Y respectively. A *rectangle* is a set $A \subseteq Z$ of the form $A = S \times T$ for some $S \in \mathcal{A}_X$ and $T \in \mathcal{A}_Y$. We let \mathcal{A}_Z be the *product algebra* on Z , that is, the smallest algebra of subsets of Z which contains all rectangles. Any capacity σ on (Z, \mathcal{A}_Z) will be defined a *product capacity*. Its *marginals on X and Y* will be respectively the capacities μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) defined as follows: for all $S \in \mathcal{A}_X$ and $T \in \mathcal{A}_Y$,

$$\mu(S) = \sigma(S \times Y), \quad \nu(T) = \sigma(X \times T). \quad (10)$$

Vice versa, suppose that we are given two capacities μ and ν respectively defined on (X, \mathcal{A}_X) and (Y, \mathcal{A}_Y) . The problem of defining the *independent product* of μ and ν has been discussed by Hendon, Jacobsen, Sloth and Tranæs (henceforth HJST) in [20]. As they observe, the following definition gives the natural generalization of the additive notion of independence.

Definition 2 A product capacity σ on (Z, \mathcal{A}_Z) is called the *independent product* of μ and ν if it satisfies for every $S \in \mathcal{A}_X, T \in \mathcal{A}_Y$,

$$\sigma(S \times T) = \mu(S) \nu(T) \quad (11)$$

One immediately checks that if σ satisfies definition 2 then μ and ν are its marginals, a property that we definitely want an independent product to satisfy. It is well known from measure theory that if both capacities are additive then there is only *one* independent product capacity. This is not the case if either capacity is not additive. In fact in such a case there will be a *set* of such products, as the following result testifies.

Proposition 4 (HJST, [20]) Suppose that μ and ν are respectively capacities on (X, \mathcal{A}_X) and (Y, \mathcal{A}_Y) . Then there are product capacities σ_* and σ^* such that for every product capacity σ (satisfying (i), (ii) and (iv)) we have $\sigma_* \leq \sigma \leq \sigma^*$. Furthermore σ_* and σ^* are defined by

$$\begin{aligned} \sigma^*(A) &\equiv \inf \{ \mu(S) \nu(T) : A \subseteq S \times T, S \in \mathcal{A}_X, T \in \mathcal{A}_Y \}, \\ \sigma_*(A) &\equiv \sup \{ \mu(S) \nu(T) : S \times T \subseteq A, S \in \mathcal{A}_X, T \in \mathcal{A}_Y \}. \end{aligned}$$

³As usual, an *algebra* (or *field*) is a family of subsets of a space which is closed with respect to complements and finite unions (hence also finite intersections).

It is very easy to construct examples which show that the distance $\sigma^*(A) - \sigma_*(A)$ can be quite large on sets A which are not rectangles. (Clearly, all independent product capacities agree on rectangles.)

Independence has a very interesting implication on the hereditarity of properties from products to marginals, contained in the following lemma. The proof is straightforward.

Lemma 2 *Suppose that σ on (Z, \mathcal{A}_Z) is an independent product of μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) . Then if σ is a convex capacity (resp. a belief function), both μ and ν are convex capacities (resp. belief functions).*

Emphatically, except when both marginals are additive, the implication in the lemma does *not* run the other way. That is, an independent product of two belief functions does not have to be a belief function itself, and so on.

As we shall see, there seems to be no sensible general solution to the non-uniqueness problem. One way to restrict the set of products is by imposing additional constraints that they have to satisfy. The following property will turn out to be very important. However its statement requires the introduction of two additional concepts.

Definition 3 *Let $f : Z \rightarrow \mathbf{R}$. We say that f has comonotonic x -sections if for every $x, x' \in X$, $f(x, \cdot) : Y \rightarrow \mathbf{R}$ and $f(x', \cdot) : Y \rightarrow \mathbf{R}$ are comonotonic functions. Comonotonicity of y -sections is defined analogously. f is called slice-comonotonic if it has both comonotonic x -sections and y -sections.*

Slice-comonotonic functions will play a key role in our results. As the name suggests, a function is slice-comonotonic if all its vertical “slices” are a comonotonic family, and the same for its horizontal slices. They will be discussed in greater detail in subsection 3.2.

Definition 4 *A set $A \in \mathcal{A}_Z$ is said to be comonotonic if its characteristic function has comonotonic x -sections.*

Since one can prove (see Denneberg [4, Example 4.5]) that two characteristic functions are comonotonic only if the sets are ordered by inclusion, an equivalent restatement of the definition is the following: A set is comonotonic if the projections on Y of its x -sections form a chain. Interestingly, if a set A satisfies such property then also the projections on X of its y -sections form a chain, so that the characteristic function of A is actually slice-comonotonic. This justifies my usage of the general term *comonotonic*. We now have the terminology to state

Definition 5 *A capacity σ on (Z, \mathcal{A}_Z) with marginals μ and ν is said to satisfy the Fubini property (or a Fubini-independent product) if for every comonotonic set $A \in \mathcal{A}_Z$,*

$$\sigma(A) = \int_Y \int_X 1_A d\mu(x) d\nu(y). \quad (12)$$

Why I chose to give this property such a name will be seen shortly, in subsection 3.1. Since every rectangle is also a comonotonic set, this property implies independence. In general it is stronger, as it imposes agreement on a larger class of sets. However, except for some cases to be discussed below, it is still weak enough to allow a set of possible products of given marginals, as the following example shows.

Example 1 Suppose that $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$, $\mu(x_i) = \alpha_i > 0$ and $\nu(y_i) = \beta_i > 0$ for $i = 1, 2$, where

$$\alpha_1 + \alpha_2 < 1 \quad \text{and} \quad \beta_1 + \beta_2 < 1.$$

For simplicity of notation let $A = (x_1, y_1)$, $B = (x_1, y_2)$, $C = (x_2, y_1)$ and $D = (x_2, y_2)$. Let σ be a Fubini-independent product of μ and ν , with Möbius transform φ . Take a comonotonic set which is not cartesian like $A \cup B \cup C$. By (4) $\sigma(A \cup B \cup C)$ is equal to

$$[\varphi(A) + \varphi(B) + \varphi(C) + \varphi(A \cup B) + \varphi(A \cup C)] + \varphi(B \cup C) + \varphi(A \cup B \cup C).$$

Since σ is an independent product, μ and ν determine the value of φ on all singletons and all rectangles made with two adjacent singletons (like $A \cup B$, or $A \cup C$). Hence the sum in brackets will have to be equal to, say, h for all independent product capacities. Since σ has the Fubini property, also the value of $\sigma(A \cup B \cup C)$ is determined by μ and ν to be equal to, say, k . But now for any pair $(\gamma_1, \gamma_2) \in \mathbf{R}^2$ such that⁴

$$\gamma_1 + \gamma_2 = k - h$$

we can construct an independent product capacity σ such that $\varphi(B \cup C) = \gamma_1$ and $\varphi(A \cup B \cup C) = \gamma_2$, and σ has the Fubini property by construction. However all such σ will differ in the value they associate to $B \cup C$, a non-comonotonic set.

As it turns out, some sharper uniqueness results can be obtained when we require that product capacities be of a specific form, e.g., a belief function, or when one of the marginals is additive. These will be presented and discussed in section 5.

3 Choquet Integration on Product Spaces: A Fubini Theorem

3.1 The Main Result

It is easy to convince oneself that in general the Fubini theorem is false for capacities (for an example see the proof of fact 1 in the next section). One way to proceed in trying to formulate a Fubini-like result is to restrict the set of capacities to which the

⁴It is a consequence of theorem 5.2 below that $k - h = 0$, so that γ_1 and γ_2 must have opposite signs.

theorem should apply, say by requiring that they are convex, or belief functions, and so on. Unfortunately so far it does not seem that anything general can be said in this context (see, e.g., the discussion on products in Chapter 12 of Denneberg [4] and remark 3 below). There is, however, another possibility: to restrict the set of functions to which we wish the result to be applied. Though less aesthetically appealing (and possibly, not as useful), this way of proceeding leads to an extremely simple result, which will be presented in this section. Basically it amounts to extending the results of the comonotonic additivity of the Choquet integral to functions on product spaces.

The key property which we will require of functions is the slice comonotonicity presented in definition 3. The measurability condition is standard in the literature on finitely additive product measures (see, e.g., Marinacci [26]). It is easy to see that an \mathcal{A}_Z -measurable function will have \mathcal{A}_X -measurable y -sections and \mathcal{A}_Y -measurable x -sections.

Theorem 3.1 *Suppose that $f : Z \rightarrow \mathbf{R}$ is bounded, \mathcal{A}_Z -measurable and slice-comonotonic. Then for every pair of capacities μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) , if σ is a product capacity on (Z, \mathcal{A}_Z) satisfying the Fubini property, we have*

$$\begin{aligned} \int_Z f(x, y) d\sigma(x, y) &= \int_Y \int_X f(x, y) d\mu(x) d\nu(y) \\ &= \int_X \int_Y f(x, y) d\nu(y) d\mu(x). \end{aligned} \quad (13)$$

Notice that the theorem holds for every pair of capacities. The proof of theorem 3.1 follows immediately from two lemmas, each one interesting for its own sake. The first shows that iterated integration of slice-comonotonic functions gives the same result regardless of the order in which we iterate.

