

# Bipolar and bivariate models in multi-criteria decision analysis: descriptive and constructive approaches

Michel GRABISCH

Université Paris I - Panthéon-Sorbonne

Salvatore GRECO

Facoltà di Economia, Università di Catania, Italy

Marc PIRLOT<sup>\*†</sup>

Faculté Polytechnique de Mons, Belgium

December 20, 2006

## Abstract

Multi-criteria decision analysis studies decision problems in which the alternatives are evaluated on several dimensions or viewpoints. In the problems we consider in this paper, the scales used for assessing the alternatives with respect to a viewpoint are bipolar and univariate or unipolar and bivariate. In the former case, the scale is divided in two zones by a neutral point; a positive feeling is associated to the zone above the neutral point and a negative feeling to the zone below this point. On unipolar bivariate scales, an alternative can receive both a positive and a negative evaluation, reflecting contradictory feelings or stimuli. The paper discusses procedures and models that have been proposed to aggregate multi-criteria evaluations when the scale of each criterion is of one of the two types above. We present both a constructive and a descriptive view on this question; the descriptive approach is concerned with characterizations of models of preference, while the constructive approach aims at building preferences by questioning the decision maker. We show that these views are complementary.

**Keywords:** Multiple criteria, Decision analysis, Preference, Bipolar models, Choquet integral.

---

<sup>\*</sup>The authors are indebted to Denis Bouyssou for helpful comments on an earlier version of this work. They also want to thank Didier Dubois for his insightful reading of the manuscript resulting in a number of suggestions that hopefully helped to improve this paper. The usual caveat applies.

<sup>†</sup>Corresponding author: email [Marc.Pirlot@fpms.ac.be](mailto:Marc.Pirlot@fpms.ac.be)



# 1 Introduction

Multi-criteria decision aid (or support) is concerned with the process of helping a decision maker (henceforth DM) understand a decision problem and clarify his preferences; helping him is required since the decision is obscured by the presence of several objectives and the fact that the alternatives show contrasted behaviour with respect to these objectives. The alternatives are usually described by their evaluation on a set of attributes; the latter are selected in such a way as to provide a description of the alternatives in all the aspects that are relevant for the decision that has to be made. Very often, at some stage of the process, each alternative has been associated a vector, the coordinates of which represent the assessments of the alternative on a set of criteria that encode the objectives of the DM. A crucial technical operation at that stage consists in producing a synthesis of the various aspects relevant to the decision; this is often called *aggregation*. The result of this operation may take different forms; it can be an evaluation of each alternative on a super-criterion; it can also be a relation on the set of alternatives, a pair of alternatives being in the relation if the comparison of the performances of both elements of the pair on the relevant criteria shows that the first element should be preferred over the second one.

The points of view on which the alternatives have been assessed may often be considered as being bipolar; for instance, there may exist a natural neutral point on the evaluation scale associated to the criteria; above this “zero” lies the zone of evaluations corresponding to a good performance and below, the zone corresponding to bad performance. Taking into account the bipolar nature of the scale often amounts to treating the good and the bad performances differently.

An example of the considered situation is the evaluation of a car on the basis of features such as maximum speed, acceleration, fuel consumption, price and so on. From a psychological viewpoint these features constitute a set of stimuli that combine to form the DM’s preferences on the set of alternatives. From the decision analysis point of view, these features are the attributes on which a comprehensive evaluation is based. Modelling the DM’s preferences on each attribute, amounts to translate the effect of the stimulus into a level on the scale associated to the corresponding criterion. The simplest way to do so is to consider a *unipolar scale* on which it is not possible to identify a reference point separating positive and negative evaluations. A *bipolar scale* is appropriate when the DM is able to say whether he perceives each alternative positively or negatively with respect to each viewpoint; for each criterion we have only a positive or a negative evaluation on a bipolar but univariate scale. In this case the positive or the negative sign of the evaluation derives from the comparison with a neutral reference point, such that all the values over the reference point are considered positive and all the values under the reference point are considered negative. For example, if we consider the maximum speed of a car, the reference neutral level can be 150 km/h. In this

case, a maximum speed over 150 km/h is considered positive, while a maximum speed under 150 km/h is considered negative.

