INTEGRATION IN POSSIBILITY THEORY

GERT DE COOMAN

Abstract. The paper discusses integration in possibility theory, both in an ordinal and in a numerical (behavioral) context. It is shown that in an ordinal context, the fuzzy integral has an important part in at least three areas: the extension of possibility measures to larger domains, the construction of product measures from marginals and the definition of conditional possibilities.

In a numerical (behavioral) context, integration can be used to extend upper probabilities to upper previsions. It is argued that the role of the fuzzy integral in this context is limited, as it can only be used to define a coherent upper prevision if the associated upper probability is $0-1$-valued, in which case it moreover coincides with the Choquet integral. These results are valid for arbitrary coherent upper probabilities, and therefore also relevant for possibility theory. It follows from the discussion that in a numerical context, the Choquet integral is better suited than the fuzzy integral for producing coherent upper previsions starting from possibility measures. At the same time, alternative expressions for the Choquet integral associated with a possibility measure are derived.

1. Introduction

The concept of integration in the theory of possibility measures, or supremum preserving set functions, is one of long-standing interest. Mainly two types of integral have been associated with them: the fuzzy integral [14, 22, 30, 32, 47, 51] and the Choquet integral [4, 26, 30, 32]. Important issues in the combination of an integral with a set function are its interpretation and its justification: what does a particular type of integration do, what is the meaning of its result, and why do we want to use a particular type of integration given this or that interpretation of the set function?

Possibility measures can be given a number of different interpretations, and it has been argued that these set functions have both a qualitative or ordinal aspect, and a quantitative or numerical one [33]. In this paper, we want to address the issues of interpretation and justification for integration of possibility measures in both the numerical and the ordinal context. In summary, we shall try and answer the following related questions: under what interpretation of possibility measures is which type of integration the most appropriate; why; and what exactly is it used for?

The paper is organized as follows. Preliminary material, such as basic definitions and notational conventions, has been gathered in Section 2. In Section 3 we investigate the role of the fuzzy integral in ordinal possibility theory; in particular in conditioning, and the problem of forming products from marginals. We also show that the form of the fuzzy integral can be justified from first principles, by imposing a number of basic properties which are deemed desirable for the integration process. The next sections deal with numerical uncertainty theory, on a behavioral interpretation. In Section 4 we give a brief survey of the basic ideas of this theory, and we discuss the roles Choquet integration and fuzzy integration have in extending upper probabilities to upper previsions. Section 5 is concerned with the special

Postdoctoral Fellow of the Fund for Scientific Research – Flanders (Belgium) (FWO).

case of possibility measures as numerical uncertainty models, and the relevance there of Choquet and fuzzy integration. Finally, we indicate in Section 6 that the Choquet integral is also naturally associated with numerical possibility measures in the context of random sets. Proofs of new results in the ordinal context have been gathered in the Appendix. Proofs of new results in the numerical context have been omitted, as they will be published elsewhere.

We also want to point out that, although this discussion is mainly concerned with possibility theory, we derive a number of results about fuzzy integrals and Choquet integrals of more general validity and appeal, most notably in Sections 4.4, 4.5 and 5.

2. Possibility measures, Choquet integrals and fuzzy integrals

In this section, we introduce the basic notions on which the later discussion will be built: possibility measures, Choquet integrals and fuzzy integrals. We also fix some basic notation.

2.1. **Possibility measures.** A *possibility measure* [14, 30, 32, 66] Π on a non-empty set Ω is a (set) function from the power class $\wp(\Omega)$ of Ω to the real unit interval $[0,1]$ that is *supremum preserving*: for any family $\{A_j : j \in J\}$ of subsets of Ω,

$$\Pi(\bigcup_{j\in J} A_j) = \sup_{j\in J} \Pi(A_j).$$

In particular ($J = \emptyset$), this implies that $\Pi(\emptyset) = 0$. Π is called *normal* if $\Pi(\Omega) = 1$.

The name 'possibility measure' was coined by Zadeh in 1978 [66], when he introduced these set functions in the context of fuzzy set theory as models representing degrees of ease, or physical possibility. They – or related notions – had nevertheless been studied before that time, in a number of different contexts, most notably by Shackle [53] as epistemic models for potential surprise; by Cohen [5] as 'Baconian probabilities' in the context of legal reasoning; by Shafer [54] as 'consonant plausibility functions' in his evidence theory; and by Shilkret [55] in mathematical measure theory. Levi [38, 40] seems to have been the first to relate what he calls 'Shackle measures' to probability theory and inductive reasoning. In later papers [39, 41], he compares his interpretation to the ones given by Cohen and Shafer. Giles [35] gives an interpretation of possibility measures in terms of upper betting rates. His approach is related to Levi's. Together with Walley's work on the behavioral theory of imprecise probabilities [60], it forms the starting point of the discussion in Sections 4 and 5. Important contributions to the theory of possibility measures, or *possibility theory*, were made by Dubois and Prade [30, 32]. They were the first to argue that there are two branches of possibility theory, a quantitative (numerical) theory of epistemic upper probability along the lines of Shackle, Levi and Giles, and a qualitative (ordinal) theory which they have, in recent years, devoted most attention to. A good overview of their work and that of many others in possibility theory can be found in [33]. In the ordinal or qualitative branch of possibility theory, the issue of integration has been studied in some detail by De Cooman in a series of three papers [14, 15, 16], and that work is the basis for the discussion in Section 3.

The values that a possibility measure Π assumes on the singletons of Ω determine a (point) function π from Ω to $[0,1]$ called the *distribution* of Π: $\pi(\omega) = \Pi(\{\omega\})$ for all $\omega \in \Omega$. A possibility measure is completely determined by its distribution, since for any $A \subseteq \Omega$, $\Pi(A) = \sup_{\omega\in A} \pi(\omega)$.

Any possibility measure is in particular *maxitive* in the sense that it satisfies $\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$ for all subsets A and B of Ω. On the other hand, a maxitive set function is not necessarily a possibility measure in the sense defined above, unless the set Ω is finite. An interesting discussion of the relation between maxitive set functions and possibility measures can be found in [50].

2.2. **The Choquet integral.** The first type of integral that we shall work with, was introduced by Choquet [4] in 1953. He associated it with a type of non-additive measures he called capacities. A good overview of the more recent advances in the field of Choquet integration and non-additive measure can be found in the book by Denneberg [26].

Let μ be a monotone set function[1] defined on a field $\mathcal{V}$ of subsets of Ω, that is, μ is a map taking elements of $\mathcal{V}$ to the set $\mathbb{R}$ of real numbers satisfying:

1. $\mu(\emptyset) = 0$;
2. if $A \subseteq B$ then $\mu(A) \leq \mu(B)$, for all A and B in $\mathcal{V}$.

Note that μ is positive, that is, $\mu(A) \geq 0$ for all $A \in \mathcal{V}$, and bounded, since all its values $\mu(A)$ are dominated by the real number $\mu(\Omega)$.

A real-valued function X on Ω will be called *$\mathcal{V}$-measurable* if the set $\{X \geq x\} = \{\omega \in \Omega \colon X(\omega) \geq x\}$ is in $\mathcal{V}$ for all $x \in \mathbb{R}$. It will be called *bounded* if its supremum $\sup[X] = \sup\{X(\omega) \colon \omega \in \Omega\}$ and its infimum $\inf[X] = \inf\{X(\omega) \colon \omega \in \Omega\}$ are real numbers (that is, are not infinite). The set of the bounded $\mathcal{V}$-measurable $\Omega - \mathbb{R}$-maps will be denoted by $\mathcal{B}(\Omega, \mathcal{V})$.

For any $X \in \mathcal{B}(\Omega, \mathcal{V})$ and $A \in \mathcal{V}$, the *Choquet integral*[2] [4, 26] of X on A with respect to μ is defined as:

$$(C)\int_A X \mathrm{d}\mu = \int_{-\infty}^{0} [\mu(A \cap \{X \geq x\}) - \mu(A)]\mathrm{d}x + \int_{0}^{+\infty} \mu(A \cap \{X \geq x\})\mathrm{d}x,$$

where the integrals on the right-hand side are improper Riemann integrals, which are guaranteed to exist because X is bounded, μ is bounded, and their integrands are non-increasing. Alternative expressions are:

$$\begin{aligned}(C)\int_A X \mathrm{d}\mu &= \inf[X]\mu(A) + \int_{\inf[X]}^{\sup[X]} \mu(A \cap \{X \geq x\})\mathrm{d}x \\ &= \sup[X]\mu(A) - \int_{\inf[X]}^{\sup[X]} [\mu(A) - \mu(A \cap \{X \geq x\})]\mathrm{d}x.\end{aligned}$$

If $A = \Omega$, we also write $(C)\int X \mathrm{d}\mu$ rather than $(C)\int_\Omega X \mathrm{d}\mu$. At the same time, it holds that $(C)\int_A X \mathrm{d}\mu = (C)\int X \mathrm{d}\mu_A$, where μ_A is the set function on $\mathcal{V}$ defined by $\mu_A(B) = \mu(A \cap B)$. For non-negative X we get:

$$(C)\int_A X \mathrm{d}\mu = \int_0^{+\infty} \mu(A \cap \{X \geq x\})\mathrm{d}x.$$

The Choquet integral can also be written as an (improper) Riemann-Stieltjes integral: $(C)\int X \mathrm{d}\mu = \int_{-\infty}^{+\infty} x \mathrm{d}F_X(x)$, where $F_X(x) = \mu(\{X < x\})$. It follows that if μ is a measure (a countably additive positive set function), the Choquet integral coincides with the Lebesgue integral. For (finitely) additive set functions, also called charges, the Choquet integral also coincides with the so-called Dunford integral [1, 34].

[1] Following Denneberg [26], we take the term *set function* for μ to imply that μ vanishes at the empty set: $\mu(\emptyset) = 0$.

[2] This is the so-called *asymmetric* Choquet integral, due to Choquet [4]. There is also a *symmetric* Choquet integral, which seems to have been first introduced by Šipoš in 1979. More details can be found in [26, 51].

2.3. **The fuzzy integral.** The second type of integral we shall study is the fuzzy integral, introduced by Sugeno [57] in 1974. The work of Sugeno was taken up by Ralescu and Adams [52] in 1980, and since then the integral has been significantly generalized, and used in many practical applications. More information about the history of this integral can be found elsewhere in this volume.

For this integral, we give a definition that is somewhat more general than the one originally given by Sugeno [57]. We only consider integration of *non-negative* bounded $\mathcal{V}$-measurable maps. The set of these maps will be denoted by $\mathcal{B}^+(\Omega, \mathcal{V})$.

Consider a binary operator $\circledast$ on the set $\mathbb{R}^+$ of non-negative reals, taking values in $\mathbb{R}^+$, and satisfying:

1. $0 \circledast x = x \circledast 0 = 0$ [zero 0]
2. if $x \leq u$ and $y \leq v$ then $x \circledast y \leq u \circledast v$ [isotonicity].

Special instances of such an operator are for instance the algebraic product $\times$ and the minimum operator min.

For $X \in \mathcal{B}^+(\Omega, \mathcal{V})$ and $A \in \mathcal{V}$, the *fuzzy $\circledast$-integral* of X on A with respect to μ is defined as[3]

$$(F_\circledast) \int_A X \mathrm{d}\mu = \sup_{x \geq 0} \left[x \circledast \mu(A \cap \{X \geq x\}) \right].$$

It is a straightforward modification of the Choquet integral, where summation is replaced by supremum, and multiplication by $\circledast$. If we take $\circledast = \min$, we recover the usual definition of the fuzzy integral. The case $\circledast = \times$ (algebraic product) yields the *Shilkret integral*, whose combination with maxitive set functions was first studied by Shilkret [55]. Interesting special cases of the fuzzy $\circledast$-integral, for operators $\circledast$ with more specific properties, have been studied by Weber [65], who only considered $[0,1]$-valued X and so-called strict triangular norms for $\circledast$, and by Mesiar [47].

3. The fuzzy integral in ordinal possibility theory

In this section, we consider possibility measures in an ordinal context. Roughly speaking, this means that we are only interested in questions of the following type: is an event A more 'possible' than an event B, or in terms of the possibility measure Π, is $\Pi(B) \leq \Pi(A)$? We are *not* interested in how much more possible A is than B, which could be expressed by the difference $\Pi(A) - \Pi(B)$, or the quotient $\Pi(A)/\Pi(B)$. In fact, we do not even assume that asking such a question has any meaning.

To stress that only the ordering of the values of Π is important, we shall for the purposes of this section define a (K-valued) possibility measure Π as a map taking subsets of Ω to elements of a *complete chain*[4] $(K, \leq)$, such that, for any family $\{A_j : j \in J\}$ of subsets of Ω,

$$\Pi(\bigcup_{j \in J} A_j) = \sup_{j \in J} \Pi(A_j),$$

where the supremum on the right-hand side is the supremum in the complete chain $(K, \leq)$. Note that $\Pi(\emptyset) = 0_K$, where 0_K denotes the smallest element of $(K, \leq)$. We call Π *normal* if $\Pi(\Omega) = 1_K$, where 1_K is the greatest element of $(K, \leq)$. Special cases of interest obtain when $K = [0,1]$ (the real unit interval), or when K has only a finite number of elements. We assume that $(K, \leq)$ is non-trivial, in that $0_K < 1_K$.

[3] This integral is non-negative and finite.

[4] Actually, the ordering $\leq$ on K does not even need to be linear, and we could just as well let an ordinal possibility measure take values in a complete *lattice* rather than a complete chain. For more detailed discussions, see [9, 10, 14, 15, 16, 22, 23].

Distributions have meaning in this more general context as well: the *distribution* π of Π is the map from Ω to K defined by $\pi(\omega) = \Pi(\{\omega\})$ for all $\omega \in \Omega$. It completely characterizes Π, as for any $A \subseteq \Omega$: $\Pi(A) = \sup_{\omega \in A} \pi(\omega)$.

Under the ordinal interpretation, possibility measures are convenient representations of comparative possibility relations. A binary relation $\preceq$ on $\wp(\Omega)$ is called a *comparative* (or quantitative) *possibility relation* if,

(CP1) if $A \preceq B$ and $B \preceq C$ then $A \preceq C$ [transitivity]
(CP2) if $A \subseteq B$ then $A \preceq B$ [monotonicity]
(CP3) $\Omega \not\preceq \emptyset$ [non-triviality]
(CP4) $A \preceq B$ or $B \preceq A$ [completeness]
(CP5) if for all $j \in J$, $A_j \preceq B$, then $\bigcup_{j \in J} A_j \preceq B$ [union-closedness].

where A, B, C and A_j, $j \in J$ are arbitrary subsets of Ω, and J is an arbitrary non-empty index set. Such relations (without the completeness axiom) were studied in detail in [9, 11, 12, 13]. Lewis [42, 43] and, independently, Dubois [27] have given and studied different but equivalent definitions in a finitary context (where only finite index sets J are considered). It is easily verified (see [9, 12, 13] for more information) that a binary relation $\preceq$ on $\wp(\Omega)$ is a comparative possibility relation if and only if there exists a complete chain $(K, \leq)$ (with $0_K < 1_K$) and a normal K-valued possibility measure on Ω such that for all A and B in $\wp(\Omega)$: $A \preceq B \Leftrightarrow \Pi(A) \leq \Pi(B)$.