Lemma 3 *Suppose that $f : Z \rightarrow \mathbf{R}$ is bounded, \mathcal{A}_Z -measurable and slice-comonotonic. Then for every pair of capacities μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) ,*

$$\int_Y \int_X f(x, y) d\mu(x) d\nu(y) = \int_X \int_Y f(x, y) d\nu(y) d\mu(x). \quad (14)$$

Proof. Suppose first that f is a simple function, i.e., it has a finite range. Since f is slice-comonotonic one can find a measurable partition $\{S_1, \dots, S_m\}$ of X and a measurable partition $\{T_1, \dots, T_n\}$ of Y such that for every $y \in Y$ $f(\cdot, y)$ is constant on every S_i and for every $i = 1, \dots, m - 1$, and every $x \in S_i$, $x' \in S_{i+1}$: $f(x, y) \geq f(x', y)$. Symmetrically for every $x \in X$, $f(x, \cdot)$ is constant on T_j and for every $j = 1, \dots, n - 1$, and every $y \in T_j$, $y' \in T_{j+1}$: $f(x, y) \geq f(x, y')$. Let x_i (resp. y_j) be an element of S_i (resp. T_j).

The fact that $f(x_i, \cdot) \geq f(x_{i+1}, \cdot)$ implies, by the monotonicity of the Choquet integral (property (iv) of proposition 2), that

$$F(x_i) \equiv \int_Y f(x_i, y) d\nu(y)$$

is decreasing in i . That is, $F : X \rightarrow \mathbf{R}$ is comonotonic with every $f(\cdot, y_j)$. Moreover, F is clearly bounded and \mathcal{A}_X -measurable. Lemma 1 then implies that there is a probability measure P on (X, \mathcal{A}_X) such that for all $j = 1, \dots, n$,

$$\int_X F(x) d\mu(x) = \int_X F(x) dP(x) \quad \text{and} \quad \int_X f(x, y_j) d\mu(x) = \int_X f(x, y_j) dP(x).$$

By the same token, if we let

$$G(y_j) \equiv \int_X f(x, y_j) d\mu(x)$$

we have that $G : Y \rightarrow \mathbf{R}$ is comonotonic with every $f(x_i, \cdot)$, bounded and \mathcal{A}_Y -measurable, so that there is a probability Q on (Y, \mathcal{A}_Y) such that for every $i = 1, \dots, m$,

$$\int_Y G(y) d\nu(y) = \int_Y G(y) dQ(y) \quad \text{and} \quad \int_Y f(x_i, y) d\nu(y) = \int_Y f(x_i, y) dQ(y).$$

But

$$\begin{aligned} \int_Y \int_X f(x, y) d\mu(x) d\nu(y) &= \int_Y G(y) d\nu(y) \\ &= \int_Y G(y) dQ(y) \\ &= \int_Y \int_X f(x, y) dP(x) dQ(y), \end{aligned} \tag{15}$$

and on the other hand

$$\begin{aligned} \int_X \int_Y f(x, y) d\nu(y) d\mu(x) &= \int_X F(x) d\mu(x) \\ &= \int_X F(x) dP(x) \\ &= \int_X \int_Y f(x, y) dQ(y) dP(x), \end{aligned} \tag{16}$$

and (14) follows from joining (15) with (16) by means of the standard Fubini theorem for additive measures.

Suppose now that f is not a simple function. Following Denneberg [4, lemma 6.2 and exercise 4.14] for $p \in \mathbf{N}$ and $r \in \mathbf{R}$ define the transformation

$$u_p(r) \equiv \sup\{k/2^p : k \in \mathbf{Z}, k/2^p \leq r\}.$$

Let $f_p = u_p \circ f$. Then $\{f_p\}_{p=1}^\infty$ is an increasing sequence of \mathcal{A}_Z -measurable simple functions converging uniformly to f . Also f_p is slice-comonotonic for every p . In fact one can easily see that, since u_p is a monotonic transformation, the fact that $f(x, \cdot)$ and $f(x', \cdot)$ are comonotonic immediately implies that $f_p(x, \cdot)$ and $f_p(x', \cdot)$ are comonotonic (see proposition 5 below). The same can be said for y -sections. By the argument of the previous paragraph we thus have

$$\int_Y \int_X f_p(x, y) d\mu(x) d\nu(y) = \int_X \int_Y f_p(x, y) d\nu(y) d\mu(x) \tag{17}$$

or, adapting the notation above in a straightforward manner,

$$\int_Y G_p(y) d\nu(y) = \int_X F_p(x) d\mu(x).$$

But now notice that $G_p(y)$ converges uniformly to $G(y)$. In fact, by the definition of \bar{u}_p , for all $y \in Y$ we have

$$f(\cdot, y) - 1/2^p \leq f_p(\cdot, y) \leq f(\cdot, y)$$

which implies

$$\int_X f(x, y) d\mu(x) - 1/2^p \leq \int_X f_p(x, y) d\mu(x) \leq \int_X f(x, y) d\mu(x),$$

which is what I claimed. Analogously we can see that $F_p(x)$ converges uniformly to $F(x)$. But this immediately implies that F and G are respectively \mathcal{A}_X - and \mathcal{A}_Y -measurable (they are also obviously bounded) and

$$\int_Y G_p(y) d\nu(y) \xrightarrow{p \rightarrow \infty} \int_Y G(y) d\nu(y)$$

and analogously

$$\int_X F_p(x) d\mu(x) \xrightarrow{p \rightarrow \infty} \int_X F(x) d\mu(x),$$

so taking the limit as $p \rightarrow \infty$ on both sides of (17) yields

$$\int_Y \int_X f(x, y) d\mu(x) d\nu(y) = \int_X \int_Y f(x, y) d\nu(y) d\mu(x).$$

Notice that an immediate implication of this result is that, since the indicator function of a comonotonic set is by definition slice-comonotonic, our choice of the order of integration when stating the Fubini property (on the right-hand side of equation (12)) was immaterial.

The second part of the proof of theorem 3.1 shows that for all Fubini-independent products of μ and ν , the integral with respect to the product of a slice-comonotonic function is equal to the iterated integrals, thus justifying my usage of the name ‘‘Fubini property’’ for such products (see also corollary 3.3 below). Before moving on to that, however, notice by inspecting again equation (5) that only sets⁵ $A \in \mathcal{U}_f$ play some role in the calculation of the integral of a function f . Moreover, as the following simple result shows, if f has comonotonic sections all sets in \mathcal{U}_f will be comonotonic.

Lemma 4 *Let $f : Z \rightarrow \mathbf{R}$ be a bounded, \mathcal{A}_Z -measurable function with comonotonic x - or y -sections. Then every $A \in \mathcal{U}_f$ is a comonotonic set.*

⁵See the proof of lemma 8 for the definition of \mathcal{U}_f .

Proof: Suppose that $A \in \mathcal{U}_f$ and A is not comonotonic. Then we can find $\alpha \in \mathbf{R}$ such that either $A = \{(x, y) \in Z : f(x, y) \geq \alpha\}$ or $A = \{(x, y) \in Z : f(x, y) > \alpha\}$. Suppose the first case obtains. Since A is not comonotonic we can find $x, x' \in X$ and $y, y' \in Y$ such that:

$$f(x, y) \geq \alpha \quad \text{and} \quad f(x, y') < \alpha$$

and

$$f(x', y) < \alpha \quad \text{and} \quad f(x', y') \geq \alpha.$$

But this clearly contradicts the hypothesis that f has comonotonic x - or y -sections. The case with the strict inequality is handled analogously.

We can thus state and prove:

Lemma 5 *Suppose that $f : Z \rightarrow \mathbf{R}$ is bounded, \mathcal{A}_Z -measurable and has comonotonic x -sections. Then for every pair of capacities μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) , if σ is a Fubini-independent product capacity of μ and ν we have*

$$\int_Z f(x, y) d\sigma(x, y) = \int_Y \int_X f(x, y) d\mu(x) d\nu(y). \quad (18)$$

Proof: As in the proof of lemma 3 let us start with the case of a simple f . Then we can find a finite chain of sets $\{A_1, \dots, A_n\}$, $n \geq 1$, such that $A_1 = Z$ and $A_{i+1} \subset A_i$ for $i = 1, \dots, n-1$, and a vector $\alpha \in \mathbf{R}^n$ such that $\alpha_1 < \dots < \alpha_n$, so that

$$f = \alpha_1 1_{A_1} + \sum_{i=2}^n (\alpha_i - \alpha_{i-1}) 1_{A_i}.$$

Clearly (cfr. the proof of lemma 1) we can find a probability measure Q on \mathcal{A}_Y and a family (with finitely many distinct elements) $\{P_y\}_{y \in Y}$ of probability measures on \mathcal{A}_X , such that

$$\int_Y \int_X f(x, y) d\mu(x) d\nu(y) = \int_Y \int_X f(x, y) dP_y(x) dQ(y).$$

Define a capacity λ on \mathcal{U}_f as follows. For every $A \in \mathcal{U}_f$,

$$\lambda(A) = \int_Y \int_X 1_A(x, y) dP_y(x) dQ(y).$$

It is easy to see that

$$\int_Z f(x, y) d\lambda(x, y) = \int_Y \int_X f(x, y) d\mu(x) d\nu(y).$$

But now observe that in this case $\mathcal{U}_f = \{A_1, \dots, A_n\}$. Also, the set $\{f, 1_{A_1}, \dots, 1_{A_n}\}$ forms a comonotonic class. In fact every pair 1_{A_i} and 1_{A_j} is comonotonic because the sets are ordered by inclusion. To see that f and 1_{A_i} are comonotonic for every i , suppose not. Then we can find $(x, y), (x', y') \in Z$ such that $f(x, y) \geq f(x', y')$ and $1_{A_i}(x, y) < 1_{A_i}(x', y')$. The last inequality implies immediately that $(x, y) \notin A_i$ and $(x', y') \in A_i$.