In some cases, a DM might be unable to make a synthesis of possibly contradictory stimuli relative to a single viewpoint, which leads to using two scales for the evaluation of the alternatives on such a viewpoint: one for the positive stimulus and one for the negative stimulus associated to the same alternative. In a *bivariate context*, one can consider two different reference points, one for the positive values and another for the negative ones. Coming back to the example of the car, a maximum speed of 120 km/h can be considered as an undesirable level while a maximum speed of 180 km/h is a desirable level. Now, consider a car with a maximum speed of 150 km/h. In this case we have a positive evaluation, because the maximum speed is larger than the lower level of 120 km/h, and at the same time a negative evaluation, because the maximum speed is smaller than the higher level of 180 km/h. In case the DM is not able to make a synthesis of the good and the bad aspects on each viewpoint, two evaluations are needed for each criterion: one related to the positive part and another to the negative part.

Also with respect to the comprehensive evaluation one can consider a unipolar setting, a bipolar setting or a bivariate setting. In a unipolar setting, for each alternative  $x$  there is only one comprehensive evaluation  $CE(x)$  and it is not meaningful to distinguish between positive and negative comprehensive evaluations. In a bipolar setting there is only one comprehensive evaluation  $CE(x)$ , but it is meaningful to distinguish between a positive comprehensive evaluation ( $CE(x) > 0$ ), a negative comprehensive evaluation ( $CE(x) < 0$ ) or a neutral comprehensive evaluation ( $CE(x) = 0$ ). Nevertheless in a bipolar setting, it is not possible to decompose the comprehensive evaluation  $CE(x)$  as the net result of the aggregation of a comprehensive positive evaluation and a negative comprehensive evaluation. In a bivariate setting there are two evaluation axes, one for a comprehensive positive evaluation and the other for a comprehensive negative evaluation. Therefore, for each alternative  $x$ , the evaluations with respect to the considered criteria are aggregated in a comprehensive positive evaluation  $CPE(x)$ , a comprehensive negative evaluation  $CNE(x)$  and a comprehensive final evaluation  $CE(x)$  resulting from a further aggregation of  $CPE(x)$  and  $CNE(x)$ , representing the final net evaluation.

In this paper, we mainly examine procedures and models that have been proposed for aggregating multi-criteria information when the scales of the criteria are viewed as bipolar and univariate or unipolar and bivariate.

There are two different approaches for analyzing aggregation procedures. One is the so-called descriptive approach, that tries to model the observed behaviour of decision makers. It states conditions under which a preference can be represented in a given family of models. Relying on such results it proposes methods for eliciting the parameters of a particular model in a specific family of models. The second approach is called constructive. It proposes procedures for building a

preference relation taking into account the available information. In section 2, we briefly review the notion of a bipolar scale as a scale of evaluation of alternatives relatively to a particular point of view (criterion); we distinguish two ways of representing bipolarity: the bipolar univariate model and the (unipolar) bivariate model. Section 3 analyzes, in a descriptive perspective, the notion of bipolarity in a variety of preference models; models for the aggregation of bivariate scales are also described. In section 4, we present a number of procedures for constructing a preference based on assessments on bipolar univariate scales. Section 5 offers a broad view on all the models that can be thought of to describe preferences resulting from the aggregation of bivariate scales associated to criteria; the case where the global or comprehensive evaluation itself is bivariate is also envisaged. The paper ends with some comments on the complementary views brought by the descriptive and the constructive approach.

## 2 Bipolar scales in MCDA

In this section, we describe two ways of modelling the preferences of a DM on a criterion, taking into account the bipolar nature of the affect.