3.1. **An ordinal version of the fuzzy integral.** In this ordinal context, events $A \subseteq \Omega$ can be considered as special K-valued maps on Ω, assuming only the values 0_K and 1_K. More specifically, A can be identified with its K-valued indicator function $I_A \colon \Omega \to K$, defined by

$$I_A(\omega) = \begin{cases} 1_K & \text{if } \omega \in A \\ 0_K & \text{if } \omega \notin A. \end{cases}$$

The K-valued possibility measure Π can then be reinterpreted as a function defined on the subset $\{I_A \colon A \subseteq \Omega\}$ of the set K^Ω of all maps from Ω to K. Integration will be considered as a way to extend this function to a map defined on the entire domain K^Ω. Since we are working with a general complete chain K rather than a set of real numbers, Choquet integration has no part in ordinal possibility theory. But we now suggest a very general 'truth-functional' way of performing the extension of a possibility measure from events to elements of K^Ω, which will lead back to *fuzzy integration* if we impose a number of conditions that we feel such an extension should satisfy.

As a first step, consider the extension of Π to *simple* $\Omega - K$-maps, that is, maps assuming only a finite number of values. Let s be such a map, and denote its different values generically by $s_1, \dots, s_n$ (where $n \geq 1$) and the corresponding domains by $D_k = s^{-1}(\{s_k\})$, $k = 1, \dots, n$.

We want to write s as some combination in terms of the values s_k and the indicator functions I_{D_k} of the domains D_k. Consider two binary operators $\boxdot$ and $\boxplus$ on K, and assume that $\boxplus$ is commutative and associative, so that the following expression makes sense:

$$\boxplus_{k=1}^{n} [s_k \boxdot I_{D_k}(\omega)]. \tag{1}$$

This expression will be identical to $s(\omega)$ for all $\omega \in \Omega$ and for any choice of simple map s if and only if $\boxdot$ and $\boxplus$ satisfy the following properties:

$$\lambda \boxdot 1_K = \lambda \text{ and } \lambda \boxplus (\mu \boxdot 0_K) = \lambda \tag{2}$$

for all λ and μ in K. We assume from now on that (2) indeed holds true.

For a simple map s and a subset A of Ω, we draw inspiration from (1) to define

$$R(s;A) = \boxplus_{k=1}^{n}[s_k \boxdot \Pi(A \cap D_k)].$$

In this approach, which reminds us of the path taken in Lebesgue integration [2] and is similar to the approaches to fuzzy integration followed in [46, 51, 52, 58, 65], we define the value $R(s;A)$ of the functional $R(\cdot\,;\cdot)$ as a 'truth-functional' combination of the values s_k assumed by s and the measures $\Pi(A \cap D_k)$ of the corresponding domains.

The following proposition states that if we want the functional $R(\cdot\,;\cdot)$ to satisfy a number of reasonable basic properties, the choice for $\boxdot$ and $\boxplus$ is rather limited. Before we can formulate it, however, we need to introduce some classes of operations on complete chains $(K, \leq)$, which will be used rather frequently in the text. More details about such operations can be found in [21]. A *triangular seminorm* $\odot$ on K is a binary operation on K that is isotonic (non-decreasing) in both places, and has unit 1_K: $1_K \odot \lambda = \lambda \odot 1_K = \lambda$ for all λ in K. Any triangular seminorm $\odot$ has zero 0_K: $\lambda \odot 0_K = 0_K \odot \lambda = 0_K$ for all λ in K. If $\odot$ is also commutative and associative, it is called a triangular *norm*. Triangular (semi)*co*norms are defined similarly, but have unit 0_K rather than 1_K.

Proposition 1. *Assume that the operators $\boxplus$ and $\boxdot$ satisfy* (2) *and that Ω has at least two elements. The functional $R(\cdot\,;\cdot)$ satisfies the following properties for every choice of the K-valued possibility measure Π:*

1. *Let s be a simple map, and let $\{F_\ell \colon \ell = 1, \ldots, m\}$ be any partition of Ω such that s is constant on its elements. Let t_ℓ be the constant value of s on F_ℓ, for $\ell = 1, \ldots, m$. Then $R(s;A) = \boxplus_{\ell=1}^{m}[t_\ell \boxdot \Pi(A \cap F_\ell)]$;*
2. *$R(I_A;\Omega) = \Pi(A)$ for all $A \subseteq \Omega$;*
3. *$R(\cdot\,;\cdot)$ is increasing in both arguments: if the simple map s is point-wise dominated by the simple map t and $A \subseteq B \subseteq \Omega$ then $R(s;A) \leq R(t;B)$;*

if and only if $\boxplus = \max$ and $\boxdot$ is a triangular seminorm on K.

A proof of this proposition is given in the Appendix. The second condition merely states that we want to use the functional $R(\cdot\,;\Omega)$ as a way to extend Π from events to simple maps. If we call a *representing partition* for a simple map s any partition of Ω such that s is constant on its elements, then the first condition is essentially an invariance property of the functional $R(\cdot\,;\cdot)$ under refining of the representing partition: $R(\cdot\,;\cdot)$ can be defined using any of these representing partitions. A completely analogous property holds for the Lebesgue integral of simple maps associated with additive measures.

It will be assumed from now on that $\boxplus = \max$ and that $\boxdot$ is a triangular seminorm. It follows from the assumptions that $R(1_K;A) = \Pi(A)$ for all $A \subseteq \Omega$.

As a second step, we want to extend the functional $R(\cdot\,;\cdot)$ from simple maps s to arbitrary $\Omega - K$-maps h. This will be done by approximating h from below by simple maps. If we use the notation $\leq$ for 'is point-wise dominated by', then for all $\omega \in \Omega$:

$$h(\omega) = \sup\{s(\omega) \colon s \text{ is simple and } s \leq h\}.$$

So we define the integral of h on A with respect to Π as the element of K given by

$$(F_\boxdot)\int_A h \mathrm{d}\Pi = \sup\{R(s;A) \colon s \text{ is simple and } s \leq h\}.$$

It follows from Proposition 1 (second and third conditions) that for any simple map s,

$$(F_\boxdot)\int_A s \mathrm{d}\Pi = R(s;A)$$

and in particular for an event $A \subseteq \Omega$, $(F_{\boxdot}) \int_{\Omega} I_A \mathrm{d}\Pi = R(I_A; \Omega) = \Pi(A)$, so this integral indeed provides us with a way to extend Π from events to (simple and other) K-valued maps on Ω. It turns out that the integral which we have defined here is very closely related to the fuzzy integral. It is shown in the Appendix that

$$(F_{\boxdot}) \int_A h \mathrm{d}\Pi = \sup_{\alpha \in K} [\alpha \boxdot \Pi(A \cap \{h \geq \alpha\})]. \tag{3}$$

A more indirect proof was also given in [22]. The argument here shows that in an ordinal possibilistic context, it is possible to 'construct' this integral from first principles by imposing a number of basic (and natural) requirements. For $K = [0,1]$ and arbitrary seminorms $\boxdot$ this type of integral associated with arbitrary monotone set functions was first studied by Suárez García and Gil Álvarez [56] as a generalization of Sugeno's fuzzy integral. De Cooman *et al.* [22, 23] extended their work to monotone set functions taking values in a complete lattice, and showed that these integrals have very nice properties when associated with possibility measures. The present construction of this integral from first principles in the context of possibility theory is new, however. A different characterization of the fuzzy integral associated with $[0,1]$-valued possibility measures was given by Mesiar [47].

The argument above also shows that it is possible to justify the form (3) of the fuzzy integral in purely order-theoretic terms. This need not mean that the form of the integral is an order-invariant, however. For this to be the case, the integral must retain its form under all order-automorphisms. Let ϕ be an order-automorphism of $(K, \leq)$. Then it is verified that

$$\phi\left((F_{\boxdot}) \int_A h \mathrm{d}\Pi\right) = (F_{\boxdot_\phi}) \int_A (\phi \circ h) \mathrm{d}(\phi \circ \Pi),$$

where $\lambda \boxdot_\phi \mu = \phi[\phi^{-1}(\lambda) \boxdot \phi^{-1}(\mu)]$ for all λ and μ in K. So, for order-invariance of the form of the integral we need that $\boxdot_\phi = \boxdot$ for all possible order-automorphisms ϕ. This is the case for $\boxdot = \min$, which brings us back to Sugeno's original definition of the fuzzy integral [57].

In the rest of Section 3, we impose an additional condition on the triangular seminorm $\boxdot$: it must be *completely* distributive (on both sides) with respect to supremum, so for all λ and any family $\{\mu_j : j \in J\}$ in K,

$$\lambda \boxdot \left(\sup_{j \in J} \mu_j\right) = \sup_{j \in J} [\lambda \boxdot \mu_j] \text{ and } \left(\sup_{j \in J} \mu_j\right) \boxdot \lambda = \sup_{j \in J} [\mu_j \boxdot \lambda]. \tag{4}$$

An other way of saying this is that the partial maps of $\boxdot$ should be supremum preserving. When $(K, \leq)$ is the real unit interval $([0,1], \leq)$, this means that the partial maps of $\boxdot$ should be left-continuous. If K is finite, (4) holds for any triangular seminorm $\boxdot$.

The fuzzy integral can in this case also be written as

$$(F_{\boxdot}) \int_A h \mathrm{d}\Pi = \sup_{\omega \in A} [h(\omega) \boxdot \pi(\omega)]. \tag{5}$$

This seems to have been first observed by Dubois and Prade [29] in the case that $K = [0,1]$ and $\boxdot = \min$. We shall also write this integral as $(F_{\boxdot}) \int_A h(\omega) \mathrm{d}\Pi(\omega)$ to indicate which dummy variable is used in the supremum. A proof of (5) can be found in [22], together with a number of interesting properties of the fuzzy integral associated with a possibility measure. The most interesting among these is the following exchangeability formula. For any family $\{h_j : j \in J\}$ of $\Omega - K$-maps, let $\sup_{j \in J} h_j$ denote their point-wise supremum. Then

$$(F_{\boxdot}) \int_A \sup_{j \in J} h_j \mathrm{d}\Pi = \sup_{j \in J} (F_{\boxdot}) \int_A h_j \mathrm{d}\Pi.$$

This shows that the fuzzy integral can be used to extend a possibility measure, which is essentially a supremum preserving map (set function) on events, to a supremum preserving map (functional) on $\Omega - K$-maps. We want to mention in passing that there are integrals which are related to the fuzzy integral, but which can be used to turn an infimum preserving set function (also called a *necessity measure*) into an infimum preserving functional [22], a possibility measure into an infimum preserving functional, and a necessity measure into a supremum preserving functional [20]. It is also worth mentioning that quite some work [3, 7, 14, 15, 37, 48, 58] has been done on the converse, Radon-Nikodym-like problem: given some supremum preserving map σ on a subset of K^Ω – which can also be called a collection of fuzzy events –, find a K-valued possibility measure Π and a binary operation $\boxdot$ on K that generate that map σ through fuzzy integration in the sense that $\sigma(h) = (F_\boxdot) \int_\Omega h \mathrm{d}\Pi$ for all h in the domain of σ.

In a series of three papers on possibility measures [14, 15, 16], we have argued that the fuzzy integral has in important part in an ordinal theory of possibility. The following subsections are intended as a very brief survey of the more striking points of this argument.

3.2. **Product measures and multiple integrals.** The fuzzy integral can be used to construct product possibility measures from marginals. Consider two *normal* K-valued possibility measures Π_1 and Π_2 with distributions π_1 and π_2 on the respective sets Ω_1 and Ω_2. A *product* of Π_1 and Π_2 will be any normal K-valued possibility measure Π on $\Omega_1 \times \Omega_2$ with *marginals* Π_1 and Π_2: for all $A_1 \subseteq \Omega_1$ and $A_2 \subseteq \Omega_2$:

$$\Pi(A_1 \times \Omega_2) = \Pi_1(A_1) \text{ and } \Pi(\Omega_1 \times A_2) = \Pi_2(A_2).$$

The point-wise greatest, and therefore least restrictive, such product is the normal possibility measure $\Pi_1 \times_{\min} \Pi_2$ whose distribution $\pi_1 \times_{\min} \pi_2 \colon \Omega_1 \times \Omega_2 \to K$ is given by

$$(\pi_1 \times_{\min} \pi_2)(\omega_1, \omega_2) = \min\{\pi_1(\omega_1), \pi_2(\omega_2)\}.$$

Note that on rectangles $A_1 \times A_2$:

$$(\Pi_1 \times_{\min} \Pi_2)(A_1 \times A_2) = \min\{\Pi_1(A_1), \Pi_2(A_2)\}.$$

There is another way leading to this product, which closely follows the standard measure-theoretic construction of products from marginals using integration (see for instance [2]). Consider an arbitrary $E \subseteq \Omega_1 \times \Omega_2$. In order to find the measure of E, we can proceed as follows: (i) cut E along the line $\{\omega_1\} \times \Omega_2$, where $\omega_1 \in \Omega_1$, to obtain the subset $\omega_1 E = \{\nu_2 \in \Omega_2 \colon (\omega_1, \nu_2) \in E\}$ of Ω_2; (ii) take its measure $\Pi_2(\omega_1 E)$; (iii) this defines a function on Ω_1 which can be integrated using a fuzzy integral associated with the possibility measure Π_1. If we do this for every $E \subseteq \Omega_1 \times \Omega_2$, we actually construct a K-valued set function ρ_{21} on $\wp(\Omega_1 \times \Omega_2)$, with

$$\rho_{21}(E) = (F_\boxdot) \int_{\Omega_1} \Pi_2(\omega_1 E) \mathrm{d}\Pi_1(\omega_1) = \sup_{(\omega_1, \omega_2) \in E} [\pi_2(\omega_2) \boxdot \pi_1(\omega_1)].$$

In a completely symmetrical way, we may construct the set function ρ_{12}, where

$$\rho_{12}(E) = (F_\boxdot) \int_{\Omega_2} \Pi_1(E \omega_2) \mathrm{d}\Pi_2(\omega_2) = \sup_{(\omega_1, \omega_2) \in E} [\pi_1(\omega_1) \boxdot \pi_2(\omega_2)].$$

Both ρ_{12} and ρ_{21} are possibility measures on $\Omega_1 \times \Omega_2$ with marginals Π_1 and Π_2, and therefore products of Π_1 and Π_2. ρ_{12} and ρ_{21} are completely determined by the values they assume on rectangles $A_1 \times A_2$, and in particular we have

$$\rho_{21}(A_1 \times A_2) = \Pi_2(A_2) \boxdot \Pi_1(A_1) \text{ and } \rho_{12}(A_1 \times A_2) = \Pi_1(A_1) \boxdot \Pi_2(A_2).$$

They are identical if $\boxdot$ is commutative. In that case the common value is denoted by $\Pi_1 \times_{\boxdot} \Pi_2$ and its distribution by $\pi_1 \times_{\boxdot} \pi_2$. The choice $\boxdot = \min$ yields the product $\Pi_1 \times_{\min} \Pi_2$ mentioned above. For this choice, the product is also an order-invariant.