But this implies $f(x, y) < \alpha \leq f(x', y')$, a contradiction. Given that f and 1_{A_i} are comonotonic (so that also $f(\cdot, y)$ and $1_{A_i}(\cdot, y)$ are comonotonic for every $y \in Y$), for every i we must have

$$\int_Y \int_X 1_{A_i}(x, y) d\mu(x) d\nu(y) = \int_Y \int_X 1_{A_i}(x, y) dP_y(x) dQ(y) = \lambda(A_i),$$

and the fact that σ satisfies the Fubini property then implies that $\sigma(A) = \lambda(A)$ for all $A \in \mathcal{U}_f$. So we can use lemma 4 to conclude that

$$\int_Z f(x, y) d\sigma(x, y) = \int_Z f(x, y) d\lambda(x, y) = \int_Y \int_X f(x, y) d\mu(x) d\nu(y),$$

which is what we wanted to prove.

Suppose now that f is not simple. Define the sequence $\{f_p\}_{p=1}^\infty$ as in the proof of lemma 3. Since f has comonotonic x -sections, each f_p will have comonotonic x -sections. Applying the result we just proved, we have that for every p ,

$$\int_Z f_p(x, y) d\sigma(x, y) = \int_Y \int_X f_p(x, y) d\mu(x) d\nu(y). \quad (19)$$

From the proof of lemma 3 we know that

$$\int_Y \int_X f_p(x, y) d\mu(x) d\nu(y) \xrightarrow{p \rightarrow \infty} \int_Y \int_X f(x, y) d\mu(x) d\nu(y).$$

As for the left-hand side, notice that f_p converges to f uniformly on Z , in fact for every $(x, y) \in Z$ we have

$$f(x, y) - 1/2^p \leq f_p(x, y) \leq f(x, y).$$

This implies that

$$\int_Z f(x, y) d\sigma(x, y) - 1/2^p \leq \int_Z f_p(x, y) d\sigma(x, y) \leq \int_Z f(x, y) d\sigma(x, y),$$

so that we immediately get

$$\int_Z f_p(x, y) d\sigma(x, y) \xrightarrow{p \rightarrow \infty} \int_Z f(x, y) d\sigma(x, y).$$

Hence the result follows from taking the limit as $p \rightarrow \infty$ on both sides of (19).

Remark 2 Clearly we could symmetrically prove that if f has comonotonic y -sections then we have

$$\int_Z f(x, y) d\sigma(x, y) = \int_X \int_Y f(x, y) d\nu(y) d\mu(x).$$

Also notice that lemma 3 is immediately implied by two applications of lemma 5. I preferred to state and prove it independently first of all because I believe the proof given above gives better insight as to why the result is true, and second to underline that the result does not depend on the existence of a product capacity with the Fubini property.

3.2 A Discussion

A natural question which arises at this point is: How tight are the assumptions in the theorem? They evidently seem to be just sufficient, but to what extent can we hope to obtain more general results? Regarding lemma 3, it seems clear that the assumption of slice-comonotonicity of f is not necessary. It is easy to provide examples of functions which do not satisfy slice-comonotonicity and pairs of capacities such that the iterated integrals coincide, the following being a simple one.

Example 2 Suppose that the capacities μ and ν express total ignorance in the following extreme sense: For all $S \subseteq X$ such that $S \neq X$ (resp. $T \subseteq Y$ such that $T \neq Y$) we have $\mu(S) = 0$ (resp. $\nu(T) = 0$). That is, $\mu = u_X$ and $\nu = u_Y$ according to (1). Then it is easy to see that for every f

$$\int_Y \int_X f(x, y) d\mu(x) d\nu(y) = \min_{y \in Y} [\min_{x \in X} f(x, y)] \quad (20)$$

and

$$\int_X \int_Y f(x, y) d\nu(y) d\mu(x) = \min_{x \in X} [\min_{y \in Y} f(x, y)], \quad (21)$$

and the right-hand sides of (20) and (21) clearly coincide.

On the other hand the assumption of slice-comonotonicity is necessary in a weaker sense. If a function f is not slice-comonotonic then there are pairs of capacities which are such that the iterated integrals do not give the same result.⁶ The next fact illustrates how one such pair can be constructed in the finite case (but see remark 3 below).

Fact 1 *Suppose that X and Y are finite and $f : Z \rightarrow \mathbf{R}$ is not slice-comonotonic, say because it does not have comonotonic x -sections. Then there are capacities μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) such that*

$$\int_Y \int_X f(x, y) d\mu(x) d\nu(y) \neq \int_X \int_Y f(x, y) d\nu(y) d\mu(x).$$

Proof: For every $x \in X$ find the $y \in Y$ which minimizes $f(x, y)$ and denote it by $y_n(x)$. If $y_n(\cdot)$ is constant on X then for $x \in X$ find $y \in Y \setminus \{y_n(x)\}$ which minimizes $f(x, y)$ and denote it by $y_{n-1}(x)$. If $y_{n-1}(\cdot)$ is again constant on X consider the third worst element $y_{n-2}(\cdot)$, and so on. Since f has non-comonotonic x -sections there must be a $p \geq 0$ such that $y_{n-p}(\cdot)$ is not constant on X . Relabel Y so that $y_{n-p+1}(x), \dots, y_n(x)$ (the choice of x does not matter given the way p is defined) coincide with y_{n-p+1}, \dots, y_n .

⁶This is really analogous to proving that if f and g are non-comonotonic functions on some set Ω then one can find a capacity σ on Ω such that their Choquet integral with respect to σ is not additive.

Let μ be a uniform probability measure on X (that is, $\mu(x) = 1/m$ for every $x \in X$) and let $\nu = u_T$ where $T = \{y_1, \dots, y_{n-p}\}$. Then it is easy to see that for every x ,

$$\begin{aligned} \int_Y f(x, y) d\nu(y) &= \min_{y \in T} f(x, y) \\ &= f(x, y_{n-p}(x)) \end{aligned}$$

so that

$$\int_X \int_Y f(x, y) d\nu(y) d\mu(x) = \frac{1}{m} \sum_{i=1}^m f(x_i, y_{n-p}(x_i)). \quad (22)$$

On the other hand it is clear that for every y , $\int_X f(x, y) d\mu(x) = (1/m) \sum_i f(x_i, y)$, hence

$$\int_Y \int_X f(x, y) d\mu(x) d\nu(y) = \frac{1}{m} \min_{y \in T} \left[\sum_{i=1}^m f(x_i, y) \right]. \quad (23)$$

Consider the matrix $[f(x_i, y_j)]$ for $i = 1, \dots, m$ and $j = 1, \dots, n-p$. We are done if we show that the average of the minima for each column (i.e., the right-hand side of (22)) is different from the minimum of the averages of each row (the right-hand side of (23)).

In general let A be any $r \times t$ matrix. Let \underline{a}_j be the minimum of column $j = 1, \dots, t$ and let \bar{a}_i be the average of row $i = 1, \dots, r$. Then I claim that

$$\sum_{j=1}^t \underline{a}_j = \min_i \bar{a}_i \quad (24)$$

iff there is a row such that all the minima \underline{a}_j lie on it. One implication is immediate. For the other suppose that there is no row such that all minima lie on it. Then we can find j^* and j^{**} such that \underline{a}_{j^*} lies on some⁷ row i^* and $\underline{a}_{j^{**}}$ lies on some row $i^{**} \neq i^*$. So we have

$$\underline{a}_{j^*} + \underline{a}_{j^{**}} < a_{i^*, j^*} + a_{i^{**}, j^{**}}$$

which clearly implies that

$$\sum_j \underline{a}_j < \sum_j a_{i^*, j}$$

The claim then follows from observing that for every i we also have

$$\sum_j \underline{a}_j \leq \sum_j a_{i, j} = t \bar{a}_i.$$

But our choice of p was such that the condition of the claim above is not satisfied. So we can conclude that μ and ν give us the desired result.

⁷For brevity I'm assuming here that the minima for each column are unique. It is immediate to see how the argument should be generalized to take care of multiple minima.

Remark 3 It should be observed that in fact 1 I used a belief function and a probability measure. Thus the example can also be taken to show that total monotonicity of the two capacities is not enough to prove a Fubini result for functions which are not slice-comonotonic. Actually something even stronger is true. Given a function with, say, non-comonotonic x -sections one could construct a different⁸ counterexample in which both marginals have a support of two points (where by support I mean the smallest subset with weight 1), the marginal on X is additive and the marginal on Y is almost additive, in the sense that all but $\epsilon > 0$ weight is divided among the two points, and yet the iterated integrals do not coincide. This construction would work whatever the size of the space, so it can be used to prove the validity of the statement in fact 1 also in infinite spaces.

As a consequence of fact 1 and remark 3 we have the following immediate strengthening of lemma 3.