### 2.1 Two possible models

Many experiments in psychology have shown that our way of evaluating alternatives, objects and making decision is guided by *affect*. According to Slovic [37], affect is the “*specific quality of “goodness” and “badness”, as it is felt consciously or not by the decision maker, and demarcating a positive or negative quality of stimulus*”. This definition clearly reflects the *bipolar* nature of the affect, since it is built on two opposite poles (good/bad, positive/negative). Hence, the way we represent quantities on a scale in a given model of decision making, such as scores, preferences, etc., should reflect this bipolar nature. Measurement theory [24, 31] introduces scales as homomorphisms between a set of objects of interest equipped with some (e.g., binary) relation  $\succeq$  and the set of real numbers equipped with another relation  $\geq$  (e.g., the usual order). However, the distinction between bipolar scales and unipolar scales (i.e., with only one pole, which could be positive or negative) is not considered in this theory.

There are basically two ways of representing a bipolar notion on a scale: the *bipolar univariate model*, called Type I, (Osgood *et al.* [28]), and the *unipolar bivariate model*, called Type II (Cacioppo *et al.*, [8]). Scales of Type I consist of a single axis with a central 0 value, ranging from negative values to positive ones (see Fig. 1). On this scale, the

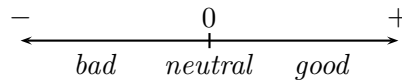


Figure 1: The bipolar univariate model

intensity of the affect is encoded by a number greater than 0 if it is felt as good or positive in some sense, and smaller than 0 if it is felt as bad or negative. The central 0 value is used to encode the neutral affect, which is neither good nor bad. More recently, the tendency is to use the *unipolar bivariate model*, introduced by Cacioppo *et al.* [8]. As shown on Fig. 2, this model uses two independent unipolar scales bounded below by 0. The horizontal axis encodes the intensity of the positive affect, while the vertical axis encodes the negative affect. Hence, one can have independent components for the positive and negative parts of the affect. The motivation for such a model is that a subject may feel at the same time for a given object (alternative, etc.) a positive feeling and a negative one. Yet the subject does not have a resulting feeling, which would be in some sense the summation of both. For example,

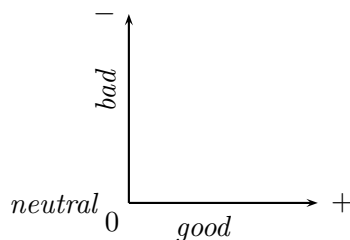


Figure 2: The unipolar bivariate model

eating chocolate gives some gustative pleasure (positive affect), while at the same time the subject may feel some greediness (negative affect)<sup>1</sup>.

## 2.2 Neutral level, boundedness

Coming back to the definition of the affect given above, besides the notion of poles, another important notion which is characteristic of the affect —and indeed a consequence of the presence of two opposite poles—, is the notion of *neutral level*. The neutral level is the point where the two sides (negative, positive) of the axis meet, i.e., it is the point which is felt by the decision maker neither with a positive nor a negative affect. Formally, let us consider a set  $A$  of objects and some binary relation  $\succsim$  on  $A$ , expressing for example the preference of a decision maker. We assume that  $\succsim$  is reflexive, complete and transitive. We consider a numerical scale  $f : (A, \succsim) \rightarrow (\mathbb{R}, \geq)$ , i.e., an order-preserving function. A *neutral level* is an object in  $A$  denoted by  $\mathbf{0}$ , such that for each object  $a \in A$ , if  $a \succ \mathbf{0}$ , then  $a$  is considered as “good” by the decision maker (positive affect), and if  $a \prec \mathbf{0}$ , then  $a$  is considered as “bad” (negative affect). Usually we set  $f(\mathbf{0}) = 0$ , which means that the value 0 on the scale represents the neutral level.

Our above definition of a neutral level is rooted in the notion of affect, and so always corresponds to the frontier between two paired and opposite notions of natural language related to the affect. Hence, a neutral level exists in particular with the following examples of binary relations: “more attractive than”, “better than”, “preferred than”, whose

<sup>1</sup>See however Section 4. A closer analysis of this example shows that it is better modeled in a multicriteria framework, that is, gustative pleasure and greediness (we may also speak of “effect on diet”) are considered as two criteria, each one being evaluated on a scale which is bipolar univariate or possibly only unipolar univariate. Indeed, “gustative pleasure” has a neutral level corresponding to the absence of taste, and two poles which could be called “delicious” and “disgusting, inedible”. Greediness in this context may be considered as unipolar, while “effect on diet” is bipolar with a neutral level separating good effects from bad effects. A good argument to be convinced that the chocolate example is not suited to unipolar bivariate scales is to replace “chocolate” by “steamed broccoli”. Many people do not consider steamed broccoli as giving a gustative pleasure, but everybody agrees that it is good for diet.

corresponding opposite pairs are respectively “attraction/repulsion”, “good/bad”, “like/dislike”. On the contrary, the relations “more permitted than”, “more priority than” are not related to affect, and thus have no neutral level in our sense (although they may have in another context, like logic. See [...] in this issue).