If $\boxdot$ is both commutative and associative (and therefore in particular a triangular norm on K), it is very easy to prove the following Fubini-like result. Given an $\Omega_1 \times \Omega_2 - K$-map h, we can consider its fuzzy integral associated with $\Pi_1 \times_{\boxdot} \Pi_2$ on the set $E \subseteq \Omega_1 \times \Omega_2$, called *double integral*:

$$
\begin{aligned}
(F_{\boxdot}) \int_E h \mathrm{d}(\Pi_1 \times_{\boxdot} \Pi_2) &= \sup_{(\omega_1,\omega_2)\in E} [h(\omega_1,\omega_2) \boxdot [(\pi_1 \times_{\boxdot} \pi_2)(\omega_1,\omega_2)]] \\
&= \sup_{(\omega_1,\omega_2)\in E} [h(\omega_1,\omega_2) \boxdot \pi_1(\omega_1) \boxdot \pi_2(\omega_2)].
\end{aligned}
$$

On the other hand, we may also integrate the partial maps of h in one space and then integrate the result in the other space. This yields the following two so-called *iterated integrals*:

$$
\begin{aligned}
(F_{\boxdot}) \iint_E h \mathrm{d}\Pi_1 \mathrm{d}\Pi_2 &= (F_{\boxdot}) \int_{\Omega_2} \left((F_{\boxdot}) \int_{E\omega_2} h(\omega_1,\omega_2) \mathrm{d}\Pi_1(\omega_1) \right) \mathrm{d}\Pi_2(\omega_2) \\
&= \sup_{\omega_2\in\Omega_2} \left[\sup_{\omega_1\in E\omega_2} [h(\omega_1,\omega_2) \boxdot \pi_1(\omega_1)] \boxdot \pi_2(\omega_2) \right] \\
&= \sup_{(\omega_1,\omega_2)\in E} [h(\omega_1,\omega_2) \boxdot \pi_1(\omega_1) \boxdot \pi_2(\omega_2)]
\end{aligned}
$$

and similarly

$$
\begin{aligned}
(F_{\boxdot}) \iint_E h \mathrm{d}\Pi_2 \mathrm{d}\Pi_1 &= (F_{\boxdot}) \int_{\Omega_1} \left((F_{\boxdot}) \int_{\omega_1 E} h(\omega_1,\omega_2) \mathrm{d}\Pi_2(\omega_2) \right) \mathrm{d}\Pi_1(\omega_1) \\
&= \sup_{\omega_1\in\Omega_1} \left[\sup_{\omega_2\in \omega_1 E} [h(\omega_1,\omega_2) \boxdot \pi_2(\omega_2)] \boxdot \pi_1(\omega_1) \right] \\
&= \sup_{(\omega_1,\omega_2)\in E} [h(\omega_1,\omega_2) \boxdot \pi_2(\omega_2) \boxdot \pi_1(\omega_1)].
\end{aligned}
$$

Then the order of integration does not matter, and both iterated integrals are equal to the double integral:

$$
(F_{\boxdot}) \int_E h \mathrm{d}(\Pi_1 \times_{\boxdot} \Pi_2) = (F_{\boxdot}) \iint_E h \mathrm{d}\Pi_1 \mathrm{d}\Pi_2 = (F_{\boxdot}) \iint_E h \mathrm{d}\Pi_2 \mathrm{d}\Pi_1.
$$

3.3. **Integral equations for conditional possibility.** Fuzzy integrals can also be used to introduce a notion of conditional possibility. Consider two variables X and Y taking values in the sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Then (X,Y) is a variable in $\mathcal{X} \times \mathcal{Y}$, and we assume that ordinal information about the values that (X,Y) may take in $\mathcal{X} \times \mathcal{Y}$ is given by the *normal* K-valued possibility measure $\Pi_{(X,Y)}$ on $\mathcal{X} \times \mathcal{Y}$ with distribution $\pi_{(X,Y)}$. The marginals Π_X on $\mathcal{X}$ and Π_Y on $\mathcal{Y}$ are normal K-valued possibility measures with respective distributions π_X and π_Y, and are defined by

$$
\Pi_X(A) = \Pi_{(X,Y)}(A \times \mathcal{Y}) \text{ and } \Pi_Y(B) = \Pi_{(X,Y)}(\mathcal{X} \times B)
$$

for $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$. The marginal Π_X provides information about the values that X assumes in $\mathcal{X}$, and similarly for Π_Y. Fix $A \subseteq \mathcal{X}$ and consider the following system of integral equations in the $\mathcal{Y} - K$-map h:

$$
(F_{\boxdot}) \int_B h \mathrm{d}\Pi_Y = \Pi_{(X,Y)}(A \times B), \quad B \subseteq \mathcal{Y}. \tag{6}
$$

These equations are the formal counterparts of the ones used in Kolmogorov's measure-theoretic account of probability theory to define the notion of conditional probability (see for instance [2, 44]). We say that two $\mathcal{Y}-K$-maps f and g are $(\Pi_Y, \boxdot)$-*equivalent* if

$$f(y) \boxdot \pi_Y(y) = g(y) \boxdot \pi_Y(y), \quad y \in \mathcal{Y},$$

and denote this as $f = g$ *a.s.* $(\Pi_Y, \boxdot)$. Obviously, $(\Pi_Y, \boxdot)$-equivalence is an equivalence relation on the set $K^{\mathcal{Y}}$ of all $\mathcal{Y}-K$-maps. If we rewrite (6) in the equivalent form:

$$h(y) \boxdot \pi_Y(y) = \Pi_{(X,Y)}(A \times \{y\}), \quad y \in \mathcal{Y}, \tag{7}$$

we see that if the system of equations (6) has solutions h, they will only be determined up to $(\Pi_Y, \boxdot)$-equivalence. Let us define the *left-residual implicator*[5] [6, 9, 21] $\mathcal{I}_{\boxdot}\colon K^2 \to K$ associated with the triangular seminorm $\boxdot$ as follows:

$$\mathcal{I}_{\boxdot}(\alpha, \beta) = \sup\{\gamma \in K\colon \gamma \boxdot \alpha \leq \beta\}, \quad (\alpha, \beta) \in K^2.$$

Then it is easy to see that (6) has a solution if and only if the map $h^g\colon \mathcal{Y} \to K$ defined by

$$\begin{aligned} h^g(y) &= \mathcal{I}_{\boxdot}(\pi_Y(y), \Pi_{(X,Y)}(A \times \{y\})) \\ &= \mathcal{I}_{\boxdot}(\Pi_{(X,Y)}(\mathcal{X} \times \{y\}), \Pi_{(X,Y)}(A \times \{y\})), \quad y \in \mathcal{Y}, \end{aligned}$$

is the (point-wise) greatest solution. We denote any solution (any map in the $(\Pi_Y, \boxdot)$-equivalence class containing h^g) by $\Pi_{X|Y}(A|\cdot)$, and call $\Pi_{X|Y}(A|y)$ the *conditional possibility that X assumes a value in $A \subseteq \mathcal{X}$ given that Y assumes the value y*. We also denote $\Pi_{X|Y}(\{x\}|y)$ by $\pi_{X|Y}(x|y)$.

It seems difficult in this general context to characterize those triangular seminorms $\boxdot$ for which the integral equations (6) are guaranteed to have solutions, irrespective of the choice of the (normal) possibility measure $\Pi_{(X,Y)}$. Giving sufficient conditions in a number of special cases is easy enough, though. To give one important general example (more specific ones are given below, see Examples 1 and 2): when the complete chain $(K, \leq)$ is order-dense and $\boxdot$ is a triangular norm, it is sufficient that the partial maps of $\boxdot$ are both supremum preserving and infimum preserving [6]. If $K = [0, 1]$, this amounts to requiring that the partial maps of $\boxdot$ should be continuous.

Let us now assume that the triangular seminorm $\boxdot$ is such that solutions of the integral equations are guaranteed. It turns out that if we fix $y \in \mathcal{Y}$, any solution $\Pi_{X|Y}(\cdot|y)$ will still be a possibility measure on $\mathcal{X}$ with distribution $\pi_{X|Y}(\cdot|y)$, but only in some weaker sense, as it were up to $(\Pi_Y, \boxdot)$-equivalence. A proof of this result can be found in [15].

Theorem 2. *1.* $\Pi_{X|Y}(\emptyset|\cdot) = 0_K$ *a.s.* $(\Pi_Y, \boxdot)$.
2. $\Pi_{X|Y}(\mathcal{X}|\cdot) = 1_K$ *a.s.* $(\Pi_Y, \boxdot)$.
3. Let $\{A_j\colon j \in J\}$ *be an arbitrary family of subsets of* $\mathcal{X}$. *Then*

$$\Pi_{X|Y}(\bigcup_{j \in J} A_j|\cdot) = \sup_{j \in J} \Pi_{X|Y}(A_j|\cdot) \text{ a.s. } (\Pi_Y, \boxdot).$$

4. Let A *be a subset of* $\mathcal{X}$. *Then*

$$\Pi_{X|Y}(A|\cdot) = \sup_{x \in A} \pi_{X|Y}(x|\cdot) \text{ a.s. } (\Pi_Y, \boxdot).$$

The following examples illustrate cases of particular interest.

[5] The difference between left-residual and right-residual implicators derives from the fact that a triangular seminorm $\boxdot$ need not be commutative. For triangular norms, both notions coincide and we simply speak about residual implicators.

Example 1. Let $\boxdot$ be the minimum operator min (or in other words the meet) on the complete chain $(K, \leq)$. Its residual implicator is given by

$$\mathcal{I}_{\min}(\alpha, \beta) = \begin{cases} 1 & \text{if } \alpha \leq \beta \\ \beta & \text{if } \alpha > \beta \end{cases} \quad (\alpha, \beta) \in K^2.$$

It is seen that the $\mathcal{Y} - K$-map h^g, defined as

$$\begin{aligned} h^g(y) &= \mathcal{I}_{\min}(\pi_Y(y), \Pi_{(X,Y)}(A \times \{y\})) \\ &= \begin{cases} \Pi_{(X,Y)}(A \times \{y\}) & \text{if } \Pi_{(X,Y)}(A \times \{y\}) < \pi_Y(y) \\ 1 & \text{if } \Pi_{(X,Y)}(A \times \{y\}) = \pi_Y(y) \end{cases} \end{aligned}$$

is indeed a solution of (7) – or (6) –, and at the same time the point-wise greatest (least restrictive) solution. The defining equation for conditional possibilities is therefore guaranteed to have solutions for this particular choice of $\boxdot$. For the conditional possibility distribution $\pi_{X|Y}(x|y)$ we find in particular that

$$\pi_{X|Y}(x|y) = \begin{cases} \pi_{(X,Y)}(x, y) & \text{if } \pi_{(X,Y)}(x, y) < \pi_Y(y) \\ 1 & \text{if } \pi_{(X,Y)}(x, y) = \pi_Y(y) \end{cases} \quad \text{a.s. } (\Pi_Y, \min),$$

and what appears in the right-hand side is precisely the Dubois-Prade conditioning rule [28, 30, 32]. Note that this rule is an order-invariant.

Example 2. Let $(K, \leq)$ be the real unit interval and let $\boxdot = \times$ be the algebraic product on $[0, 1]$. Its residual implicator $\mathcal{I}_\times$ is given by:

$$\mathcal{I}_\times(\alpha, \beta) = \begin{cases} 1 & \text{if } \alpha \leq \beta \\ \beta/\alpha & \text{if } \alpha > \beta \end{cases} \quad (\alpha, \beta) \in [0, 1]^2.$$

The $\mathcal{Y} - [0, 1]$-map h^g, defined as

$$\begin{aligned} h^g(y) &= \mathcal{I}_\times(\pi_Y(y), \Pi_{(X,Y)}(A \times \{y\})) \\ &= \begin{cases} \Pi_{(X,Y)}(A \times \{y\})/\pi_Y(y) & \text{if } \pi_Y(y) > 0 \\ 1 & \text{if } \pi_Y(y) = 0 \end{cases} \end{aligned}$$

is the greatest solution of the defining equation for conditional possibilities (7) – or (6). For the conditional possibility distribution $\pi_{X|Y}(x|y)$ we find

$$\pi_{X|Y}(x|y) = \begin{cases} \pi_{(X,Y)}(x, y)/\pi_Y(y) & \text{if } \pi_Y(y) > 0 \\ 1 & \text{if } \pi_Y(y) = 0 \end{cases} \quad \text{a.s. } (\Pi_Y, \times),$$

and what appears in the right-hand side is essentially Dempster's conditioning rule [25, 54]. Interestingly, we have that two $\mathcal{Y} - [0, 1]$-maps f and g are $(\Pi_Y, \times)$-equivalent if and only if they only differ on a set with possibility zero:

$$f = g \text{ a.s. } (\Pi_Y, \times) \Leftrightarrow \Pi_Y(\{y \in \mathcal{Y} \colon f(y) \neq g(y)\}) = 0.$$

As a corollary to Theorem 2, we find that $\Pi_{X|Y}(\cdot|y)$ is uniquely determined, and a possibility measure with distribution $\pi_{X|Y}(\cdot|y)$, for all y such that $\pi_Y(y) > 0$. This can of course also be directly verified using the expressions given above. We want to point out that for general triangular seminorms $\boxdot$ on $[0, 1]$, $(\Pi_Y, \boxdot)$-equivalence does not coincide with (but is weaker than) the notion of 'equality except on a set with possibility zero'. De Baets *et al.* [8] have shown that the only continuous triangular *norms* for which the solutions (7) are uniquely determined and possibility measures except on a set with possibility zero, are the *strict* triangular norms, which are essentially identical to the algebraic product, up to an order-automorphism. Nevertheless, we want to stress that by replacing 'equality except on a set with possibility zero' with the weaker notion of equivalence introduced above, we find similar but weaker results for more general class of triangular *semi*norms.