Corollary 3.2 *Equation (14) will hold for every pair of capacities μ and ν if and only if the (bounded and \mathcal{A}_Z -measurable) f is slice-comonotonic.*

Coming to lemma 5, given our brief discussion in remark 2 it is not surprising to see that the assumption that f has comonotonic x -sections is not necessary: Just use the capacities in example 2 and notice that $\sigma = u_Z$ satisfies the Fubini property. However we can prove a result analogous to corollary 3.2 by modifying slightly the proof of fact 1 to show that for every f without comonotonic x -sections one can find a pair of capacities and a Fubini-independent product such that equation (18) does not hold.⁹

Likewise, the assumption that σ satisfies the Fubini property is not necessary to insure that the equation

$$\int_Z f(x, y) d\sigma(x, y) = \int_Y \int_X f(x, y) d\mu(x) d\nu(y) \quad (25)$$

holds for every f with comonotonic x -sections. One immediately sees this by considering the case where f is a constant and equal, say, to some value $c \in \mathbf{R}$. Then, whatever σ , (25) will obviously hold as both sides will be equal to c by properties (i) and (ii) in proposition 2. Again, however, we can prove that the satisfaction of the Fubini property is necessary in the following weak sense: Suppose that σ is not Fubini-independent. Then there is a function f with comonotonic x -sections such that its integral with respect to σ is different from the iterated integral. To see that, let $A \in \mathcal{A}_Z$ be a comonotonic set such that

$$\sigma(A) \neq \int_Y \int_X 1_A(x, y) d\mu(x) d\nu(y),$$

and take $f = 1_A$. Clearly f has the desired properties. We have thus proved the following stronger version of lemma 5.

⁸And definitely less illuminating. This is why I chose to present instead the construction in fact 1.

⁹If μ and ν are defined as in that proof one can show (see theorem 5.2 below) that there is only one belief function σ which satisfies the Fubini property (see definition 8), and the integral with respect to that σ will be equal to the right-hand side of equation (22).

Corollary 3.3 *If σ is a product capacity on (Z, \mathcal{A}_Z) , with marginals μ and ν , equation (18) will hold for every function $f : Z \rightarrow \mathbf{R}$ bounded, \mathcal{A}_Z -measurable and with comonotonic x -sections, if and only if σ satisfies the Fubini property.*

Another interesting question to be posed is: How large is the class of slice-comonotonic functions? The following proposition might help answer this question. The notation is the same used in section 2.

Proposition 5 (Denneberg, [4] Proposition 4.5) *For two bounded functions $f, g : \Omega \rightarrow \mathbf{R}$, the following are equivalent:*

- (i) *f and g are comonotonic;*
- (ii) *There exists a bounded function $h : \Omega \rightarrow \mathbf{R}$ and non-decreasing functions $u, v : \mathbf{R} \rightarrow \mathbf{R}$ such that*

$$f = u \circ h \quad \text{and} \quad g = v \circ h.$$

From this we immediately have that if f is slice-comonotonic then for every $x \in X$ (resp. $y \in Y$) $f(x, \cdot)$ (resp. $f(\cdot, y)$) will be a monotone transformation of some function $g : Y \rightarrow \mathbf{R}$ (resp. $h : X \rightarrow \mathbf{R}$), or equivalently all the $f(x, \cdot)$ (resp. $f(\cdot, y)$) will be monotone transformations of each other. This in particular implies that any f which is monotone in each argument will be slice-comonotonic.

Hence we can conclude that the class of slice-comonotonic functions is quite large, and it certainly includes many functions used in economic applications. However the negative result presented in fact 1, and the discussion in remark 3, also have the relevant implication that violations of slice-comonotonicity cannot be taken lightly, as they show that small amounts of non-additivity in beliefs can bring forth violations of Fubini's theorem.

3.3 An Interesting Implication

There is a consequence of the Fubini theorem which, though almost obvious, is quite relevant to applications in different fields, and therefore deserves spending a few words on. Imagine having a "sequence" of two spaces, (X, \mathcal{A}_X) and (Y, \mathcal{A}_Y) , and let f and g be two real-valued functions, respectively defined on X and Y (that is, two random variables). We can extend f to $\bar{f} : Z \rightarrow \mathbf{R}$ and g to $\bar{g} : Z \rightarrow \mathbf{R}$ in the obvious way, i.e., for every $(x, y) \in Z$ let

$$\bar{f}(x, y) = f(x), \quad \bar{g}(x, y) = g(y).$$

Suppose that μ and ν are defined on X and Y respectively, and let σ be a product capacity which has them as marginals. First of all we might be interested in the “expected value” of \bar{f} and \bar{g} . We have

$$\int_Z \bar{f}(x, y) d\sigma(x, y) = \int_X f(x) d\mu(x),$$

and

$$\int_Z \bar{g}(x, y) d\sigma(x, y) = \int_Y g(y) d\nu(y).$$

Now we might be interested in saying something about the expectation of $\bar{f} + \bar{g}$, that is, the function on Z which is defined by $(\bar{f} + \bar{g})(x, y) = f(x) + g(y)$. This is the topic of the following result. Clearly it can immediately be generalized to any finite sum of random variables.

Corollary 3.4 *If σ is a product capacity satisfying the Fubini property then*

$$\int_Z (\bar{f} + \bar{g})(x, y) d\sigma(x, y) = \int_X f(x) d\mu(x) + \int_Y g(y) d\nu(y). \quad (26)$$

Proof: $\bar{f} + \bar{g}$ is slice-comonotonic, hence theorem 3.1 and two applications of property (iv) of proposition 2 imply that

$$\begin{aligned} \int_Z (\bar{f} + \bar{g})(x, y) d\sigma(x, y) &= \int_Z (f(x) + g(y)) d\sigma(x, y), \\ &= \int_Y \int_X (f(x) + g(y)) d\mu(x) d\nu(y) \\ &= \int_Y [(\int_X f(x) d\mu(x)) + g(y)] d\nu(y) \\ &= \int_X f(x) d\mu(x) + \int_Y g(y) d\nu(y). \end{aligned}$$

This suggests that the notion of Fubini independence has very important applications. Thus, for instance, if two random variables are independent in the sense that their distributions satisfy the Fubini property, then the expectation of their sum is equal to the sum of the expectations. Thus my expectation of profit from two investments which are Fubini-independent will be the sum of the expected profits taken separately, even if my beliefs about these investments are represented by capacities.

As we did above, we might wonder to what extent the assumption of Fubini independence is necessary to prove this result. The following simple example will help understand the problem.

Example 3 In the set-up and notation of example 1, assume that $f(x_1) = 4$, $f(x_2) = 1$ and $g(y_1) = 2$, $g(y_2) = 0$. Then $(\bar{f} + \bar{g})$ has the following result matrix

	y_1	y_2
x_1	6	4
x_2	3	1

Let $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1/3$. Suppose that $\sigma(A) = 1/9$, $\sigma(A \cup B) = 1/3$ and $\sigma(A \cup B \cup C) = 1/3$. Then σ can be (extended to) an independent product, but it violates the Fubini property (it would satisfy it only if $\sigma(A \cup B \cup C) = 5/9$). We have that

$$\int_X f(x) d\mu(x) = 4(1/3) + (2/3) = 2,$$

and

$$\int_Y g(y) d\nu(y) = 2(1/3),$$

so that their sum is $2(1 + 1/3)$. On the other hand

$$\int_Z (\bar{f} + \bar{g})(x, y) d\sigma(x, y) = 2(1/9) + 1(1/3) + 2(1/3) + 1 = 2(1 + 1/9).$$

It turns out that, once more, Fubini independence is necessary in a weak sense for corollary 3.4. That is, if σ is an independent product which is not Fubini-independent, then one can find two functions f and g such that (26) fails for $\bar{f} + \bar{g}$.

Proposition 6 *Suppose that σ is an independent product of μ and ν which does not have the Fubini property. Then there are $f : X \rightarrow \mathbf{R}$ and $g : Y \rightarrow \mathbf{R}$ such that (26) fails.*

Proof. I will just sketch the argument here, leaving it to the interested reader to fill in the (rather tedious) details. Suppose that $A \in \mathcal{A}_Z$ is a comonotonic set such that (12) fails. Since A is product measurable, it can be written as a disjoint union of finitely many rectangles. Hence we can focus on constructing two simple functions f and g , which implies that it is without any loss in generality to prove the result for the case of finite X and Y . As a consequence we can also assume that A is smallest, in the sense that all subsets of A (in \mathcal{A}_Z) are such that (12) holds. We can then use the fact that A is comonotonic to reorder X and Y so that the x -sections of A form a decreasing class (in the order on X), and the same for the y -sections. Now it is quite simple to see that (as we did in example 3) we can construct two nonincreasing functions f and g such that A is a level set of $\bar{f} + \bar{g}$, and such that the integral with respect to σ does not coincide with the iterated integral.

4 Products of Convex Capacities

In this and in the following section we examine what kind of conclusions can be obtained when we impose stronger requirements on either marginals or their products. Specifically,

we will discuss here the case of convex capacities, and belief functions in the next section.¹⁰

First of all let me review some interesting results about convex capacities,¹¹ i.e., capacities which satisfy properties (i)-(iii).

Definition 6 Let σ be a capacity on (Ω, \mathcal{A}) , the core of σ , denoted $\mathcal{C}(\sigma)$, is the set of all (finitely additive) probabilities on (Ω, \mathcal{A}) which pointwise dominate σ , i.e.,

$$\mathcal{C}(\sigma) \equiv \{P : P \text{ is a probability on } (\Omega, \mathcal{A}), P(A) \geq \sigma(A) \text{ for all } A \in \mathcal{A}\}.$$

In general the core of a capacity might be empty, but Shapley [35] proved that every convex capacity has a non-empty core. Also Schmeidler provided the following nice characterization of the Choquet integral of a function on a convex capacity.