Another interesting feature of scales is boundedness. We say that a scale is *bounded above* (resp. *below*) if there exists an object  $\bar{a}$  such that  $\bar{a} \succsim a$  for every  $a \in A$  (resp. there exists  $\underline{a}$  such that  $\underline{a} \precsim a$  for every  $a \in A$ ). If a unipolar scale is bounded below, which is most often the case, we denote the lower bound by  $\mathbf{0}$ , and we usually set  $f(\mathbf{0}) = 0$ . When an upper bound exists for a scale (unipolar or bipolar), we may denote it by  $\mathbf{1}$ , and set  $f(\mathbf{1}) = 1$ . In the rest of the paper, the lower bound of a bipolar scale is denoted by  $-\mathbf{1}$  when it exists, and we set  $f(-\mathbf{1}) = -1$ . In the previous examples, relations “more attractive than”, “better than”, “preferred than” and “more priority than” are not bounded since it is always possible to find something more attractive, or better, or preferred to, or more priority than a given object. Besides, relation “more permitted than” is bounded, with greatest element “completely permitted”.

### 2.3 Net predisposition and the CPT model

Let us consider a unipolar bivariate model, taking as illustration the above “chocolate example”. We said that the subject feels at the same time a negative and a positive affect, without mixing them in a kind of overall resulting feeling. The simultaneous presence of positive and negative affects makes any decision difficult to make. Suppose now that the subject is forced to act, for example, he must decide whether to buy or not to buy some chocolate in a shop, or to order or not some desert with chocolate in a restaurant. Then the subject resolves the ambiguity caused by the presence of positive and negative affects, by mapping them onto a bipolar univariate scale. The resulting feeling is called the *net predisposition*. Then, if no other factor can influence the decision, the subject will act according to the position of the net predisposition with respect to the neutral level of the scale.

Experiments in psychology have shown that the net predisposition has a typical shape when represented as a function of the positive and negative affects (see Fig. 3). One can remark that the slope is steeper on the negative part, since in general the negative affect has more weight than the positive affect. This is clearly shown by point  $A$ , where both positive and negative affects have their maximal value, but the net predisposition is strictly negative.

Cumulative Prospect Theory (CPT) introduced by Tversky and Kahnemann [40] is an example of decision making model computing a net predisposition in a simple way. This model will be detailed in Section 4, and is expressed as follows:

$$\text{CPT}(f) := \mathcal{C}_{\mu_1}(f^+) - \mathcal{C}_{\mu_2}(f^-)$$

where  $f$  is a function expressing scores,  $f^+, f^-$  being their positive and negative parts, and  $\mathcal{C}_{\mu_i}$  is a function aggregating scores into an

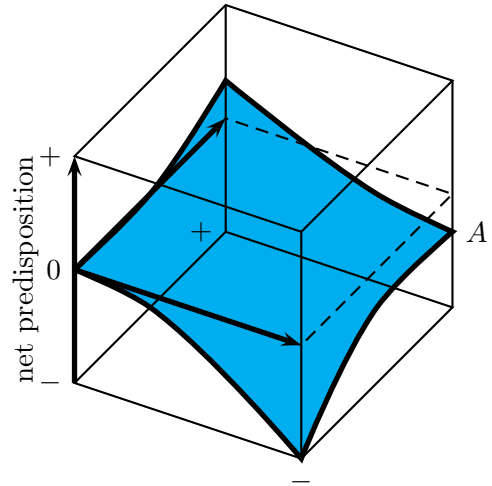


Figure 3: Net predisposition as a function of positive and negative affects

overall score. Hence,  $\mathcal{C}_{\mu_1}(f^+)$  represents the intensity of positive affect, while  $\mathcal{C}_{\mu_2}(f^-)$  represents the negative one, and the net predisposition is simply computed as a difference of the two parts. This is represented on Fig. 4. Note that this model is a very simple example of net predisposition, and does not exhibit a steeper slope for the negative part.