The discussion above serves to illustrate the role of the fuzzy integral in the ordinal version of possibility theory: it plays a role similar to the Lebesgue integral in probability theory, and it allows the construction of a mathematical theory of possibility measures and possibilistic conditioning (and also independence, for more details see [16]) that is formally similar to Kolmogorov's measure-theoretic approach to probability theory.

4. Integration in numerical uncertainty theory

In the rest of this paper, we study the significance of Choquet integration and fuzzy integration for normal possibility measures, when they are considered as numerical uncertainty models. There are a number of such models, with different interpretations, in which it makes perfect sense to do that. Firstly, there is the behavioral theory of imprecise probabilities, which was brought to a synthesis by Walley [59, 60, 61], and where a normal possibility measure turns out to be a coherent *upper probability*, or in other words, a consistent system of upper betting rates. In the present section, we discuss the role of (Choquet and fuzzy) integration in imprecise probabilities in general. We look at the special case of possibility measures in the next section.

Secondly, there is the Bayesian sensitivity analysis, or robust Bayesian, interpretation. Many Bayesians take the view that all uncertainty should be modeled by probability measures. But given that, it may be difficult in practice to pinpoint such a probability measure exactly. For this reason, classes of probability measures are sometimes considered, of which it is assumed that the (unknown) ideal probability measure is a member. The main concern then is that whatever is inferred from the model should be robust under small changes in the model (whence the name for the interpretation). Upper probabilities – and in particular normal possibility measures – arise as upper envelopes of classes of probabilities. It turns out that from a formal mathematical point of view, the behavioral interpretation of imprecise probabilities and the sensitivity analysis interpretation often lead to the same results. For this reason, we do not treat this interpretation separately, but rather alongside with imprecise probabilities.

Random sets [25, 45] are a third numerical model normal possibility measures fit well into [19, 31, 49]. Integration in this context is discussed in Section 6.

Finally, normal possibility measures also have a part in Shafer's theory of evidence, as consonant plausibility functions [54]. Unfortunately, it is not at all obvious what interpretation should be given to integration in this context, unless plausibility functions are given the random set interpretation or are interpreted as (coherent) systems of upper betting rates. We shall therefore not be explicitly concerned with evidence theory in this paper.

We now turn to integration in the behavioral theory of imprecise probabilities. In order to keep this text reasonably self-contained, we begin with a brief discussion of the behavioral interpretation of upper probabilities and upper previsions, and of the important notions of coherence and natural extension. For a more detailed exposition, we refer to Walley's book on the subject [60].

4.1. **Upper probabilities.** Let us assume that we have a subject who is uncertain about something; say for the sake of the argument that he is uncertain about the outcome of an experiment. We denote by Ω the set of all possible outcomes, also called *possibility space*. Although our subject is uncertain about which element of Ω will be the actual outcome, his knowledge about the world in general and the experiment in particular may cause him to maintain certain beliefs about this outcome, and he may indicate the strengths of these beliefs by engaging in gambles about it. The simplest gambles he may engage in are bets on or against events $A \subseteq \Omega$, and he may indicate which betting rates are acceptable to him by specifying his upper and lower probabilities for A.

The subject's *lower probability* $\underline{P}(A)$ of the event $A \subseteq \Omega$ is his marginally highest rate for betting *on* A, that is, the highest number x such that he will be disposed to bet *on* A at any rate $r < x$. Betting on A at rate r yields a gain $1-r$ if A occurs and a loss r if it doesn't. (We assume that gains and losses are expressed in units of some linear utility.) The *upper probability* $\overline{P}(A)$ is the smallest number y such that our subject will be disposed to bet *against* A (or on the complementary event co$A$) at any rate $r < 1 - y$. In other words, $1 - \overline{P}(A)$ is the marginally highest rate for betting on co$A$: $\underline{P}(\text{co}A) = 1 - \overline{P}(A)$. Since this relationship between upper and lower probabilities is generally valid, we shall restrict ourselves to upper probabilities.

Let $G(A) = \overline{P}(A) - I_A$, where I_A is the (real-valued) indicator function of A. $G(A)$ is a real-valued function on Ω. The value of $G(A)$ in ω gives the net reward from a bet against A at betting rate $1 - \overline{P}(A)$, if the actual outcome of the experiment turns out to be ω: the result is a gain of $\overline{P}(A)$ if A does not occur ($\omega \notin A$) and a loss of $1 - \overline{P}(A)$ if A does occur ($\omega \in A$). Thus, $\overline{P}(A)$ could also be defined as the subject's infimum betting rate for taking bets on A, or equivalently, as the marginally lowest price for which the subject accepts to sell the gamble I_A.

If we have an upper probability $\overline{P}$ defined on a set of events $\mathcal{A}$, the behavioral interpretation of $\overline{P}$ is that the gamble $G(A)$ is marginally acceptable for all $A \in \mathcal{A}$, and that non-negative linear combinations of such gambles are also at least marginally acceptable. (By *marginally* acceptable we mean that $G(A) + \epsilon$ is acceptable for any $\epsilon > 0$.)

Because upper probabilities represent a subject's behavioural dispositions regarding bets against events, they should satisfy a number of consistency criteria. These can be summarized through the following requirement. Let $\overline{P}$ be a subject's upper probability assessment on a collection $\mathcal{A}$ of events. Then for any natural number[6] $n \geq 1$ and non-negative real numbers $\lambda_o, \ldots, \lambda_n$, for any events $A_o, A_1, \ldots, A_n$ in $\mathcal{A}$ it must hold that:

$$\sup\left[\sum_{k=1}^{n} \lambda_k G(A_k) - \lambda_o G(A_o)\right] \geq 0, \tag{8}$$

where, as indicated before, we denote by $\sup[X]$ the supremum of a bounded real-valued function X over its domain Ω.

To see why this condition is a significant consistency criterion, let us first look at the case $\lambda_o = 0$. If the condition fails, there are non-negative $\lambda_1, \ldots, \lambda_n$ and $A_1, \ldots, A_n$ in $\mathcal{A}$ such that the net reward of the (at least) marginally acceptable corresponding non-negative linear combination of gambles $G(A_1), \ldots, G(A_n)$ is strictly negative: $\sum_{k=1}^{n} \lambda_k G(A_k) < -\epsilon$ for some $\epsilon > 0$. In other words, there is a non-negative linear combination of marginally acceptable gambles which is certain to produce an overall loss! This is clearly an undesirable situation, which can only be avoided by imposing the above condition for $\lambda_o = 0$. For this reason, if the condition holds for $\lambda_o = 0$, $\overline{P}$ is said to *avoid sure loss* on $\mathcal{A}$.

---

[6] It is immaterial whether we take $n \geq 0$ or $n \geq 1$ in this definition.

Next, we turn to the case $\lambda_o > 0$. If the condition fails, there are non-negative $\lambda_1, \dots, \lambda_n$ and $A_o, A_1, \dots, A_n$ in $\mathcal{A}$ and $\epsilon > 0$ such that

$$\sum_{k=1}^{n} \lambda_k G(A_k) \leq \lambda_o[(\overline{P}(A_o) - \epsilon) - I_{A_o}].$$

Since the left-hand side is (at least marginally) acceptable, the right-hand side will be (at least marginally) acceptable too, as it represents a gamble with a higher net reward. This means that the subject should be effectively disposed to take bets on A_o at a (marginal) rate $\overline{P}(A_o) - \epsilon$, which is strictly lower than his infimum betting rate $\overline{P}(A_o)$: in specifying $\overline{P}(A_o)$ the subject did not take into account the implications of his other upper probability assessments. This produces a kind of logical inconsistency, which is not as bad as incurring a sure loss, but should nevertheless be avoided and/or corrected. If the condition holds for all $\lambda_o \geq 0$, we say that $\overline{P}$ *is coherent* on $\mathcal{A}$. Note that coherence implies avoiding a sure loss.

Coherence implies the following properties, which we shall need hereafter. Let $\overline{P}$ be a coherent upper probability defined on an arbitrary set $\mathcal{A}$ of events:

1. if $\emptyset \in \mathcal{A}$ then $\overline{P}(\emptyset) = 0$;
2. if $\Omega \in \mathcal{A}$ then $\overline{P}(\Omega) = 1$;
3. $0 \leq \overline{P}(A) \leq 1$;
4. if $A \subseteq B$ then $\overline{P}(A) \leq \overline{P}(B)$ [monotonicity];
5. if $\text{co}A \in \mathcal{A}$ then $\overline{P}(A) + \overline{P}(\text{co}A) \geq 1$, or equivalently, $\underline{P}(A) \leq \overline{P}(A)$;

for any A and B in $\mathcal{A}$. A much more detailed discussion of the consequences of coherence can be found in [60].

The following examples will serve to illustrate the notion of coherence. They also introduce a number of notions that will play a central role in the rest of the paper.

Example 3. Let $\overline{P}$ be an upper probability, defined on a field $\mathcal{V}$ of subsets of Ω, that is *2-alternating*:

1. $\overline{P}(\emptyset) = 0$ and $\overline{P}(\Omega) = 1$;
2. $\overline{P}(A \cup B) + \overline{P}(A \cap B) \leq \overline{P}(A) + \overline{P}(B)$, for all A and B in $\mathcal{V}$.

Then $\overline{P}$ is coherent (for a proof, see [59]). Note that any upper probability $\overline{P}$ that is *maxitive*, that is, $\overline{P}(A \cup B) = \max\{\overline{P}(A), \overline{P}(B)\}$ for A and B in $\mathcal{V}$, and has $\overline{P}(\emptyset) = 0$ and $\overline{P}(\Omega) = 1$ is 2-alternating and therefore coherent. As a consequence, normal possibility measures are coherent (and in particular avoid sure loss) when interpreted as upper probabilities, and can be meaningfully used in a behavioral theory of imprecise probabilities. This seems to have been first noted by Giles [35].

Example 4. Let $\overline{P}$ be an upper probability, defined on a field $\mathcal{V}$ of subsets of Ω, that only assumes the values 0 and 1. Then $\overline{P}$ is coherent if and only if the subset $\mathcal{I} = \{A \in \mathcal{V} \colon \overline{P}(A) = 0\}$ is an *ideal*, that is,

1. $\mathcal{I}$ is decreasing: if $A \in \mathcal{V}$ and $B \in \mathcal{I}$ then $A \subseteq B$ implies $A \in \mathcal{I}$;
2. $\mathcal{I}$ is closed under (finite) unions: if $A \in \mathcal{I}$ and $B \in \mathcal{I}$ then $A \cup B \in \mathcal{I}$,

that is furthermore *proper*: $\mathcal{I} \neq \mathcal{V}$ or equivalently, $\Omega \notin \mathcal{I}$ (for a proof, see [60]). In that case $\overline{P}$ is maxitive and therefore 2-alternating. The proper ideal $\mathcal{I}$ is the set of events which our subject is absolutely certain will not occur, since he is disposed to bet against them at any odds. If $\mathcal{V} = \wp(\Omega)$, $\overline{P}$ will be a normal possibility measure if and only if the associated proper ideal $\mathcal{I}$ is complete, that is, closed under arbitrary unions. In that case, the proper ideal $\mathcal{I}$ is a principal ideal, which means that there is some $O \subset \Omega$ such that $\mathcal{I} = \{A \subseteq \Omega \colon A \subseteq O\}$, and the possibility measure $\overline{P}$ has distribution $I_{\text{co}O}$. If $O = \emptyset$, $\overline{P}$ is called the *vacuous* upper probability, and we have that $\overline{P}(A) = 1$ unless $A = \emptyset$. $\overline{P}$ is then the greatest coherent upper probability defined on all events and it is a vacuous belief model with minimal behavioral implications: if our subject has $\overline{P}$ as his upper probability, this means that he is disposed to bet against any event $A \neq \emptyset$ only at the trivial rate 0, so that the reward function $G(A) = 1 - I_A$ is non-negative (meaning he cannot loose from such a bet).

Example 5. Let P be a *finitely additive probability (measure)* – or probability charge [1] – on a field $\mathcal{V}$ of subsets of Ω. In other words, P is a map from $\mathcal{V}$ to $[0, 1]$ which satisfies the following properties:

1. $P(\emptyset) = 0$ and $P(\Omega) = 1$;
2. $P(A \cup B) + P(A \cap B) = P(A) + P(B)$ for A and B in $\mathcal{V}$.

Then P is in particular 2-alternating and therefore coherent when interpreted as an upper probability on $\mathcal{V}$. If P is $0-1$-valued, then its associated proper ideal $\mathcal{I} = \{A \in \mathcal{V} \colon P(A) = 0\}$ is a prime (or maximal) ideal: for every $A \in \mathcal{V}$, either A or its set-theoretic complement co$A$ belongs to $\mathcal{I}$.

It follows from the properties of a finitely additive probability P that for any A in its domain $\mathcal{V}$, $P(\text{co}A) = 1 - P(A)$: $P(A)$ is at the same time the upper and the lower probability of A, and is therefore a *fair* betting rate in the sense of de Finetti [24]: $P(A)$ is the unique number x such that our subject will be disposed to bet on A at any rate $r < x$ and against A at any rate $s < 1 - x$.

Many Bayesians assume that whatever a subject's state of information, he can always specify his beliefs in terms of betting rates that are fair, or precise. In allowing a subject's lower and upper betting rates to be different, the behavioral theory of upper and lower (or imprecise) probabilities makes no such strong and stringent assumptions about his beliefs (for a more detailed discussion, see [60, Chapter 5]). Nevertheless, finitely additive probabilities do have an important role in the imprecise theory, as we shall see further on. We denote by $\mathbb{P}(\mathcal{V})$ the set of all finitely additive probabilities on the field $\mathcal{V}$.

4.2. **Upper previsions.** Up to now, we have looked at our subject's behavior with respect to very simple gambles, namely betting on events, where the reward functions assume only two values, depending on whether the events occur or not. It turns out that a much more complete model of our subject's beliefs can be obtained if we look at his behavior with respect to more complicated gambles, which may distinguish between any possible number of situations, rather than just two. In general, the reward for such a gamble can be represented by a bounded real-valued map X on Ω, where $X(\omega)$ gives the (possibly negative) reward (in units of some linear utility) when the outcome of the experiment turns out to be ω. We shall identify gambles with their reward functions, and call *gamble* on Ω any bounded real-valued map on Ω. The set of all gambles on Ω will be denoted by $\mathcal{L}(\Omega)$.

The *upper prevision* $\overline{P}(X)$ of a gamble X is defined as the infimum selling price for X, that is, the marginally lowest price for which a subject is disposed to sell the gamble X. It is the smallest number x such that the subject is disposed to sell X for all prices $p > x$. Similarly, his *lower prevision* $\underline{P}(X)$ for X is defined as the supremum buying price for X. Since buying a gamble X for a price p is the same thing as selling $-X$ for the price $-p$, we must have the following general relationship between lower and upper previsions: $\underline{P}(X) = -\overline{P}(-X)$. For this reason, we restrict ourselves to upper previsions.