Proposition 7 (Schmeidler, [31] Proposition) Given any bounded \mathcal{A} -measurable function $f : \Omega \rightarrow \mathbf{R}$, σ is a convex capacity on (Ω, \mathcal{A}) if and only if

$$\int_{\Omega} f(\omega) d\sigma(\omega) = \min_{P \in \mathcal{C}(\sigma)} \int_{\Omega} f(\omega) dP(\omega). \quad (27)$$

If Ω is a finite set (say $\Omega = \{\omega_1, \dots, \omega_n\}$) then something more can be said about the shape of $\mathcal{C}(\sigma)$. Consider the set Π of all the $n!$ possible permutations of $\{1, \dots, n\}$. For every $\pi \in \Pi$ let P_{π} be the additive measure defined by

$$P_{\pi}(\omega_{\pi(i)}) = \sigma(\{\omega_{\pi(1)}, \dots, \omega_{\pi(i)}\}) - \sigma(\{\omega_{\pi(1)}, \dots, \omega_{\pi(i-1)}\}), \text{ for all } i = 1, \dots, n.$$

Shapley proved [35] that $\mathcal{C}(\sigma)$ is the convex hull of the P_{π} for $\pi \in \Pi$. Going back to formula (6) we can notice that the measure P , equivalent to σ for integration purposes, which we obtained in lemma 1 is just one of the P_{π} . So we can interpret Schmeidler's result as saying that the Choquet integral with respect to a convex σ is an operator which selects to integrate a function f with respect to the measure in the class $\{P_{\pi}\}_{\pi \in \Pi}$ which minimizes the integral. In this sense we say that the Choquet integral with respect to a convex capacity reflects a very "pessimistic" attitude.

Coming back to our product space set-up, one can immediately verify that in general there is more than one *convex* independent product of two convex capacities μ and ν . An interesting question is whether there are many Fubini-independent convex products. The example below shows that the answer is also yes.

Example 4 Going back again to example 1, suppose that $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1/3$. As we know, all Fubini independent products will be identical on comonotonic sets. Hence any *convex* Fubini independent product σ will be, say, equal to $5/9$ on $A \cup B \cup C$ or to $1/3$ on $A \cup C$. But one can easily check that any value between $2/9$ and $1/3$ for $\sigma(A \cup D)$ and $\sigma(B \cup C)$ can be assumed by any such capacity.

¹⁰Remember lemma 2 implies that requiring the product to have either property will automatically imply it for both marginals.

¹¹For other results on convex capacities and belief functions see Chateauneuf and Jaffray [1].

We might try to obtain a unique independent product by invoking considerations different from the ones we discussed so far. For instance, Gilboa and Schmeidler in [17] suggested that for convex capacities the following notion seems intuitively appealing.¹²

Definition 7 Let μ and ν be two convex capacities on (X, \mathcal{A}_X) and (Y, \mathcal{A}_Y) respectively. We call their core-independent product the set-function σ_C defined as follows:

$$\sigma_C \equiv \inf\{P \times Q : P \in \mathcal{C}(\mu), Q \in \mathcal{C}(\nu)\}$$

that is, σ_C is the lower envelope of the set of product measures generated by choosing a probability from each of the cores of the marginals. Denote by $\mathcal{C}(\mu) \times \mathcal{C}(\nu)$ such set.

Notice that since both $\mathcal{C}(\mu)$ and $\mathcal{C}(\nu)$ are weak* compact by well-known arguments,¹³ the infimum in the definition is attained, and we can write “minimum” instead. As a side remark, notice that $\mathcal{C}(\mu) \times \mathcal{C}(\nu)$ is generally smaller than the core of σ_C , as the latter will contain also measures which allow for correlation. As it turns out, σ_C is a Fubini-independent convex product capacity.

Proposition 8 Let μ and ν be two convex capacities respectively defined on (X, \mathcal{A}_X) and (Y, \mathcal{A}_Y) , then σ_C is a convex independent product capacity on (Z, \mathcal{A}_Z) . Moreover it has the Fubini property.

Proof: The first part descends easily from the definition and proposition 7. For the second, I prove the statement by showing that σ_C must satisfy equation (18) for a slice-comonotonic f . As we observed above the probabilities P and Q which we obtained in the proof of lemma 3 respectively belong to $\mathcal{C}(\mu)$ and $\mathcal{C}(\nu)$, so that $R = P \times Q \in \mathcal{C}(\mu) \times \mathcal{C}(\nu)$. Thus by the reasoning in the proof of lemma 5 $\int \int f d\mu d\nu = \int f dR$ from which we get

$$\int_Z f(x, y) d\sigma_C \leq \int_Y \int_X f(x, y) d\mu(x) d\nu(y).$$

Now we have to prove that strict inequality cannot occur. Suppose it did, then there must be P' and Q' , respectively in $\mathcal{C}(\mu)$ and $\mathcal{C}(\nu)$, such that if $R' = P' \times Q'$,

$$\begin{aligned} \int_Y \int_X f(x, y) dP'(x) dQ'(y) &= \int_Z f(x, y) dR'(x, y) \\ &< \int_Z f(x, y) dR(x, y) \\ &= \int_Y \int_X f(x, y) dP(x) dQ(y). \end{aligned}$$

For every $y_j \in Y$ let $G(y_j)$ be as defined in the proof of lemma 3. Let

$$G'(y_j) \equiv \int_X f(x, y_j) dP'(x).$$

¹²The same notion had been proposed some years earlier by Walley and Fine [37].

¹³Remember that we define the core to be a set of *finitely* additive measures. The last statement would be false if we restricted to countably additive measures.

Since μ is a convex capacity we must have that $G(y_j) \leq G'(y_j)$ for every $y_j \in Y$. So property (iii) of proposition 2 and the fact that ν is a convex capacity imply that

$$\int_Y G(y) dQ(y) = \int_Y G(y) d\nu(y) \leq \int_Y G'(y) d\nu(y) \leq \int_Y G'(y) dQ'(y),$$

and we get a contradiction.

This result provides an explanation of some results of Dow and Werlang and Marinacci. Dow and Werlang argue ([7] in the example on page 10) that if product beliefs are core-independent then the expectation of a sum of random variables is the sum of the expectations. This is an immediate consequence of the fact that core independence implies Fubini independence and corollary 3.4. Analogously Marinacci's theorem 4.6.1 in [28] goes through for exactly the same reason.

5 Products of Belief Functions

5.1 The Möbius-Independent Product Belief Function

Stronger conclusions can be drawn once we impose the additional requirement that the product capacity be a belief function. By lemma 2 this implies that both marginals are belief functions. This might happen either if we are given a product belief function to start with, or if we are given a couple of marginals which are belief functions, and we are interested in an independent product which is a belief function itself. As it is tradition in the literature on belief functions, except where otherwise noted in this subsection I will assume that X and Y (hence Z) are finite and that $\mathcal{A}_X = 2^X$ and $\mathcal{A}_Y = 2^Y$ (hence $\mathcal{A}_Z = 2^Z$).

The following notion of independent product of two belief functions μ and ν was suggested by Walley and Fine [37] and Hendon *et al.* [20].

Definition 8 *Given a belief function μ on (X, \mathcal{A}_X) and a belief function ν on (Y, \mathcal{A}_Y) , with Möbius transforms χ and ψ respectively,¹⁴ let σ_M be the unique capacity on (Z, \mathcal{A}_Z) which has Möbius transform φ defined as follows, for every $A \in \mathcal{A}_Z$,*

$$\varphi(A) = \begin{cases} \chi(S)\psi(T) & \text{if } A = S \times T \text{ for } S \in \mathcal{A}_X, T \in \mathcal{A}_Y, \\ 0 & \text{otherwise.} \end{cases}$$

σ_M will be called the Möbius-independent product of μ and ν .

¹⁴Which exist by proposition 1. For obvious reasons I prefer this to the more cumbersome notation φ_μ and φ_ν .

One can easily check that σ_M thus defined is indeed an independent product belief function, i.e., it satisfies properties (i)-(ii), (iii') and (iv). This definition can be generalized to the infinite case using the infinite space generalization of proposition 1 due to Gilboa and Schmeidler.

Remark 4 One might be tempted to use the definition also for any pair μ and ν (I was, at least). However Rainer Dyckerhoff provided me with a nice example which shows that if μ and ν are not belief functions then σ_M will in general violate monotonicity, so it will not be a capacity.

There is a nice decision-theoretic explanation of the specific form of σ_M . I remarked earlier that the Möbius transform of a belief function is non-negative, and it is a simple consequence of equation (4) that it is actually a probability measure on the relevant set of subsets. As Shafer suggested in [32], one can then interpret $\varphi(A)$ as the *specific weight* which is assigned to A and cannot be further divided among its subsets. For instance if σ is a probability measure all such weight is divided among singletons and no weight is allocated to larger sets, so that in a sense we can interpret a probability as an epistemic status in which beliefs are extremely precise and specified. σ_M assigns *no* specific weight to non-rectangles. Hence we can interpret σ_M as reflecting a situation in which the independence of μ and ν is beyond any doubt, or, in a terminology which is widely used by decision theorists, there is no “uncertainty” about their independence. Other justifications can be given in some special cases, but they will be discussed in subsection 5.2. Of course, one can argue that the results to follow provide by themselves a justification of Möbius independence.

Hendon *et al.* present some interesting results about σ_M . First of all they show that we can find an exact formula for σ_M in terms of μ and ν . For every $A \in \mathcal{A}_Z$ let \mathcal{J}_A be the set of all possible *families* of cartesian subsets of A . That is, abusing notation somewhat, every $J \in \mathcal{J}_A$ is a set $\{X_j \times Y_j\}_{j \in J}$ with $X_j \in \mathcal{A}_X$ and $Y_j \in \mathcal{A}_Y$. Then HJST prove [20, Proposition 3] that for all $A \in \mathcal{A}_Z$,

$$\sigma_M(A) = \max_{J \in \mathcal{J}_A} \sum_{I \subseteq J, I \neq \emptyset} (-1)^{|I|+1} \mu(\cap_{i \in I} X_i) \nu(\cap_{i \in I} Y_i). \quad (28)$$

Using this fact, HJST proceed to show that σ_M has a very interesting property: it is the *smallest* of all the independent product belief functions.