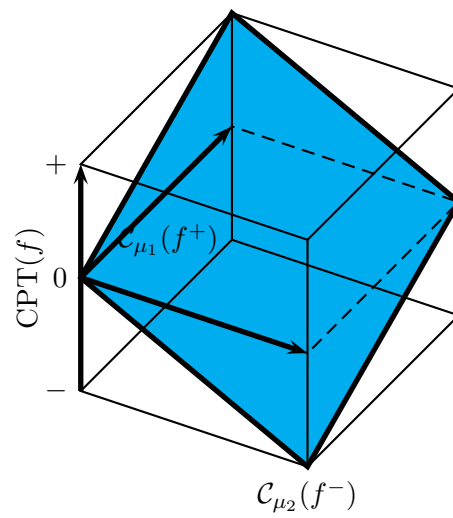


Figure 4: The CPT model viewed as a particular way of computing the net predisposition

### 3 A descriptive approach based on conjoint measurement

As anticipated in the introduction, the descriptive approach to MCDM, is a behaviourist one: it aims to describe the behaviour of a DM facing a decision problem. The only thing that can actually be observed is the preference that the DM expresses by choosing an object instead of another. This preference is usually represented by a relation on a set of objects (alternatives) and constitutes the main primitive of the theory. The role of the models is to relate the observable preference of the DM to a description of the objects in terms of their attributes. The conditions under which a model is applicable are formulated in terms of properties of the preference; they may thus be tested in practical situations by asking the DM whether he prefers or not some alternative over another alternative (most of the time, the compared alternatives are chosen in such a way that some answers or sequences of answers imply that a property is not satisfied and, consequently, lead to reject the model).

#### 3.1 The additive value function model

Suppose that in a given context, the alternatives can be described by means of  $n$  relevant attributes; an alternative is thus well represented by a vector  $x = (x_1, \dots, x_n)$ , where  $x_i$  describes the alternative on attribute  $i$ ;  $x_i$  ranges in a set  $X_i$ , that we call the domain of attribute  $i$ . The set of all attributes is denoted by  $N = \{1, \dots, n\}$ . The set  $X_i$  can be a set of numbers but it can also be a set of labels, ordered (e.g. “bad”, “average”, “good”) or unstructured (i.e. nominal labels). A preference  $\succsim$  is supposed to be a relation that contains all pairs  $(x, y)$  of alternatives (each of them corresponding to a  $n$  dimensional vector) such that the DM would choose  $x$  rather than  $y$ . Any combination of levels in  $X_i$  is supposed to correspond to an alternative, so that the set of all alternatives  $X$  can be identified with the Cartesian product  $\prod_{i=1}^n X_i$ . The following notation will be useful in the sequel; let  $a = (a_1, \dots, a_i, \dots, a_n)$  be any alternative; we denote by  $a_{-i}$  the  $n - 1$  components vector having the same components as  $a$  except on the  $i^{\text{th}}$  dimension that is lacking;  $X_{-i}$  is the set of all such vectors. We denote by  $(x_i, a_{-i})$  the alternative  $(a_1, \dots, x_i, \dots, a_n)$ , that is obtained from  $a$  by substituting its  $i^{\text{th}}$  component value by a level  $x_i \in X_i$ . In the same spirit, for any subset of attributes  $I \subset N$ , we denote by  $a_I$  (resp.  $a_{-I}$ ) the vector of components of  $a$  corresponding to the attributes in  $I$  (resp. in the complement  $-I$  of  $I$  in  $N$ );  $(x_I, a_{-I})$  is thus the alternative that has the same evaluations as  $x$  on the attributes in  $I$  and the same evaluations as  $a$  on the attributes that do not belong to  $I$ .