Consider the real-valued function $G(X) = \overline{P}(X) - X$, which represents the net reward from selling X for the price $\overline{P}(X)$: when the actual outcome of the experiment turns out to be ω, the selling transaction results in a (possibly negative) gain $G(X)(\omega) = \overline{P}(X) - X(\omega)$. If we have an upper prevision $\overline{P}$ defined on a set of gambles $\mathcal{K}$, then the behavioral interpretation of $\overline{P}$ is that the gamble $G(X)$ is marginally acceptable for all $X \in \mathcal{K}$, or in other words that the subject accepts the gamble $G(X) + \epsilon$ for any $\epsilon > 0$, and that any non-negative linear combination of such gambles is again (at least) marginally acceptable.

When the gamble X is $0-1$-valued, and therefore an indicator function I_A of some subset A of Ω, we can identify the upper prevision $\overline{P}(I_A)$ and the upper probability $\overline{P}(A)$. Upper probabilities are special upper previsions defined on (indicator functions of) events. In what follows we shall also often identify (use the same notation for) an event A and its indicator function I_A.

Since an upper prevision $\overline{P}$ defined on a set of gambles $\mathcal{K}$ represents a subject's dispositions to sell gambles, it should satisfy a number of consistency (or rationality) criteria. The conditions we impose are natural generalizations of the ones imposed on upper probabilities: for any natural number[7] $n \geq 1$ and non-negative real numbers $\lambda_o, \ldots, \lambda_n$, and for any gambles $X_o, X_1, \ldots, X_n$ in $\mathcal{K}$ it must hold that

$$\sup\left[\sum_{k=1}^{n} \lambda_k G(X_k) - \lambda_o G(X_o)\right] \geq 0. \tag{9}$$

If this condition holds for $\lambda_o = 0$, we say that the upper prevision $\overline{P}$ on $\mathcal{K}$ *avoids sure loss*; we say that $\overline{P}$ is *coherent* if the condition holds for any $\lambda_o \geq 0$. The justification of avoiding sure loss and coherence as rationality criteria for upper previsions is essentially the same as the one for upper probabilities given above.

Coherent upper previsions have the following interesting properties, which we shall have occasion to use further on. If $\overline{P}$ is a coherent upper prevision defined on an arbitrary set $\mathcal{K}$ of gambles, it satisfies:

1. $\inf[X] \leq \overline{P}(X) \leq \sup[X]$;
2. if $X \leq Y$ then $\overline{P}(X) \leq \overline{P}(Y)$ [monotonicity];
3. if $X + Y \in \mathcal{K}$ then $\overline{P}(X + Y) \leq \overline{P}(X) + \overline{P}(Y)$ [subadditivity];
4. if $\lambda > 0$ and $\lambda X \in \mathcal{K}$ then $\overline{P}(\lambda X) = \lambda\overline{P}(X)$ [positive homogeneity];
5. if $-X \in \mathcal{K}$ then $\overline{P}(X) + \overline{P}(-X) \geq 0$, or equivalently, $\underline{P}(X) \leq \overline{P}(X)$;

for all X and Y in $\mathcal{K}$. Again, a more detailed discussion of the consequences of coherence can be found in [60].

4.3. **Extension of upper probabilities to upper previsions.** Any integral associated with an upper probability can be interpreted as a way to extend the upper probability to an upper prevision, using the information contained in the upper probability. In a behavioral context, a minimal requirement that such an integration process must satisfy, is that it should yield upper previsions that are *coherent*, especially if we start out with coherent upper probabilities.

In the rest of this section, we discuss a number of ways to extend an upper probability defined on a field $\mathcal{V}$ of subsets of Ω to

- an upper prevision on $\mathcal{B}(\Omega, \mathcal{V})$: for natural extension and Choquet integration;
- an upper prevision on $\mathcal{B}^+(\Omega, \mathcal{V})$: for fuzzy integration and in particular Shilkret integration.

Natural extension allows us to extend the upper probability $\overline{P}$ by taking only two things into account: (i) the information contained in $\overline{P}$, and (ii) the requirement of coherence. Consider any gamble X on Ω. Assume that p is our subject's infimum selling price for X, which means that the gamble $p - X$ is marginally acceptable to him. Coherence requires that this new assessment should be compatible with the upper probability assessments made previously, in the sense that for any $n \geq 0$,

[7]Again, it is immaterial whether we take $n \geq 0$ or $n \geq 1$ in this definition.

any non-negative real $\lambda_1, \dots, \lambda_n$ and any $A_1, \dots, A_n$ in $\mathcal{V}$:

$$p \leq \sup \left[X + \sum_{k=1}^{n} \lambda_k G(A_k) \right].$$

If this were not the case, there would be some $\epsilon > 0$ such that

$$(p - \epsilon) - X \geq \sum_{k=1}^{n} \lambda_k G(A_k).$$

The left-hand side represents an (at least) marginally acceptable gamble, since it dominates a non-negative linear combination of marginally acceptable gambles, which should be (at least) marginally acceptable. This implies that our subject's upper probability assessments imply that he should be disposed to sell X for a price $p - \epsilon$ strictly lower than p, and this is in conflict with his assessment of p as an *infimum* selling price for X. If we take into account all the assessments implicit in the upper probability $\overline{P}$, we find that coherence imposes the following upper bound on p: $p \leq \overline{E}_N(X)$, where

$$\overline{E}_N(X) = \inf \left\{ \sup \left[X + \sum_{k=1}^{n} \lambda_k G(A_k) \right] : n \geq 0, \lambda_k \geq 0, A_k \in \mathcal{V} \right\}. \qquad (10)$$

The functional $\overline{E}_N$ defined by (10) is called the *natural extension* of the upper probability $\overline{P}$. Note that the natural extension is defined for any gamble X on Ω. In the sequel, we shall restrict our attention to gambles that are $\mathcal{V}$-measurable, because we want to compare natural extension with Choquet and fuzzy (or Shilkret) integration. The procedure of natural extension can also be generalized to extend any upper prevision from its domain to a larger set of gambles.

Natural extension derives it importance from the following result, which is a special case of a theorem due to Walley [60, Theorem 3.1.2].

Theorem 3. *Let $\overline{P}$ be an upper probability on a field $\mathcal{V}$ of subsets of Ω. Assume that $\overline{P}$ avoids sure loss. Let $\overline{E}_N$ be the natural extension of $\overline{P}$.*

1. *$\overline{E}_N$ is a coherent upper prevision on $\mathcal{B}(\Omega, \mathcal{V})$ – or on $\mathcal{L}(\Omega)$ for that matter.*
2. *$\overline{E}_N$ is the greatest coherent upper prevision that is dominated by $\overline{P}$ on its domain $\mathcal{V}$.*
3. *$\overline{P}$ is coherent if and only if it coincides with its natural extension $\overline{E}_N$ on its domain $\mathcal{V}$.*
4. *If $\overline{P}$ is coherent then $\overline{E}_N$ is the greatest coherent upper prevision that coincides with $\overline{P}$ on its domain $\mathcal{V}$.*

This result implies that natural extension is least-committal in the following sense: any other coherent extension of a coherent upper probability $\overline{P}$ implies a disposition to sell gambles X for a price that is at least as low as $\overline{E}_N(X)$, and therefore has behavioral implications that are at least as strong. Natural extension allows us to extend a coherent $\overline{P}$ taking into account only the information in $\overline{P}$ and the requirements imposed by coherence. At the same time, if $\overline{P}$ is an upper probability that avoids sure loss but is not coherent, natural extension corrects and extends it into a coherent upper prevision, again in a manner which has minimal behavioral implications.

Example 6. Let $\overline{P}$ be the vacuous upper probability defined on $\wp(\Omega)$, that is, $\overline{P}(A) = 1$ if $A \neq \emptyset$ and $\overline{P}(\emptyset) = 0$ (see also Example 4). It is coherent, and its natural extension to $\mathcal{L}(\Omega)$ is given by $\overline{E}_N(X) = \sup[X]$, for any gamble X on Ω. This coherent upper prevision is called the *vacuous* upper prevision, and models a vacuous information state: since for any X the corresponding $G(X) =$

$\overline{E}_N(X) - X = \sup[X] - X \geq 0$, the reward for selling X for a price $\overline{E}_N(X)$ is always non-negative, whatever the outcome of the experiment, and $\overline{E}_N(X)$ is the marginally lowest price for which this is guaranteed.

Example 7. Let P be a finitely additive probability on a field $\mathcal{V}$ of subsets of Ω. When we interpret P as an upper probability on $\mathcal{V}$, its natural extension to $\mathcal{B}(\Omega, \mathcal{V})$ can be calculated by Choquet integration (for a proof, see [60, Section 3.2]). For any $X \in \mathcal{B}(\Omega, \mathcal{V})$:

$$\overline{E}_N(X) = (C)\int_\Omega X \mathrm{d}P = \int_{-\infty}^{+\infty} x \mathrm{d}F_X(x) = \inf[X] + \int_{\inf[X]}^{\sup[X]} P(\{X \geq x\})\mathrm{d}x, \tag{11}$$

where $F_X(x) = P(\{X < x\}) = 1 - P(\{X \geq x\})$. From now on, we denote this natural extension by $E_P(X)$. It should be noted that for any $A \in \mathcal{V}$, $E_P(I_A) = P(A)$, so E_P extends P; and that for all $X \in \mathcal{B}(\Omega, \mathcal{V})$ such that $-X \in \mathcal{B}(\Omega, \mathcal{V})$, $E_P(X) = -E_P(-X)$, so the upper prevision E_P coincides with the conjugate lower prevision and is therefore precise on $\mathcal{B}(\Omega, \mathcal{V}) \cap -\mathcal{B}(\Omega, \mathcal{V})$. Moreover, we have for $\lambda \in \mathbb{R}$ and X and Y in $\mathcal{B}(\Omega, \mathcal{V})$ that

1. $\inf[X] \leq E_P(X) \leq \sup[X]$;
2. if $\lambda X \in \mathcal{B}(\Omega, \mathcal{V})$ then $E_P(\lambda X) = \lambda E_P(X)$;
3. if $X + Y \in \mathcal{B}(\Omega, \mathcal{V})$ then $E_P(X + Y) = E_P(X) + E_P(Y)$.

$E_P(X)$ is also called the *expectation* of X, or the *prevision* of X (in the sense of de Finetti [24]). We shall have more to say about precise previsions in Section 4.6.

The functionals E_P on $\mathcal{B}(\Omega, \mathcal{V})$ associated with finitely additive probabilities P on $\mathcal{V}$ play an important part in the following alternative characterization of natural extension. Let $\overline{P}$ be an upper probability on $\mathcal{V}$, and denote by $\mathcal{M}(\overline{P})$ the set of all dominated finitely additive probabilities on $\mathcal{V}$:

$$\mathcal{M}(\overline{P}) = \{P \in \mathbb{P}(\mathcal{V}) \colon (\forall A \in \mathcal{V})(P(A) \leq \overline{P}(A))\}.$$

Walley [60, Theorems 3.3.3 and 3.4.1] has shown that $\overline{P}$ avoids sure loss if and only if $\mathcal{M}(\overline{P}) \neq \emptyset$, and that for any $X \in \mathcal{B}(\Omega, \mathcal{V})$ the natural extension $\overline{E}_N(X)$ of $\overline{P}$ to X is given by:

$$\overline{E}_N(X) = \max\{E_P(X) \colon P \in \mathcal{M}(\overline{P})\}. \tag{12}$$

In particular, we find for $A \in \mathcal{V}$ that

$$\overline{E}_N(A) = \overline{E}_N(I_A) = \max\{P(A) \colon P \in \mathcal{M}(\overline{P})\}. \tag{13}$$

As a consequence, we see, using Theorem 3, that an upper probability $\overline{P}$ on $\mathcal{V}$ is coherent if and only if it is the *upper envelope* of its set $\mathcal{M}(\overline{P})$ of dominated finitely additive probabilities. These results provide a formal link between the behavioral theory of imprecise probabilities and Bayesian sensitivity analysis.

4.4. **Coherence of the Choquet functional.** Can we use Choquet integration to coherently extend the coherent upper probability $\overline{P}$ on a field $\mathcal{V}$ to an upper prevision defined on $\mathcal{B}(\Omega, \mathcal{V})$? In other words, let us define the *Choquet functional* $\overline{E}_C$ on $\mathcal{B}(\Omega, \mathcal{V})$ as follows: for any $\mathcal{V}$-measurable gamble X,

$$\overline{E}_C(X) = (C)\int_\Omega X \mathrm{d}\overline{P} = \inf[X] + \int_{\inf[X]}^{\sup[X]} \overline{P}(\{X \geq x\})\mathrm{d}x.$$

For any $A \in \mathcal{V}$, $\overline{E}_C(I_A) = \overline{P}(A)$, so the Choquet functional $\overline{E}_C$ is an extension of $\overline{P}$ to $\mathcal{B}(\Omega, \mathcal{V})$. But is $\overline{E}_C$ coherent when interpreted as an upper prevision on $\mathcal{B}(\Omega, \mathcal{V})$, and if so, what is its relation to the natural extension $\overline{E}_N$? By combining (11) and (12) we find that for any $X \in \mathcal{B}(\Omega, \mathcal{V})$:

$$\begin{aligned}\overline{E}_N(X) &= \sup\{E_P(X) \colon P \in \mathcal{M}(\overline{P})\} \\ &= \inf[X] + \sup_{P \in \mathcal{M}(\overline{P})} \int_{\inf[X]}^{\sup[X]} P(\{X \geq x\}) \mathrm{d}x.\end{aligned}$$

If we exchange the supremum and integral operators, we get

$$\begin{aligned}&\leq \inf[X] + \int_{\inf[X]}^{\sup[X]} \sup\{P(\{X \geq x\}) \colon P \in \mathcal{M}(\overline{P})\} \mathrm{d}x \\ &= \inf[X] + \int_{\inf[X]}^{\sup[X]} \overline{E}_N(\{X \geq x\}) \mathrm{d}x \\ &= \inf[X] + \int_{\inf[X]}^{\sup[X]} \overline{P}(\{X \geq x\}) \mathrm{d}x = \overline{E}_C(X), \qquad (14)\end{aligned}$$

since we also made the assumption that $\overline{P}$ is coherent, and therefore coincides with its natural extension $\overline{E}_N$ on $\mathcal{V}$. We see that we can use Choquet integration to calculate the natural extension to $\mathcal{B}(\Omega, \mathcal{V})$ of a coherent upper probability on $\mathcal{V}$, *provided that* the above inequality is an equality, that is, provided that we can actually interchange the supremum and integral operators in the above expression. If, on the other hand, the above inequality is strict ($\overline{E}_N(X) < \overline{E}_C(X)$) for some $X \in \mathcal{B}(\Omega, \mathcal{V})$, then the Choquet functional $\overline{E}_C$ on $\mathcal{B}(\Omega, \mathcal{V})$ is *not coherent* when interpreted as an upper prevision, because it strictly dominates the greatest coherent extension $\overline{E}_N$ of $\overline{P}$ to $\mathcal{B}(\Omega, \mathcal{V})$. In other words, the Choquet functional $\overline{E}_C$ is only coherent when it coincides with the natural extension $\overline{E}_N$! This observation is the starting point for the following important result, proven by Walley [59].