Proposition 9 (HJST, [20] Proposition 3) *Let σ be the Möbius-independent product of the belief functions μ and ν . Then for every other independent product belief function λ of μ and ν we have*

$$\lambda \geq \sigma_M.$$

Obviously, by property (vi) of proposition 2, this immediately implies that for *all* functions f we have

$$\int_Z f d\lambda \geq \int_Z f d\sigma_M. \quad (29)$$

Notice that we cannot immediately draw a similar conclusion for σ_C , because even if μ and ν are belief functions σ_C is not necessarily a belief function. Yet we can use the result to establish that $\mathcal{C}(\sigma) \supseteq \mathcal{C}(\mu) \times \mathcal{C}(\nu)$. In fact every $R \in \mathcal{C}(\mu) \times \mathcal{C}(\nu)$ is an independent product belief function, so that $R \geq \sigma$. Given that R is also a probability we immediately have that $R \in \mathcal{C}(\sigma)$ as well, proving the claim. Using this we can conclude :

Corollary 5.1 *If the marginals μ and ν are belief functions then*

$$\sigma_C \geq \sigma_M,$$

so that (29) also holds for σ_C .

Proof: By definition $\sigma_C = \inf\{R : R \in \mathcal{C}(\mu) \times \mathcal{C}(\nu)\}$, while by proposition 7 we have $\sigma_M = \inf\{R : R \in \mathcal{C}(\sigma_M)\}$. The result then follows from $\mathcal{C}(\sigma_M) \supseteq \mathcal{C}(\mu) \times \mathcal{C}(\nu)$.

The next step is clearly checking whether σ_M has the Fubini property.¹⁵

Proposition 10 *Let σ_M be the Möbius-independent product of belief functions μ and ν . Then σ_M is Fubini-independent.*

Proof: I prove the equivalent statement that σ_M satisfies equation (18) for every f with comonotonic x -sections. Given that f has comonotonic x -sections it is possible to relabel the elements of Y as in the first part of the proof of lemma 3. Let \mathcal{C}_Z be the set of all rectangles in \mathcal{A}_Z . By the definition of σ and proposition 3 we have

$$\begin{aligned} \int_Z f(x, y) d\sigma_M(x, y) &= \sum_{A \in \mathcal{A}_Z} \varphi(A) \left[\min_{(x, y) \in A} f(x, y) \right] \\ &= \sum_{A \in \mathcal{C}_Z} \varphi(A) \left[\min_{(x, y) \in A} f(x, y) \right] \\ &= \sum_{(S, T) \in \mathcal{A}_X \times \mathcal{A}_Y} \chi(S) \psi(T) \left[\min_{x \in S} f(x, y_{m(T)}) \right] \\ &= \sum_{T \in \mathcal{A}_Y} \psi(T) \sum_{S \in \mathcal{A}_X} \chi(S) \left[\min_{x \in S} f(x, y_{m(T)}) \right], \end{aligned} \quad (30)$$

where $m(T)$ is the largest index j such that $y_j \in T$. On the other hand

$$\begin{aligned} \int_Y \int_X f(x, y) d\mu(x) d\nu(y) &= \sum_{T \in \mathcal{A}_Y} \psi(T) \min_{y \in T} \left(\int_X f(x, y) d\mu(x) \right) \\ &= \sum_{T \in \mathcal{A}_Y} \psi(T) \min_{y \in T} \left(\sum_{S \in \mathcal{A}_X} \chi(S) \left[\min_{x \in S} f(x, y) \right] \right) \\ &= \sum_{T \in \mathcal{A}_Y} \psi(T) \sum_{S \in \mathcal{A}_X} \chi(S) \left[\min_{x \in S} f(x, y_{m(T)}) \right]. \end{aligned} \quad (31)$$

¹⁵This result can be easily extended to infinite products using the same technique that we saw in the proof of lemmas 3 and 5.

Now we conclude by noticing that the right-hand side of (30) is the same as the right-hand side of (31).

But something more is true. Surprisingly, it turns out that σ_M is the *only* independent product belief function with the Fubini property. Some additional terminology will help in proving this result.

Definition 9 *Suppose that $A \in \mathcal{A}_Z$. Then we call its comonotonic hull (and label $C(A)$) a smallest comonotonic set which contains A . That is, $C(A) \supseteq A$ and for all comonotonic $B \supseteq A$ we have $C(A) \subseteq B$.*

Obviously I am abusing notation, as there is more than one comonotonic hull for non-comonotonic sets. One could however write down an explicit procedure to obtain a specific comonotonic hull for all such sets A and define $C(A)$ to be the result of such procedure, so as to make the definition rigorous. Notice that $C(A) = A$ for all comonotonic sets, and in particular $C(A)$ can be a rectangle if and only if A is rectangle. Now we can state and prove:

Theorem 5.2 *Suppose that σ_M is the Möbius-independent product of belief functions μ and ν . Let λ be any other independent product belief function. Then if $\sigma_M(A) = \lambda(A)$ for all comonotonic sets A it must be the case that $\sigma_M = \lambda$. In other words, σ_M is the only Fubini-independent product belief function.*

Proof: As usual let ξ and φ be the Möbius transforms of λ and σ respectively. The proof is by induction on the cardinality of the comonotonic hull. If A has $|C(A)| = 1$ then clearly $\xi(A) = \varphi(A)$ by the fact that λ is an independent product.

Suppose that $\xi(A) = \varphi(A)$ for all A such that $|C(A)| \leq n - 1$. Let A have $|C(A)| = n$ and suppose that $C(A)$ is not a rectangle (that is, A is not a rectangle). I want to prove that $\xi(A) = \varphi(A)$: Since λ and σ coincide on all comonotonic sets we must have $\lambda(C(A)) = \sigma(C(A))$, that is,

$$\sum_{B \subseteq C(A)} \xi(B) = \sum_{B \subseteq C(A)} \varphi(B),$$

which, given the induction hypothesis, amounts to

$$\sum_{\substack{B \subseteq C(A) \\ |C(B)| = n}} \xi(B) = \sum_{\substack{B \subseteq C(A) \\ |C(B)| = n}} \varphi(B).$$

But any such B cannot be a rectangle, for no rectangle with cardinality n can be contained in $C(A)$ (and we assumed that $C(A)$ was not a rectangle). Hence, since we have $\xi(B) \geq 0 = \varphi(B)$ for all B to which the sum above is extended, we can only have equality if $\xi(B) = \varphi(B)$ for all such B , including A itself.

If on the other hand $C(A)$ is a rectangle, then $A = C(A)$ and we must have $\lambda(A) = \sigma(A)$. For all $B \subseteq A$ such that $B \neq A$ we must have $|C(B)| \leq n - 1$, hence the induction hypothesis implies $\xi(B) = \varphi(B)$. This in turn implies $\xi(A) = \varphi(A)$.

Thus we see that, in the finite case,¹⁶ uniqueness can be obtained when we require the product to be a belief function with the Fubini property.

5.2 Another Story for Möbius independence

Möbius independence has an arguably natural justification when the belief functions μ and ν are the distributions induced by random correspondences. To avoid excessively burdening the exposition with additional technical detail (connected with the generalization of proposition 1), I shall again assume that X and Y are finite. However all the results to follow can be generalized immediately to the infinite case.

Let a probability space (Ω, \mathcal{A}, P) be given, X and Y be two finite sets, and suppose that $\mathbf{M} : \Omega \rightarrow \mathcal{A}_X$ and $\mathbf{N} : \Omega \rightarrow \mathcal{A}_Y$ are two \mathcal{A} -measurable *correspondences* (i.e., multi-valued functions).

Definition 10 For every \mathcal{A} -measurable correspondence $\mathbf{L} : \Omega \rightarrow Z$, where Z is a finite set, call the distribution of \mathbf{L} on Z the set-function λ on \mathcal{A}_Z defined as follows, for all $A \subseteq Z$,

$$\lambda(A) \equiv P(\{\omega \in \Omega : \mathbf{L}(\omega) \subseteq A\}). \quad (32)$$

Such distribution will be a probability if and only if the correspondence \mathbf{L} is a *function*. Otherwise it is going to be a belief function. This is easily seen by remarking that the Möbius transform of λ is just: for $A \subseteq Z$,

$$\xi(A) = P(\{\omega \in \Omega : \mathbf{L}(\omega) = A\}) \geq 0. \quad (33)$$

Label μ and ν the distributions of \mathbf{M} and \mathbf{N} respectively, and let χ and ψ be their respective Möbius transforms.

As it is well-known, when \mathbf{M} and \mathbf{N} are functions (that is, they are random variables) we say that they are *independent* if for every pair¹⁷ $S \subseteq X, T \subseteq Y$,

$$P(\mathbf{M} \in S, \mathbf{N} \in T) = P(\mathbf{M} \in S) P(\mathbf{N} \in T). \quad (34)$$

When we are discussing random correspondences rather than random variables, this definition does not make sense because the event $\mathbf{M} \in S$ (or $\mathbf{N} \in T$) can be undefined. So the question is how to extend (34) to random correspondences. The following extension immediately springs to mind.

¹⁶Unfortunately I do not see any obvious way of extending the proof of this result to products of spaces with cardinality higher than denumerable.