The model of preferences that is dominant is the *additive value function model*. A preference  $\succsim$  on  $X$  is representable in the additive value function model (also called additive utility model) if there are

functions  $u_i : X_i \rightarrow \mathbb{R}$  such that

$$x \succsim y \Leftrightarrow \sum_{i=1}^n u_i(x_i) \geq \sum_{i=1}^n u_i(y_i). \quad (1)$$

Sufficient conditions on  $\succsim$  are known from the literature, guaranteeing that a preference is representable in the above model [10, 24, 45]. Some necessary conditions are obvious, among which the fact that  $\succsim$  must be a *weak order* (a reflexive, complete and transitive relation) and that  $\succsim$  must fulfill the full *preference independence property*, i.e., for all  $x, y, a, b$  in  $X$  and for all subsets of attributes  $I$ :

$$(x_I, a_{-I}) \succsim (y_I, a_{-I}) \Rightarrow (x_I, b_{-I}) \succsim (y_I, b_{-I}). \quad (2)$$

The particular case in which  $I$  is a single attribute  $\{i\}$ , gives rise to the property known as *weak preference independence*, i.e., for all  $x, y, a, b$  in  $X$  and for all attribute  $i \in N$ :

$$(x_i, a_{-i}) \succsim (y_i, a_{-i}) \Rightarrow (x_i, b_{-i}) \succsim (y_i, b_{-i}). \quad (3)$$

Preference independence allows for *ceteris paribus* reasoning. This implies in particular, when preferences are supposed to be weak orders, that the functions  $u_i$  in model (1) are numerical representations of relations  $\succsim_i$  on  $X_i$  called *marginal preferences* and defined by:

$$x_i \succsim_i y_i \Leftrightarrow \forall a_{-i} \in X_{-i}, (x_i, a_{-i}) \succsim (y_i, a_{-i}). \quad (4)$$

A typical result in conjoint measurement is the following (see [24, theorem 13, p. 302] for additional definitions and proof).

**Theorem 1** *Let  $\succsim$  be a relation on the Cartesian product  $X = \prod_{i=1}^n X_i$ , with  $n \geq 3$ . If  $\succsim$  is an independent weak order that satisfies two additional conditions (restricted solvability and an Archimedean property) and provided the preference  $\succsim$  is non-degenerated (at least three components are “essential”), then there exist real-valued functions  $u_i$  on  $X_i$  such that for all  $x, y \in X$ ,*

$$x \succsim y \Leftrightarrow \sum_{i=1}^n u_i(x_i) \geq \sum_{i=1}^n u_i(y_i).$$

*Moreover, the functions  $u_i$  are unique up to a positive affine transformation, i.e. if  $u'_i$  is another such family of functions, then there exist numbers  $\alpha > 0$  and  $\beta_i \in \mathbb{R}$  for  $i \in N$ , such that:*

$$u'_i = \alpha u_i + \beta_i.$$

The practical interest of such results is that they provide a sound basis to the elicitation process of the model: in the present case, the independence hypothesis allows us to look for pairs of indifferent alternatives that differ only on two dimensions (all their other levels being equal) and this leads to building the  $u_i$  functions used in the model; once these functions have been obtained, we know that we have

found the “right” ones since theorem 1 states that they are (essentially) unique<sup>2</sup>.

### 3.2 Bipolarity in the additive value model

It should firstly be noted that formulating a bipolarity property in a conjoint measurement (descriptive) approach is not straightforward. Why is it so? Because, in the models alluded to above, no a priori structure, not even an ordering, is postulated on the attributes scales  $X_i$ ; the structure on  $X_i$  is eventually induced by the preference relation  $\succsim$  on  $X$ . For example, in the additive value model (1),  $\succsim$  induces a marginal preference relation  $\succsim_i$  on  $X_i$ ; the marginal preference orders the levels of the set  $X_i$  that may initially be linguistic labels. If bipolarity is to be considered relevant in a descriptive model, it must be revealed by, for instance, the marginal preferences or, in other terms, it must be “contained” in the global preference  $\succsim$ .