Theorem 4. *Let $\overline{P}$ be coherent upper probability on $\mathcal{V}$. Then for all $X \in \mathcal{B}(\Omega, \mathcal{V})$, $\overline{E}_N(X) \leq \overline{E}_C(X)$. Moreover, $\overline{E}_C$ coincides with $\overline{E}_N$ on $\mathcal{B}(\Omega, \mathcal{V})$, and is therefore coherent, if and only if $\overline{P}$ is 2-alternating.*

In other words, the Choquet functional is very useful in a behavioral context when it is associated with a 2-alternating upper probability $\overline{P}$. In that case it coincides with natural extension, and therefore allows us to extend the upper probability to an upper prevision, taking only coherence into account. In the context of Bayesian sensitivity analysis, it allows us to calculate the upper envelope of the extensions E_P of the finitely additive probabilities P in the set $\mathcal{M}(\overline{P})$: for any $X \in \mathcal{B}(\Omega, \mathcal{V})$, $(C) \int X \mathrm{d}\overline{P} = \max\{E_P(X) \colon P \in \mathcal{M}(\overline{P})\}$. A related result for countably additive probabilities and 2-alternating capacities on complete separable metrizable spaces was first proven by Huber and Strassen [36].

On the other hand, if the coherent upper probability $\overline{P}$ is not 2-alternating, Choquet integration will produce incoherence, and is therefore in a behavioral context not useful as a tool for extending $\overline{P}$ to an upper prevision.

4.5. **Coherence of the fuzzy functionals.** Can a fuzzy integral be used to coherently extend an upper probability? More precisely, consider a coherent upper probability $\overline{P}$ defined on a field $\mathcal{V}$ of subsets of Ω. We consider the *fuzzy $\circledast$-functional* $\overline{E}_\circledast$ on the set $\mathcal{B}^+(\Omega, \mathcal{V})$ of non-negative $\mathcal{V}$-measurable gambles X, defined by

$$\overline{E}_\circledast(X) = (F_\circledast) \int_\Omega X \mathrm{d}\overline{P} = \sup_{x \geq 0} \left[x \circledast \overline{P}(\{X \geq x\}) \right],$$

where $\circledast$ is an isotonic binary operator on $\mathbb{R}^+$ with zero 0, as defined in Section 2.3. We interpret $\overline{E}_\circledast$ as an upper prevision on $\mathcal{B}^+(\Omega,\mathcal{V})$ and we want to know whether $\overline{E}_\circledast$ coherently extends $\overline{P}$.

The first step we take is to show that a fuzzy $\circledast$-functional $\overline{E}_\circledast$ can only coherently extend $\overline{P}$ to $\mathcal{B}^+(\Omega,\mathcal{V})$ if it coincides everywhere on its domain with the *Shilkret functional* $\overline{E}_S$, which is the fuzzy functional with algebraic product $\times$ as its binary operation:

$$\overline{E}_S(X) = \sup_{x\geq 0} x\overline{P}(\{X \geq x\}).$$

The Shilkret functional $\overline{E}_S$ actually extends a coherent $\overline{P}$: $\overline{E}_S(I_A) = \overline{P}(A)$ for all A in $\mathcal{V}$; and for all X in $\mathcal{B}^+(\Omega,\mathcal{V})$, $0 \leq \overline{E}_S(X) \leq \sup[X]$, so $\overline{E}_S(X)$ is finite.

Proposition 5. *Assume that the fuzzy $\circledast$-functional $\overline{E}_\circledast$ extends the coherent upper probability $\overline{P}$ on $\mathcal{V}$ to a coherent upper prevision on $\mathcal{B}^+(\Omega,\mathcal{V})$. Then $\overline{E}_\circledast = \overline{E}_S$.*

We can therefore concentrate on the Shilkret functional: for which underlying coherent upper probabilities is it coherent, and if coherent, what is its relation to the Choquet functional and to natural extension? The following example looks at the special case that $\overline{P}$ assumes only the values 0 and 1.

Example 8. For a coherent (and therefore maxitive and 2-alternating) $0-1$-valued upper probability $\overline{P}$ on a field $\mathcal{V}$, the natural extension $\overline{E}_N$ to $\mathcal{B}^+(\Omega,\mathcal{V})$ can be obtained both by Choquet and by Shilkret integration. For any $\mathcal{V}$-measurable non-negative gamble X we have:

$$\overline{E}_N(X) = \overline{E}_C(X) = \overline{E}_S(X) = \inf\{a\colon \{X \geq a\} \in \mathcal{I}\},$$

where $\mathcal{I} = \{A \in \mathcal{V}\colon \overline{P}(A) = 0\}$ is the proper ideal associated with $\overline{P}$.

It turns out that in general, the natural extension lies between the Shilkret and the Choquet functionals. This follows from Theorem 4 and the following proposition.

Proposition 6. *Let $\overline{P}$ be a coherent upper probability on a field $\mathcal{V}$ of subsets of Ω. Then for all $X \in \mathcal{B}^+(\Omega,\mathcal{V})$, $\overline{E}_S(X) \leq \overline{E}_N(X)$.*

A simple proof of this result involving the well-known Markov inequality in probability theory also gives a nice interpretation for the Shilkret functional. Let P be any finitely additive probability on $\mathcal{V}$ that is dominated on $\mathcal{V}$ by $\overline{P}$, that is, $P \in \mathcal{M}(\overline{P})$. Since, for any $a \geq 0$, $aI_{\{X\geq a\}} \leq X$, it follows by applying the increasing (because coherent) operator E_P to both sides of the inequality that

$$aP(\{X \geq a\}) \leq E_P(X). \tag{15}$$

This is the famous Markov inequality. Taking the supremum over $P \in \mathcal{M}(\overline{P})$ on both sides of the inequality yields $a\overline{E}_N(\{X \geq a\}) \leq \overline{E}_N(X)$, or equivalently,

$$a\overline{P}(\{X \geq a\}) \leq \overline{E}_N(X), \tag{16}$$

since $\overline{P}$ is coherent and therefore coincides on $\mathcal{V}$ with $\overline{E}_N$. If we then take the supremum over all $a \geq 0$, we find that $\overline{E}_S(X) = \sup_{a\geq 0} a\overline{P}(\{X \geq a\}) \leq \overline{E}_N(X)$, which completes the proof of the proposition.

It follows from (15) that the Shilkret integral associated with a probability P is the closest we can get to the prevision (or natural extension, or expectation) E_P by expressions of the form $aP(\{X \geq a\})$. Similarly, (16) tells us that the Shilkret integral associated with a coherent upper probability $\overline{P}$ is the closest we can get to its natural extension by expressions of the form $a\overline{P}(\{X \geq a\})$. But it so happens that we shall only rarely be able to actually reach the natural extension in this way for all $X \in \mathcal{B}^+(\Omega,\mathcal{V})$.

Indeed, the following theorem indicates that we have identified in Example 8 the *only* case where Shilkret integration leads to coherent extension of an upper probability.

Theorem 7. *Let $\overline{P}$ be a coherent upper probability on a field $\mathcal{V}$ of subsets of Ω. Then the following statements are equivalent.*

1. *The Shilkret functional $\overline{E}_S$ is a coherent upper prevision on $\mathcal{B}^+(\Omega, \mathcal{V})$.*
2. *$\overline{P}$ only assumes the values 0 and 1.*
3. *The Shilkret functional $\overline{E}_S$ coincides with the Choquet functional $\overline{E}_C$ on $\mathcal{B}^+(\Omega, \mathcal{V})$.*
4. *The Shilkret functional $\overline{E}_S$ coincides with natural extension $\overline{E}_N$ on $\mathcal{B}^+(\Omega, \mathcal{V})$.*

As a corollary, we may conclude that the fuzzy $\circledast$-functional $\overline{E}_\circledast$ can only coherently extend a coherent upper probability $\overline{P}$ if $\overline{P}$ is $0-1$-valued. If $\overline{P}$ indeed only assumes the values 0 and 1, then $\overline{E}_\circledast$ only coherently extends $\overline{P}$ to $\mathcal{B}^+(\Omega, \mathcal{V})$ provided $\overline{E}_\circledast$ coincides on $\mathcal{B}^+(\Omega, \mathcal{V})$ with the Shilkret functional $\overline{E}_S$, and therefore also with natural extension $\overline{E}_N$ and the Choquet functional $\overline{E}_C$.

4.6. **Linear previsions.** We conclude this general section on integration in the behavioral theory of imprecise probabilities with a few remarks about precise probabilities and previsions.

The set $\mathcal{L}(\Omega)$ is a real linear space under the point-wise addition of gambles and the point-wise scalar multiplication of real numbers with gambles. Linear functionals on $\mathcal{L}(\Omega)$, that is, linear maps ϕ from $\mathcal{L}(\Omega)$ to $\mathbb{R}$ that are positive ($X \geq 0 \Rightarrow \phi(X) \geq 0$) and have unit norm ($\phi(1) = 1$) are called *linear previsions*. Any linear prevision is coherent when interpreted as an upper prevision. It is also self-conjugate in the sense that the conjugate lower prevision $-P(-X) = P(X)$ is equal to the upper prevision. Conversely, it can be shown that any self-conjugate coherent upper prevision on $\mathcal{L}(\Omega)$ is a linear prevision [60, Theorem 2.8.2]. The linear previsions are the precise probability models on $\mathcal{L}(\Omega)$ and are previsions in the sense of de Finetti [24].

The restriction of a linear prevision to any field $\mathcal{V}$ of events is a finitely additive probability on $\mathcal{V}$. Conversely, for any finitely additive probability P on the field $\wp(\Omega)$, its natural extension – or the associated Choquet functional – E_P, with for all $X \in \mathcal{L}(\Omega)$,

$$E_P(X) = (C)\int_\Omega X \mathrm{d}P = \inf[X] + \int_{\inf[X]}^{\sup[X]} P(\{X \geq x\})\mathrm{d}x,$$

is a linear prevision. Finally, if we take a linear prevision ϕ, and consider its restriction P to $\wp(\Omega)$, then its natural extension E_P is identical to the linear prevision ϕ we started out with: any linear prevision is completely determined by the values it assumes on events. This is a special case of the general representation results of continuous linear functionals by Choquet (or Dunford) integrals associated with bounded finitely additive set functions (bounded charges). More details can be found in [1, Section 4.7], [26, Chapters 6–13], [34, Chapter 3] and [60, Sections 3.2 and 3.4].

5. Fuzzy and Choquet integration in numerical possibility theory

We now apply the results of the previous section to the special case where the upper probability is a normal possibility measure Π on Ω, with distribution π. We have seen in Example 3 that a normal possibility measure is always a coherent upper probability, since it is in particular maxitive and therefore 2-alternating. It therefore makes sense to interpret the numbers $\Pi(A)$ for $A \subseteq \Omega$ as upper betting rates for taking bets on A, or equivalently, as infimum selling precise for the gambles

I_A. This was first observed by Giles [35]. If Π is not normal, then $\Pi(\Omega) < 1$ and the corresponding reward function $G(\Omega) = \Pi(\Omega) - 1 < 0$, so Π incurs a sure loss! It therefore makes no sense to consider possibility measures that are not normal in this behavioral context. More details on the relation between possibility measures and imprecise probabilities can be found in [17, 18, 19, 61, 62].

On the behavioral interpretation, a normal possibility measure, as for that matter any other maxitive coherent upper probability, models a special type of situation: for any event A we have $\max\{\Pi(A), \Pi(\mathrm{co}A)\} = 1$, so the available evidence does not warrant betting against A ($\Pi(\mathrm{co}A) = 1$) or against its complement ($\Pi(A) = 1$).

Maxitive upper probabilities arise naturally as uncertainty models whenever the available evidence leads to the specification of a (coherent) upper probability $\overline{P}$ on a collection of nested events (a chain of sets) $\mathcal{A}$. In that case the natural extension of $\overline{P}$ to $\wp(\Omega)$, or in other words the greatest, or least-committal, coherent upper probability defined for all events that extends the upper probability assessment $\overline{P}$ on $\mathcal{A}$, will be maxitive. If $\mathcal{A}$ is finite, the natural extension is always a possibility measure. If not, a number of continuity assumptions must be imposed on the assessments to ensure that the natural extension would be a possibility measure. For more details, see [18, 61] and also [63], where it is argued that this type of situation typically occurs when modeling linguistic uncertainty, where the available information consists of simple affirmative statements in natural language.

Let us now investigate what roles the fuzzy and Choquet integrals may play in turning a normal possibility measure defined on $\wp(\Omega)$ into a coherent upper prevision.

Let $\circledast$ be an isotonic binary operator on $\mathbb{R}^+$ with zero 0, as described in Section 2.3. We also assume that $\circledast$ is completely distributive with respect to supremum, or in other words that the partial maps of $\circledast$ are left-continuous: for any non-empty family $\{b_j : j \in J\}$ and any a in $\mathbb{R}^+$, $a \circledast (\sup_{j\in J} b_j) = \sup_{j\in J}[a \circledast b_j]$ and $(\sup_{j\in J} b_j) \circledast a = \sup_{j\in J}[b_j \circledast a]$. Note that for any $A \subseteq \Omega$ and any non-negative gamble X:

$$(F_\circledast)\int_A X \mathrm{d}\Pi = \sup_{\omega\in A}[X(\omega) \circledast \pi(\omega)]. \tag{17}$$

Moreover, the fuzzy $\circledast$-functional $\overline{E}_\circledast$ associated with the normal possibility measure Π, and defined on the set $\mathcal{L}^+(\Omega)$ of all non-negative gambles on Ω, is supremum preserving in the following sense: if $\{X_j : j \in J\}$ is a non-empty family of non-negative gambles such that the point-wise supremum $\sup_{j\in J} X_j$ is bounded (and therefore a non-negative gamble), then $\overline{E}_\circledast(\sup_{j\in J} X_j) = \sup_{j\in J} \overline{E}_\circledast(X_j)$. But we have seen in the previous section that $\overline{E}_\circledast$ is not coherent as an upper prevision on $\mathcal{L}^+(\Omega)$ unless $\overline{E}_\circledast$ coincides with the Shilkret functional $\overline{E}_S$ and unless Π – and therefore also π – assumes only the values zero and one. In that case, if we let $E = \{\omega \in \Omega : \pi(\omega) = 1\} \neq \emptyset$ then $\pi = I_E$ and for any non-negative gamble X, $\overline{E}_\circledast(X) = \overline{E}_S(X) = \sup_{\omega\in E} X(\omega)$, that is, $\overline{E}_\circledast$ is the vacuous upper prevision relative to the event E.