¹⁷As usual, for a random variable f I write $P(f \in A)$ rather than the more cumbersome $P(\{\omega \in \Omega : f(\omega) \in A\})$.

Definition 11 *Suppose that \mathbf{M} and \mathbf{N} are random correspondences as defined above. Then we say that they are independent if, for every pair $S \subseteq X, T \subseteq Y$,*

$$P(\mathbf{M} \subseteq S, \mathbf{N} \subseteq T) = P(\mathbf{M} \subseteq S) P(\mathbf{N} \subseteq T). \quad (35)$$

It is immediate to see which consequences this notion of independence has for the distribution λ of the vector random correspondence (\mathbf{M}, \mathbf{N}) on $Z = X \times Y$. If, in fact, λ is defined as in (32) independence of \mathbf{M} and \mathbf{N} implies that for all $S \subseteq X, T \subseteq Y$,

$$\lambda(S \times T) = P(\mathbf{M} \subseteq S, \mathbf{N} \subseteq T) = P(\mathbf{M} \subseteq S) P(\mathbf{N} \subseteq T) = \mu(S) \nu(T).$$

That is, the product belief function will be independent in the sense of definition 2. At this point the reader will be well aware that such λ is not uniquely defined in this fashion. In the terms of this section we need to have additional information of the joint behavior of \mathbf{M} and \mathbf{N} to define λ uniquely on non-cartesian subsets of Z .

But one could also consider a stronger but natural extension, which basically amounts to noticing that correspondences are functions into the set of all subsets of the range, and then applying the standard definition.

Definition 12 *Suppose that \mathbf{M} and \mathbf{N} are random correspondences as defined above. Then we say that they are strongly independent if, for every pair $\mathcal{S} \subseteq \mathcal{A}_X, \mathcal{T} \subseteq \mathcal{A}_Y$,*

$$P(\mathbf{M} \in \mathcal{S}, \mathbf{N} \in \mathcal{T}) = P(\mathbf{M} \in \mathcal{S}) P(\mathbf{N} \in \mathcal{T}). \quad (36)$$

Clearly strong independence implies independence: in fact for $S \subseteq X$ let

$$\mathcal{S} = \{U \in \mathcal{A}_X : U \subseteq S\}$$

(and analogously do for T) and substitute it in (36), so that (35) is obtained. If instead we take $\mathcal{S} = \{S\}$ and $\mathcal{T} = \{T\}$ we get

$$P(\mathbf{M} = S, \mathbf{N} = T) = \chi(S) \psi(T). \quad (37)$$

Thus we have the following natural result, which can be used to argue that the notion of strong independence for random correspondences provides another justification for the definition of Möbius-independent product belief function.

Proposition 11 *Suppose that random correspondences \mathbf{M} and \mathbf{N} are strongly independent. Let σ be the distribution on Z induced by the vector (\mathbf{M}, \mathbf{N}) . Then σ is equal to σ_M , the Möbius-independent product of μ and ν , which are the marginal distributions of \mathbf{M} and \mathbf{N} respectively.*

Proof. As usual let φ be the Möbius transform of σ . By (37) we clearly have that for every rectangle $A \in \mathcal{A}_Z$ such that $A = S \times T$ for some $S \subseteq X, T \subseteq Y$,

$$\varphi(A) = \chi(S)\psi(T).$$

One then just needs to notice that this implies that φ will have to be identically zero on non-rectangles. In fact, if \mathcal{C}_Z is the set of all rectangle subsets of Z ,

$$\begin{aligned} \sigma(Z) &= \sum_{A \in \mathcal{A}_Z} \varphi(A) \\ &= \sum_{A \in \mathcal{C}_Z} \varphi(A) + \sum_{A \in \mathcal{A}_Z \setminus \mathcal{C}_Z} \varphi(A) \end{aligned}$$

so that

$$\sum_{A \in \mathcal{A}_Z \setminus \mathcal{C}_Z} \varphi(A) = 1 - \sum_{A \in \mathcal{A}_Z} \varphi(A).$$

But

$$\begin{aligned} \sum_{A \in \mathcal{A}_Z} \varphi(A) &= \sum_{S \in \mathcal{A}_X} \sum_{T \in \mathcal{A}_Y} \chi(S)\psi(T) \\ &= \sum_{S \in \mathcal{A}_X} \chi(S) \sum_{T \in \mathcal{A}_Y} \psi(T) \\ &= \sum_{S \in \mathcal{A}_X} \chi(S) \\ &= 1 \end{aligned}$$

which immediately implies the claim, given that $\varphi \geq 0$.

In particular strong independence arises naturally when the belief functions μ and ν are the distributions induced by random correspondences on *independent* probability spaces. Suppose that $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$ are given. Let $\Omega = \Omega_1 \times \Omega_2$ be equipped with the product algebra $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$, and with the probability measure given by the independent product $P = P_1 \times P_2$. Suppose that there are correspondences $\mathbf{M} : \Omega_1 \rightarrow \mathcal{A}_X$ and $\mathbf{N} : \Omega_2 \rightarrow \mathcal{A}_Y$, with respective distributions given by μ and ν , that can be extended to Ω as $\bar{\mathbf{M}}$ and $\bar{\mathbf{N}}$ in the usual way (analogously to what we did for f and g in subsection 3.3). We might be interested in saying something about the distribution σ on $Z = X \times Y$ induced by the correspondence $(\bar{\mathbf{M}}, \bar{\mathbf{N}})$. Unsurprisingly, the hypothesis of independence on the product measure implies that $\bar{\mathbf{M}}$ and $\bar{\mathbf{N}}$ are strongly independent, so that σ will be the Möbius-independent product of μ and ν .

Corollary 5.3 *Let σ be the distribution on (Z, \mathcal{A}_Z) induced by $(\bar{\mathbf{M}}, \bar{\mathbf{N}})$, that is, for every $A \in \mathcal{A}_Z$,*

$$\sigma(A) = P(\{\omega \in \Omega : (\bar{\mathbf{M}}(\omega_1), \bar{\mathbf{N}}(\omega_2)) \subseteq A\}).$$

Then $\sigma = \sigma_M$, the Möbius-independent product of μ and ν .

Proof: Just notice that for every pair $\mathcal{S} \subseteq \mathcal{A}_X$, $\mathcal{T} \subseteq \mathcal{A}_Y$,

$$\begin{aligned} P(\bar{\mathbf{M}} \in \mathcal{S}, \bar{\mathbf{N}} \in \mathcal{T}) &= P_1(\mathbf{M} \in \mathcal{S}) P_2(\mathbf{N} \in \mathcal{T}) \\ &= P(\bar{\mathbf{M}} \in \mathcal{S}) P(\bar{\mathbf{N}} \in \mathcal{T}), \end{aligned}$$

so that $\bar{\mathbf{M}}$ and $\bar{\mathbf{N}}$ are strongly independent. Proposition 11 can then be used to obtain the result.

5.3 Some More Unicity

In subsection 5.1 we saw that unicity of the product can be obtained if we impose the following requirements: 1) the product is a belief function, 2) the product is Fubini-independent. There is another interesting case in which unicity can be obtained only by requiring the product to be independent, as well as a belief function. This is when one of the marginals, say ν , is a probability measure. Once again, the results to be presented in this section could be proved in full generality, but I will present only the finite version for brevity.

The following lemma, due to Eichberger and Kelsey [11], is the key. It says that if a product belief function has an additive marginal then it must satisfy a certain type of additivity itself.

Lemma 6 *For finite X and Y , let σ be a product belief function on $Z = X \times Y$. Then its marginal on Y , labelled ν , is a probability measure if and only if for every $A \in \mathcal{A}_Z$, if for every $y \in Y$ we let $A_y = \{z \in A : z = (x, y) \exists x \in X\}$ (i.e., the y -section of A), we have*

$$\sigma(A) = \sum_{y \in Y} \sigma(A_y). \quad (38)$$

Proof: Necessity is obvious. For sufficiency, let φ be the Möbius transform of σ . On the one hand we clearly have

$$1 = \sigma(Z) = \sum_{B \subseteq Z} \varphi(B),$$

but we also have

$$1 = \sum_{y \in Y} \nu(\{y\}) = \sum_{y \in Y} \sigma(X \times \{y\}) = \sum_{y \in Y} \sum_{B \subseteq X \times \{y\}} \varphi(B).$$

Hence

$$\sum_{B \subseteq Z} \varphi(B) = \sum_{y \in Y} \sum_{B \subseteq X \times \{y\}} \varphi(B),$$

from which we conclude that for every $B \in \mathcal{A}_Z$ which is not contained in some y -section will have $\varphi(B) = 0$. This immediately implies that for every $A \in \mathcal{A}_Z$

$$\begin{aligned}\sigma(A) &= \sum_{y \in Y} \sum_{B \subseteq A \cap (X \times \{y\})} \varphi(B), \\ &= \sum_{y \in Y} \sigma(A_y).\end{aligned}$$

It takes little effort to construct counterexamples to (38) where σ has an additive marginal but is not a belief function. Even convexity can be shown to be insufficient. We can now prove the announced unicity result.

Theorem 5.4 *Suppose that belief functions μ on (X, \mathcal{A}_X) and ν on (Y, \mathcal{A}_Y) are given, and ν is additive on \mathcal{A}_Y . Then there is only one independent product belief function σ , and it satisfies the Fubini property.*

Proof: Let σ be an independent product belief function. By independence ν is a marginal of σ . Hence lemma 6 implies that for every $A \in \mathcal{A}_Z$,

$$\sigma(A) = \sum_{y \in Y} \sigma(A_y) = \sum_{y \in Y} \mu(A_y^X) \nu(\{y\}), \quad (39)$$

where $A_y^X = \{x \in X : x \in A_y\}$ (i.e., the X -projection of A_y). The second equality in (39) follows by independence since $A_y = A_y^X \times \{y\}$. Clearly (39) implies that σ is uniquely defined, and it is immediate to verify that it also implies that σ has the Fubini property.