In a similar vein, it was shown in [18] that any upper prevision defined on the set of all non-negative gambles is at the same time supremum preserving and coherent if and only if it is of the above-mentioned form: the vacuous upper prevision relative to some event, or equivalently, the Shilkret functional (or Choquet functional or natural extension) associated with a normal possibility measure that only assumes the values 0 and 1. Results that can be related to this were proven in a different context by Shilkret, when he showed that his functional is the only one associated with maxitive set functions that extends them, is positively homogeneous and preserves countable suprema [55]. Related results, with modified positive homogeneity conditions, are discussed in [47].

It appears that the fuzzy integral has only a very minor part in the numerical theory of possibility, on a behavioral interpretation. Choquet integration, on the other hand, is very important in this context: since normal possibility measures are in particular 2-alternating, Theorem 4 tells us that it is equivalent to natural extension. In other words, we have for any gamble X in $\mathcal{L}(\Omega)$:

$$\overline{E}_N(X) = \overline{E}_C(X) = \inf[X] + \int_{\inf[X]}^{\sup[X]} \Pi(\{X \geq x\})\mathrm{d}x,$$

and $\overline{E}_C = \overline{E}_N$ is the greatest, or least-committal, coherent upper prevision that coincides with Π on all events. Note in particular that if $\mathcal{M}(\Pi)$ is the set of all finitely additive probabilities on $\wp(\Omega)$ that are dominated by Π on $\wp(\Omega)$, then $\overline{E}_C(X) = (C)\int X\mathrm{d}\Pi = \max\{E_P(X) \colon P \in \mathcal{M}(\Pi)\}$.

The following observations make the connection between possibility measures and Choquet integration even stronger. Consider the *vacuous* upper probability Γ on Ω as discussed in Example 4: for any $A \subseteq \Omega$,

$$\Gamma(A) = \begin{cases} 1 & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset. \end{cases}$$

Note that Γ is a possibility measure whose distribution is the constant $\Omega - \{1\}$-map I_Ω. We can use Γ and the Choquet integral to define, for any gamble X on Ω, the set function Γ_X in the following way. For any $A \subseteq \Omega$:

$$\begin{aligned} \Gamma_X(A) &= (C)\int_A X\mathrm{d}\Gamma \\ &= \inf[X]\Gamma(A) + \int_{\inf[X]}^{\sup[X]} \Gamma(A \cap \{X \geq x\})\mathrm{d}x \qquad (18) \\ &= \begin{cases} \sup_{\omega\in A} X(\omega) & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset, \end{cases} \qquad (19) \end{aligned}$$

that is, $\Gamma_X(A)$ is the Choquet integral of X on A with respect to Γ. Γ_X is a supremum preserving set function with distribution X, and a possibility measure if $\inf[X] \geq 0$ and $\sup[X] \leq 1$. In particular, if Π is a normal possibility measure with distribution π, then $\Pi = \Gamma_\pi$, and (18) can be rewritten as:

$$\Pi(A) = (C)\int_A \pi\mathrm{d}\Gamma = \int_0^1 \Gamma(A \cap \{\pi \geq x\})\mathrm{d}x. \qquad (20)$$

This could also be formulated as follows: π is the Radon-Nikodym derivative of Π with respect to Γ. This was first observed by Nguyen *et al.* [50].

We can use a combination of (18), (19) and (20) to derive an alternative expression for the Choquet functional associated with a possibility measure Π. For any gamble X:

$$\begin{aligned} \overline{E}_C(X) &= \inf[X] + \int_{\inf[X]}^{\sup[X]} \Pi(\{X \geq x\})\mathrm{d}x \\ &= \inf[X] + \int_{\inf[X]}^{\sup[X]} \left(\int_0^1 \Gamma(\{\pi \geq y\} \cap \{X \geq x\})\mathrm{d}y\right)\mathrm{d}x \end{aligned}$$

and since the function $f(x,y) = \Gamma(\{\pi \geq y\} \cap \{X \geq x\})$ is non-increasing and therefore Riemann-integrable on its domain, we can change the order of integration:

$$= \int_0^1 \left(\inf[X] + \int_{\inf[X]}^{\sup[X]} \Gamma(\{X \geq x\} \cap \{\pi \geq y\}) \mathrm{d}x \right) \mathrm{d}y$$

and since π is normal, $\{\pi \geq y\}$ is non-empty for all $y \in [0,1)$, so $\Gamma(\{\pi \geq y\}) = 1$,

$$\begin{aligned} &= \int_0^1 \Gamma_X(\{\pi \geq y\}) \mathrm{d}y \\ &= \int_0^1 \sup\{X(\omega) \colon \pi(\omega) \geq y\} \mathrm{d}y. \end{aligned}$$

In summary, we have just shown that

$$\inf[X] + \int_{\inf[X]}^{\sup[X]} \sup\{\pi(\omega) \colon X(\omega) \geq x\} \mathrm{d}x = \int_0^1 \sup\{X(\omega) \colon \pi(\omega) \geq x\} \mathrm{d}x. \tag{21}$$

This interesting expression was first proven by Walley [61, 62] and independently by De Cooman and Aeyels [19] who gave two proofs different from Walley's. All these proofs are arguably more complicated and less elegant than the new one given here.

Remark 1. Loosely speaking, (21) is an exchange property between π and X. This exchange property is an interesting consequence of Choquet integration that holds in more general contexts as well. Consider a monotone set function μ defined of a field $\mathcal{V}$ of subsets of Ω. Using a non-negative $\mathcal{V}$-measurable gamble X we can define a new monotone set function μ_X on $\mathcal{V}$: for A in $\mathcal{V}$,

$$\mu_X(A) = (C) \int_A X \mathrm{d}\mu = \mu(A) \inf[X] + \int_{\inf[X]}^{\sup[X]} \mu(A \cap \{X \geq x\}) \mathrm{d}x.$$

Contrary to what Denneberg states in [26, p. 132], if μ is 2-alternating then μ_X is 2-alternating as well[8] (a proof can be given through a straightforward direct verification, using the definition of 2-alternating behavior). In that case, we can consider the value $\mu_X(Y)$ of the Choquet functional associated with μ_X in the non-negative $\mathcal{V}$-measurable gamble Y: $\mu_X(Y) = (C) \int Y \mathrm{d}\mu_X$. The argument given above for $\mu = \Gamma$ can be repeated to prove the following general symmetry result, or exchange property: $\mu_X(Y) = \mu_Y(X)$, or more explicitly:

$$(C) \int Y \mathrm{d}\mu_X = (C) \int X \mathrm{d}\mu_Y,$$

with obvious notations. Using these notations, and restricting ourselves to non-negative gambles X, the exchange property (21) can be written more succinctly as $\Gamma_\pi(X) = \Gamma_X(\pi)$.

6. Possibility measures and random sets

To conclude this paper, we briefly indicate how integration can be introduced in possibility theory in the context of random set theory. That possibility measures fit nicely into the random set framework was pointed out indirectly by Shafer [54] in 1976, through the connection between a random set and the focal elements of a plausibility function, and much more explicitly by Nguyen in 1984 [49], and Dubois and Prade in 1987 [31]. But, at least to our knowledge, it was pointed out only recently [19] that the same connection also leads to an interesting justification for

[8] It turns out that this procedure for defining a new set function from an old one using Choquet integration not only preserves the 2-alternating behavior but many other properties as well. For a discussion, see [50, 64].

associating a Choquet integral with a possibility measure. Let us describe briefly how this comes about. For a more detailed discussion and explicit proofs, we refer to [19].

Consider a multi-valued mapping $\Gamma\colon [0,1] \to \wp(\Omega)$ satisfying the following basic properties:

(R1) if $x \leq y$ then $\Gamma(y) \subseteq \Gamma(x)$ [antitonicity];

(R2) $\Gamma(0) = \Omega$.

Define, for any $A \subseteq \Omega$, the *outer projection* $A^* \subseteq [0,1]$ of A on $[0,1]$ as follows:

$$A^* = \{x \in [0,1]\colon \Gamma(x) \cap A \neq \emptyset\}. \tag{22}$$

The properties of Γ imply that these outer projections are actually subintervals of $[0,1]$. Given any coherent upper probability $\overline{P}$ defined on a set of subsets of $[0,1]$ that includes these outer projections, we may define an upper probability[9] $\overline{P}^*$ on $\wp(\Omega)$ as follows: for $A \subseteq \Omega$,

$$\overline{P}^*(A) = \frac{\overline{P}(A^*)}{\overline{P}(\Omega^*)} = \frac{\overline{P}(\{x \in [0,1]\colon \Gamma(x) \cap A \neq \emptyset\})}{\overline{P}(\{x \in [0,1]\colon \Gamma(x) \neq \emptyset\})}.$$

This upper probability $\overline{P}^*$ satisfies $\overline{P}^*(\emptyset) = 0$ and $\overline{P}^*(\Omega) = 1$, and is always maxitive. It will be a normal possibility measure if $\overline{P}$ and/or Γ satisfy appropriate continuity conditions. This will in particular be the case if $\overline{P}$ is a (countably additive) probability measure P_o, defined on the Borel sets of $[0,1]$, that is furthermore absolutely continuous with respect to Lebesgue measure on $[0,1]$, and this observation leads us back to the essence of the results by Nguyen [49], and Dubois and Prade [31]. Note that P_o describes the random behavior of the variable Γ whose values are subsets of Ω, and therefore Γ is often called a random set [45].

But how can we extend these results from events to gambles? For any gamble X on Ω, we can define its *outer projection* X^* of X on $[0,1]$ as the following gamble on $[0,1]$:

$$X^*(x) = \begin{cases} \sup_{\omega \in \Gamma(x)} X(\omega) & \text{if } \Gamma(x) \neq \emptyset \\ 0 & \text{if } \Gamma(x) = \emptyset. \end{cases} \tag{23}$$

If we identify events with their indicator functions, (23) indeed generalizes (22). The properties of Γ ensure that X^* is a Borel function on $[0,1]$, that is therefore Lebesgue integrable with respect to the probability measure P_o over the compactum $[0,1]$. If we define the upper prevision P_o^* on $\mathcal{L}(\Omega)$ as follows, using Lebesgue (or Choquet) integration: $P_o^*(X) = E_{P_o}(X^*)/P_o(\Omega^*) = (C)\int_{[0,1]} X^* \mathrm{d}P_o/P_o(\Omega^*)$, then this upper prevision is a normal possibility measure Π when restricted to events, and moreover P_o^* can be obtained by Choquet integration with respect to Π: $P_o^*(X) = (C)\int_\Omega X \mathrm{d}\Pi$. P_o^* is therefore also the natural extension of Π to the set of all gambles.

7. Conclusion

The results in this paper allow us to draw this conclusion: the fuzzy integral is very important in ordinal possibility theory, and much less so in numerical possibility theory, whereas the converse is true for the Choquet integral. Indeed, Choquet integration of numerical possibility measures coincides with their natural extension, and therefore allows us to extend normal possibility measures to upper previsions, taking only the requirements of coherence into account. On the other hand, since it is based on the operations of addition and multiplication of real numbers, it seems hard to justify in a purely ordinal context. Nevertheless, we have shown that there

[9] For general Γ that need not satisfy $(R1)$ and $(R2)$ and for $\overline{P}$ that are probability measures, this is Dempster's method for defining upper probabilities [25].

is a kind of integration – fuzzy integration – that is formally related to it, which can be introduced in purely order-theoretic terms, and which is especially suited for working with ordinal possibility measures.

Acknowledgements

This paper presents research results of the Belgian Program on Inter-University Poles of Attraction initiated by the Belgian state, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with the author.

The author would like to thank Peter Walley, Dirk Aeyels, Hugo Janssen and Leen Gilbert for stimulating discussions.

Appendix A. Proofs of new results

Proof of Proposition 1. We first prove sufficiency. Let $\boxplus = \max$ and $\boxdot$ be a triangular seminorm on K. Note that since $(K, \leq)$ is a chain and since $\boxdot$ is non-decreasing, $\boxdot$ is in particular left-distributive with respect to max, in the sense that for all λ_1, λ_2 and μ in K:

$$\mu \boxdot \max\{\lambda_1, \lambda_2\} = \max\{\mu \boxdot \lambda_1, \mu \boxdot \lambda_2\}.$$

Let Π be a K-valued possibility measure on Ω. To show that the first condition is satisfied, let $s(\Omega) = \{s_1, \dots, s_n\}$ be the set of (different) values of s, and call the corresponding domains $D_k = s^{-1}(\{s_k\})$, $k = 1, \dots, n$. Then the partition $\{F_\ell \colon \ell = 1, \dots, m\}$ is finer than the partition $\{D_k \colon k = 1, \dots, n\}$. Denote $N_k = \{\ell \colon t_\ell = s_k\}$, then $\{N_k \colon k = 1, \dots, n\}$ is a partition of $\{1, \dots, m\}$ and $\bigcup_{\ell \in N_k} F_\ell = D_k$. It now follows from the commutativity and associativity of max and the left-distributivity of $\boxdot$ with respect to max that

$$\begin{aligned}
\max_{\ell=1}^{m}[t_\ell \boxdot \Pi(A \cap F_\ell)] &= \max_{k=1}^{n} \max_{\ell \in N_k}[t_\ell \boxdot \Pi(A \cap F_\ell)] \\
&= \max_{k=1}^{n} \max_{\ell \in N_k}[s_k \boxdot \Pi(A \cap F_\ell)] \\
&= \max_{k=1}^{n} \left[s_k \boxdot \max_{\ell \in N_k} \Pi(A \cap F_\ell)\right] \\
&= \max_{k=1}^{n}[s_k \boxdot \Pi(A \cap D_k)] = R(s; A).
\end{aligned}$$

We also have for arbitrary $A \subseteq \Omega$ that $R(I_A; \Omega) = \max\{1_K \boxdot \Pi(A), 0_K \boxdot \Pi(\text{co}A)\} = \Pi(A)$ (using the boundary properties of triangular seminorms) so the second condition is also satisfied. If s is a simple map and $A \subseteq B \subseteq \Omega$ then, with obvious notations, $\Pi(A \cap D_k) \leq \Pi(B \cap D_k)$ for $k = 1, \dots, n$, whence $s_k \boxdot \Pi(A \cap D_k) \leq s_k \boxdot \Pi(B \cap D_k)$ (using the isotonicity of triangular seminorms), and consequently $R(s; A) \leq R(s; B)$. If on the other hand $A \subseteq \Omega$, and s and t are simple maps, then there always is a partition $\{F_\ell \colon \ell = 1, \dots, m\}$ such that for each $\ell = 1, \dots, m$, both s and t are constant on F_ℓ. Call s_ℓ the value of s on F_ℓ and t_ℓ the value of t on F_ℓ. If s is dominated by t then $s_\ell \leq t_\ell$ for all $\ell = 1, \dots, m$. It follows from what we have proven previously that

$$R(s; A) = \max_{\ell=1}^{m}[s_\ell \boxdot \Pi(F_\ell \cap A)] \leq \max_{\ell=1}^{m}[t_\ell \boxdot \Pi(F_\ell \cap A)] = R(t; A).$$

since both max and $\boxdot$ are isotonic.

We continue with the necessity part. Let $\lambda \in K$ and choose $A \subseteq \Omega$ and Π such that $\Pi(A) = \lambda$ and $\Pi(\text{co}A) = 0_K$ (this is always possible since Ω contains more than one element). It then follows from the second condition that $\Pi(A) = R(I_A; \Omega)$, or equivalently, $\lambda = [1_K \boxdot \lambda] \boxplus [0_K \boxdot 0_K]$, whence, using the second part of (2) with $\mu = 0_K$,

$$\lambda = 1_K \boxdot \lambda. \tag{24}$$

Next, we use the third condition to show that both $\boxdot$ and $\boxplus$ must be isotonic. Consider $\lambda_1 \leq \lambda_2$ and $\mu_1 \leq \mu_2$ in K, and choose $A \subseteq B \subseteq \Omega$ and Π such that $\mu_1 = \Pi(A)$ and $\mu_2 = \Pi(B)$ (this is always possible since Ω contains at least two elements). The third condition tells us that $R(\lambda_1; A) \leq R(\lambda_2; B)$, or equivalently, $\lambda_1 \boxdot \mu_1 \leq \lambda_2 \boxdot \mu_2$, which tells us that $\boxdot$ is isotonic. On the other hand, we may choose Π such that it assumes the value 1_K on all non-empty sets, and choose a subset C of Ω such that both C and co$C$ are non-empty (possible since Ω contains more than one element). Let the simple map s_k assume the value λ_k on C and μ_k on co$C$, $k = 1, 2$, then s_1 is dominated by s_2 and it follows from the third condition that $R(s_1; \Omega) \leq R(s_2; \Omega)$, or equivalently, also using the first part of (2), $\lambda_1 \boxplus \mu_1 \leq \lambda_2 \boxplus \mu_2$, which tells us that $\boxplus$ is isotonic as well.

The isotonicity of $\boxdot$, together with (24) and the first part of (2) tells us that $\boxdot$ is a triangular seminorm on K. Consequently, for any $\mu \in K$, $\mu \boxdot 0_K = 0_K$, and if we substitute this into the second part of (2), we find that $\lambda \boxplus 0_K = \lambda$ for all $\lambda \in K$. Taken together with the commutativity, associativity and isotonicity of $\boxplus$ this tells us that $\boxplus$ is a triangular conorm on K.

We now use the first condition to show that $\boxplus = \max$. Let λ_1, λ_2 and μ be elements of K. Consider the constant $\Omega-\{\mu\}$-map, and let $\{F_1, F_2\}$ be a partition of Ω. Choose a possibility measure Π such that $\Pi(F_1) = \lambda_1$ and $\Pi(F_2) = \lambda_2$ (possible since Ω has more than one element). Then $R(\mu; \Omega) = \mu \boxdot \Pi(\Omega) = \mu \boxdot \max\{\lambda_1, \lambda_2\}$. The first condition tells us that we should have $R(\mu; \Omega) = [\mu \boxdot \Pi(F_1)] \boxplus [\mu \boxdot \Pi(F_2)] = [\mu \boxdot \lambda_1] \boxplus [\mu \boxdot \lambda_2]$ as well, whence

$$\mu \boxdot \max\{\lambda_1, \lambda_2\} = [\mu \boxdot \lambda_1] \boxplus [\mu \boxdot \lambda_2]. \tag{25}$$

If we let in particular $\lambda_1 = \lambda_2 = 1_K$ in (25) it follows from the first part of (2) that $\mu = \mu \boxplus \mu$. So $\boxplus$ must be idempotent, and the only idempotent triangular conorm is max (see [21] for a proof). □

Proof of Eq. (3). Consider, for $\alpha \in K$, the simple map s_α defined by $s(\omega) = \alpha \boxdot I_{\{h \geq \alpha\}}(\omega)$. Then $s_\alpha \leq h$ and $R(s_\alpha; A) = \alpha \boxdot \Pi(A \cap \{h \geq \alpha\})$, and consequently $\sup_{\alpha \in K}[\alpha \boxdot \Pi(A \cap \{h \geq \alpha\})] \leq \sup\{R(s; A) \colon s \leq h\}$.

Conversely, consider a simple map $s \leq h$. With obvious notations, we have for $k = 1, \ldots, n$ that $D_k \subseteq \{h \geq s_k\}$, whence $\Pi(A \cap D_k) \leq \Pi(A \cap \{h \geq s_k\})$, and therefore

$$\begin{aligned} R(s; A) &= \max_{k=1}^{n}[s_k \boxdot \Pi(A \cap D_k)] \\ &\leq \max_{k=1}^{n}[s_k \boxdot \Pi(A \cap \{h \geq s_k\})] \\ &\leq \sup_{\alpha \in K}[\alpha \boxdot \Pi(A \cap \{h \geq \alpha\})]. \end{aligned}$$

Consequently $\sup\{R(s; A) \colon s \leq h\} \leq \sup_{\alpha \in K}[\alpha \boxdot \Pi(A \cap \{h \geq \alpha\})]$. □

References

[1] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, London, 1983.

[2] C. W. Burrill. *Measure, Integration and Probability*. McGraw-Hill, New York, 1972.

[3] D. Candeloro and S. Pucci. Radon-Nikodym derivatives and conditioning in fuzzy measure theory. *Stochastica*, 11:107–120, 1987.

[4] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.

[5] L. J. Cohen. *The Probable and the Provable*. Clarendon Press, Oxford, 1977.

[6] B. De Baets. *Oplossen van vaagrelationele vergelijkingen: een ordetheoretische benadering [Solving fuzzy relational equations; an order-theoretic approach]*. PhD thesis, Universiteit Gent, Ghent, Belgium, 1995. In Dutch.

[7] B. De Baets, G. de Cooman, and E. E. Kerre. The construction of possibility measures from samples on $\mathcal{T}$-semi-partitions. *Information Sciences*, 106:3–24, 1998.

[8] B. De Baets, E. Tsiporkova, and R. Mesiar. Conditioning in possibility theory with strict order norms. *Fuzzy Sets and Systems*, 1999. to appear.
[9] G. de Cooman. *Evaluatieverzamelingen en -afbeeldingen – Een ordetheorische benadering van vaagheid en onzekerheid [Evaluation Sets and Mappings—An Order-Theoretic Approach to Vagueness and Uncertainty]*. PhD thesis, Universiteit Gent, Ghent, Belgium, 1993. In Dutch.
[10] G. de Cooman. The formal analogy between possibility and probability theory. In G. de Cooman, D. Ruan, and E. E. Kerre, editors, *Foundations and Applications of Possibility Theory – Proceedings of FAPT '95*, pages 88–98. World Scientific, Singapore, 1995.
[11] G. de Cooman. Generalized possibility and necessity measures on fields of sets. In N. C. Steele, editor, *Proceedings of the International ICSC Symposium on Fuzzy Logic (ISFL '95)*, pages A91–A98. ICSC Academic Press, Canada, 1995.
[12] G. de Cooman. Confidence relations and qualitative possibility. In *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT '96)*, pages 539–544. ELITE, Aachen, 1996.
[13] G. de Cooman. Confidence relations and ordinal information. *Information Sciences*, 104:241–278, 1997.
[14] G. de Cooman. Possibility theory I: the measure- and integral-theoretic groundwork. *International Journal of General Systems*, 25:291–323, 1997.
[15] G. de Cooman. Possibility theory II: conditional possibility. *International Journal of General Systems*, 25:325–351, 1997.
[16] G. de Cooman. Possibility theory III: possibilistic independence. *International Journal of General Systems*, 25:353–371, 1997.
[17] G. de Cooman. Possibilistic previsions. In *Proceedings of IPMU '98*, volume I, pages 2–9. Éditions EDK, Paris, 1998.
[18] G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 1998. Accepted for publication.
[19] G. de Cooman and D. Aeyels. A random set description of a possibility measure and its natural extension. *IEEE Transactions on Systems, Man and Cybernetics*, 1999. Accepted for publication.
[20] G. de Cooman and B. De Baets. Implicator and coimplicator integrals. In *Proceedings of IPMU '96*, volume III, pages 1433–1438. 1996.
[21] G. de Cooman and E. E. Kerre. Order norms on bounded partially ordered sets. *The Journal of Fuzzy Mathematics*, 2:281–310, 1994.
[22] G. de Cooman and E. E. Kerre. Possibility and necessity integrals. *Fuzzy Sets and Systems*, 77:207–227, 1996.
[23] G. de Cooman, E. E. Kerre, and F. Vanmassenhove. Possibility theory: an integral theoretic approach. *Fuzzy Sets and Systems*, 46:287–300, 1992.
[24] B. de Finetti. *Theory of Probability*, volume 1. John Wiley & Sons, Chichester, 1974. English Translation of *Teoria delle Probabilità*.
[25] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
[26] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic, Dordrecht, 1994.
[27] D. Dubois. Belief structures, possibility theory and decomposable measures on finite sets. *Computers and Artificial Intelligence*, 5:403–416, 1986.
[28] D. Dubois, L. Fariñas Del Cerro, A. Herzig, and H. Prade. An ordinal view of independence with application to plausible reasoning. In *Uncertainty in Artificial Intelligence (Proceedings of the Tenth Conference, University of Washington, Seattle, Washington, July 29–31, 1994)*, pages 195–203.
[29] D. Dubois and H. Prade. *Fuzzy Sets and Systems. Theory and Applications*. Academic Press, New York, 1980.
[30] D. Dubois and H. Prade. *Théorie des possibilités*. Masson, Paris, 1985.
[31] D. Dubois and H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279–300, 1987.
[32] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.
[33] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In Ph. Smets, editor, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1. Kluwer Academic, Dordrecht, 1998. to appear.
[34] N. Dunford and J. T. Schwartz. *Linear Operators*, volume I. Wiley Interscience, New York, 1957.
[35] R. Giles. Foundations for a theory of possibility. In M. M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 183–195. North-Holland, Amsterdam, 1982.
[36] P. J. Huber and V. Strassen. Minimax tests and the Neyman–Pearson lemma for capacities. *The Annals of Statistics*, 1:251–263, 1973.

[37] E. P. Klement and S. Weber. Generalised measures. *Fuzzy Sets and Systems*, 40:375–394, 1991.

[38] I. Levi. On potential surprise. *Ratio*, 8:107–129, 1966.

[39] I. Levi. Support and surprise: L. J. Cohen's view of inductive probability. *British Journal for the Philosophy of Science*, 30:279–292, 1979.

[40] I. Levi. Potential surprise: its role in inference and decision making. In L. J. Cohen and M. Hesse, editors, *Applications of Inductive Logic*, pages 1–27. Clarendon Press, Oxford, 1980.

[41] I. Levi. Consonance, dissonance and evidentiary mechanisms. In P. Gärdenfors, B. Hansson, and N.-E. Sahlin, editors, *Evidentiary Value; Philosophical, Judicial and Psychological Aspects of a Theory*, pages 27–43. Gleerups, Lund, 1983.

[42] D. K. Lewis. *Counterfactuals*. Basil Blackwell, Oxford, 1973.

[43] D. K. Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2, 1973.

[44] M. Loève. *Probability Theory*. D. Van Nostrand, New York, 1963.

[45] G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, New York, 1975.

[46] R. Mesiar. Choquet-like integrals. *Journal of Mathematical Analysis and Applications*, 194(2):477–488, 1995.

[47] R. Mesiar. Possibility measures based integrals – theory and applications. In G. de Cooman, D. Ruan, and E. E. Kerre, editors, *Foundations and Applications of Possibility Theory – Proceedings of FAPT '95*, pages 99–107. World Scientific, Singapore, 1995.

[48] R. Mesiar and J. Šipoš. Radon-Nikodym-like theorem for fuzzy measures. *Journal of Fuzzy Mathematics*, 1:873–878, 1993.

[49] H. T. Nguyen. On modeling of linguistic information using random sets. *Information Sciences*, 34:265–274, 1984.

[50] H. T. Nguyen, N. T. Nguyen, and T. Wang. On capacity functionals in interval probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:359–377, 1997.

[51] E. Pap. *Null-Additive Set Functions*. Kluwer Academic Publishers, Dordrecht, 1995.

[52] D. Ralescu and G. Adams. The fuzzy integral. *Journal of Mathematical Analysis and Applications*, 75:562–570, 1980.

[53] G. L. S. Shackle. *Decision, Order and Time in Human Affairs*. Cambridge University Press, Cambridge, 1961.

[54] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[55] N. Shilkret. Maxitive measures and integration. *Indagationes Mathematicae*, 33:109–116, 1971.

[56] F. Suárez García and P. Gil Álvarez. Two families of fuzzy integrals. *Fuzzy Sets and Systems*, 18:67–81, 1986.

[57] M. Sugeno. *The Theory of Fuzzy Integrals and Its Applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

[58] M. Sugeno and T. Murofushi. Pseudo-additive measures and integrals. *Journal of Mathematical Analysis and Applications*, 122:197–222, 1987.

[59] P. Walley. Coherent lower (and upper) probabilities. Statistics Research Report 22, University of Warwick, Coventry, 1981.

[60] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[61] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

[62] P. Walley. Statistical inferences based on a second-order possibility distribution. *International Journal of General Systems*, 26:337–383, 1997.

[63] P. Walley and G. de Cooman. A behavioural model for linguistic uncertainty. In P. P. Wang, editor, *Computing with Words*. John Wiley & Sons, New York, 1998. Accepted for publication.

[64] Z. Wang, G. J. Klir, and W. Wang. Monotone set functions defined by Choquet integral. *Fuzzy Sets and Systems*, 81:241–250, 1996.

[65] S. Weber. Two integrals and some modified versions — critical remarks. *Fuzzy Sets and Systems*, 20:97–105, 1986.

[66] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

Universiteit Gent, Onderzoeksgroep SYSTeMS, Technologiepark 9, 9052 Zwijnaarde, Belgium

E-mail address: gert.decooman@rug.ac.be