Remark 5 As σ is Fubini-independent, theorem 5.2 can be used to conclude that $\sigma = \sigma_M$. That is, when one of the marginals is additive σ_M is the only independent product belief function.

5.4 A Decision-Theoretic Application

Let me close by showing how the results in this work can be fruitfully used to understand and investigate a question in decision theory. As I remarked in the introduction, many decision-theoretic models have been offered which represent preferences by Choquet integrals with respect to capacities. The most important and influential model of this sort is that in Schmeidler [31]. In that model Schmeidler adopted what is usually called an ‘‘Anscombe-Aumann’’ framework. That is, he assumed that the decision maker (DM for short) has access to an independent randomizing device (over which she has additive beliefs). So every action that she can undertake is really a map from the set X of states of the world to the set $\Delta(\mathcal{C})$ of simple lotteries on the set \mathcal{C} of possible outcomes for the

DM. That is, for every $x \in X$ action f gives $f(x, \cdot)$, a probability measure with finite support on \mathcal{C} . Assume for simplicity that \mathcal{C} is finite. The axioms Schmeidler presents imply that there is a utility function $u : \mathcal{C} \rightarrow \mathbf{R}$ and a capacity μ on X such that action f will be preferred to g if and only if

$$\int_X \left(\sum_{c \in \mathcal{C}} u(c) f(x, c) \right) d\mu(x) \geq \int_X \left(\sum_{c \in \mathcal{C}} u(c) g(x, c) \right) d\mu(x). \quad (40)$$

This can be equivalently restated as follows. There are two state spaces, X and Y , the randomization space. The DM has beliefs represented by a capacity μ on X and a probability measure ν on Y . Every action is a \mathcal{C} -valued measurable function defined on $Z = X \times Y$. Thus (40) can be rewritten

$$\int_X \int_Y u(f(x, y)) d\nu(y) d\mu(x) \geq \int_X \int_Y u(g(x, y)) d\nu(y) d\mu(x). \quad (41)$$

One can easily see that this representation has some interesting features. For instance Schmeidler shows that if μ is convex then the DM will in general prefer an action which randomizes between two actions f and g to either action (a property he dubs “uncertainty aversion”). For this reason it has been generally maintained that convex non-additive beliefs are intimately connected with uncertainty averse behavior in the sense just defined.

Eichberger and Kelsey [11] however made the following observation. Suppose that we represent the DM’s preferences in a different way. Let σ be an *independent* product of μ and ν . Then an obvious representation would have the DM preferring f over g if and only if

$$\int_Z u(f(z)) d\sigma(z) \geq \int_Z u(g(z)) d\sigma(z). \quad (42)$$

The results contained in this paper tell us that in general (42) will *not* be equivalent to (41), unless u , f and g are such that $u \circ f$ and $u \circ g$ are slice-comonotonic. And this will not necessarily be the case. Eichberger and Kelsey offer an example (example 3.1) to show that with this representation we do not necessarily obtain the preference for randomization that we discussed above. Unsurprisingly, to do so they use an action-payoff function which is not slice-comonotonic. They then go on to prove (using the result in lemma 6) that if σ is a belief function then the DM will *never* prefer a randomization between two Y -independent actions (i.e., actions which only depend on the “ambiguous” space X) f and g to either of them.¹⁸

Clearly the reason for this discrepancy of results is the failure of Fubini’s theorem: The Anscombe-Aumann framework forces preferences to be represented by a specific iterated integral, whereas the product space approach conceptually allows the integral to be taken with respect to any independent product capacity. The results obtained will thus be in general different because of the failure of Fubini’s theorem for non slice-comonotonic functions. This is not the case for expected utility (when also μ is additive), as there Fubini’s theorem holds for every product-measurable function.

¹⁸In the parlance of decision theorists: preferences will have to satisfy “betweenness”. David Kelsey told me that they have later proved a generalization of this latter result to the case in which σ is just a convex capacity.

The point about the difference between the iterated (two-stage) and product (one-stage) approaches was also made by Sarin and Wakker in [30, Section 5]. They offer another example with a payoff function which violates slice-comonotonicity. However it is not true, as they claim, that this two approaches are *always* irreconcilable unless μ is additive. In fact first of all it is clear from the results of this paper that the two approaches will give exactly the same result if the payoff function is slice-comonotonic. Moreover in their example they construct the product belief in a specific way (it is the, unique as we know, product belief function), but this clearly limits the generality of their conclusions. In fact one can construct another independent product capacity¹⁹ such that the integral of the given payoff function with respect to it will be identical to the iterated integral.

References

- [1] Alain Chateauneuf and Jean-Yves Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.
- [2] Gustave Choquet. Theory of capacities. *Annales de l'Institut Fourier (Grenoble)*, 5:131–295, 1953.
- [3] Arthur P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [4] Dieter Denneberg. *Non-Additive Measure and Integral*. Kluwer, Dordrecht, 1994.
- [5] James Dow and S.R. da Costa Werlang. Nash equilibrium under Knightian uncertainty: Breaking down backward induction. Mimeo, Fundação Getulio Vargas (forthcoming in *Journal of Economic Theory*), December 1991.
- [6] James Dow and S.R. da Costa Werlang. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, 60:197–204, 1992.
- [7] James Dow and S.R. da Costa Werlang. Learning under Knightian uncertainty: The law of large numbers for non-additive probabilities. Mimeo, London Business School, November 1994.
- [8] D. Dubois and H. Prade. Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems. *Acta Psychologica*, 68:53–78, 1988.
- [9] Rainer Dyckerhoff. Product capacities, a Fubini theorem and some applications. Mimeo, University of Hamburg, December 1994.

¹⁹Clearly it will not be a belief function, but it will also be convex if the α in the example is larger than $1/3$.

- [10] Jürgen Eichberger and David Kelsey. Non-additive beliefs and game theory. Mimeo, University of Melbourne, June 1994.
- [11] Jürgen Eichberger and David Kelsey. Uncertainty aversion and preference for randomization. Mimeo, University of Birmingham, February 1995.
- [12] Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.
- [13] Larry G. Epstein and Tan Wang. Intertemporal asset pricing under knightian uncertainty. *Econometrica*, 62:283–322, 1994.
- [14] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, MA, 1995.
- [15] Paolo Ghirardato. Agency theory with non-additive uncertainty. Mimeo, University of California at Berkeley, June 1994.
- [16] Paolo Ghirardato. Coping with ignorance: Unforeseen contingencies and non-additive uncertainty. Mimeo, University of California at Berkeley, November 1994.
- [17] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1988.
- [18] Itzhak Gilboa and David Schmeidler. Canonical representation of set functions. Mimeo, Northwestern University (forthcoming in *Mathematics of Operations Research*), April 1992.
- [19] Itzhak Gilboa and David Schmeidler. Additive representations of non-additive measures and the Choquet integral. *Annals of Operations Research*, 52:43–65, 1994.
- [20] Ebbe Hendon, Hans Jørgen Jacobsen, Birgitte Sloth, and Torben Tranæs. The product of capacities and lower probabilities. Mimeo, University of Copenhagen, October 1991.
- [21] Ebbe Hendon, Hans Jørgen Jacobsen, Birgitte Sloth, and Torben Tranæs. Expected utility with lower probabilities. *Journal of Risk and Uncertainty*, 8:197–216, 1994.
- [22] Peter J. Huber. The use of Choquet capacities in statistics. *Bulletin de l'Institut International de Statistique*, 45:181–191, 1973.
- [23] J.Y. Jaffray and Peter Wakker. Decision making with belief functions: Compatibility and incompatibility with the sure-thing principle. *Journal of Risk and Uncertainty*, 8:255–271, 1994.
- [24] David Kelsey and Frank Milne. Induced preferences and decision-making under risk and uncertainty. Mimeo, University of Birmingham, December 1993.
- [25] Peter Klibanoff. Uncertainty, decision, and normal-form games. mimeo, MIT, November 1993.

- [26] Massimo Marinacci. Finitely additive and ϵ Nash equilibria. Mimeo, Northwestern University, September 1993.
- [27] Massimo Marinacci. Decomposition and representation of coalitional games. Mimeo, Northwestern University, June 1994.
- [28] Massimo Marinacci. Limit laws for non-additive probabilities. Mimeo, Northwestern University, January 1995.
- [29] Sujoy Mukerji. Decision making under subjective uncertainty and ambiguity: Epistemic foundations and economic applications of the non-additive probability model. Mimeo, Yale University, November 1993.
- [30] Rakesh Sarin and Peter P. Wakker. A simple axiomatization of nonadditive expected utility. *Econometrica*, 60:1255–1272, 1992.
- [31] David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587, 1989.
- [32] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.
- [33] Glenn Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990.
- [34] Lloyd S. Shapley. A value for n -person games. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games*, pages 307–317. Princeton University Press, Princeton, 1953.
- [35] Lloyd S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.
- [36] Mario Henrique Simonsen and S.R. da Costa Werlang. Subadditive probabilities and portfolio inertia. *Revista de Econometria*, 11:1–19, 1991.
- [37] Peter Walley and Terrence L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10:741–761, 1982.
- [38] Larry A. Wasserman. Belief functions and statistical inference. *Canadian Journal of Statistics*, 18:183–196, 1990.