In the additive value model (1) for instance, a form of bipolarity, namely bipolar univariate scales, can possibly be detected in the articulation of a pre-existing structure on the scales  $X_i$  with the marginal preferences. Consider the following example. Suppose that  $X_i$  is the set  $\mathbb{R}$  of the real numbers. Suppose further that the elicitation process of the preference  $\succsim$  of the DM yields an additive value model in which function  $u_i : (X_i = \mathbb{R}) \rightarrow \mathbb{R}$  is shaped as illustrated in figure 5. The piece of the graph corresponding to positive values of  $x_i$  is not symmetric to that corresponding to negative values with respect to the origin; in other words  $u_i$  is not an odd function. This dissymmetry of  $u_i$  w.r.t.  $x_i = 0$  can be interpreted as revealing that the scale is bipolar: a distortion in the *a priori* ordering of the real numbers is needed to correctly reflect the DM’s marginal preferences on  $X_i$  (this is similar to the modeling of the attitude of the DM in decision making under risk: risk-prone in case of gains; risk-averse in case of losses; see e.g. [30]).

This simple example prompts two observations:

- there is room for bipolarity in the context of the additive value model but this supposes taking into account some *pre-existing structure* on the scales of the attributes;
- bipolarity is related to an interpretation of the relationship between marginal preferences and an a priori structure on the scale; this interpretation focuses on the behavior of the marginal preference above and below a “special point” of the scale.

#### 3.2.1 Pre-existing structure on $X_i$

In the example above, we supposed that  $X_i = \mathbb{R}$ . Which features of the pre-existing structure on  $X_i$  make it possible to interpret the link with the marginal preference  $\succsim_i$  as a manifestation of bipolarity? The ingredients needed are the following:

---

<sup>2</sup>This is the main result on which relies the elicitation method using the so-called *standard sequences* [44]

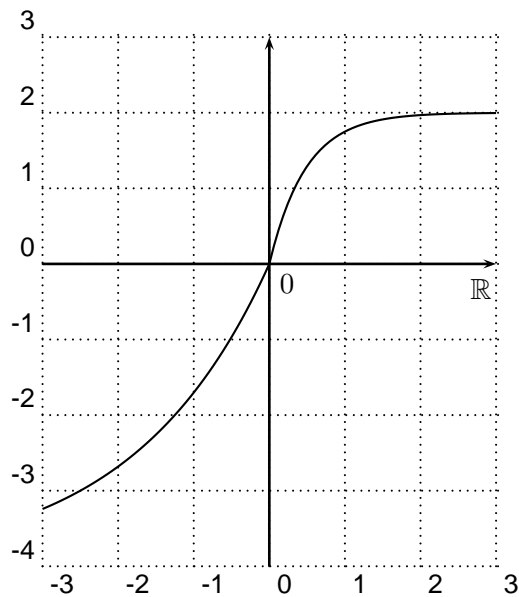


Figure 5: Example of bipolar univariate scale on the real numbers

- $X_i$  has to be an ordered set; let  $\geq_i$  denote the order on  $X_i$ ; this a priori structure is supposed to be “compatible” with marginal preferences  $\succsim_i$ , i.e., if  $x_i \geq_i y_i$ , one should not have  $y_i \succ_i x_i$  otherwise the a priori structure would be unrelated to the marginal preference;
- we need a “breakpoint” that separates  $X_i$  in two subsets  $X_i^-$  and  $X_i^+$ . Denoting the breakpoint by  $\mathbf{0}_i$  (it is actually the number 0 that plays this role in our example above), we call
  - $X_i^-$  the set of levels  $x_i < \mathbf{0}_i$  and
  - $X_i^+$  the set of levels  $x_i > \mathbf{0}_i$ ;
- finally, we need a bijective application  $\tau_i : X_i^+ \rightarrow X_i^-$  associating levels above and below  $\mathbf{0}_i$  by pairs; this application has to be antitone w.r.t. the order  $\geq_i$ , which means that, for all  $x_i, y_i \in X_i^+$ ,

$$x_i \geq_i y_i \quad \Rightarrow \quad \tau_i(y_i) \geq_i \tau_i(x_i). \quad (5)$$

Using the inverse  $\tau_i^{-1}$ , we may extend  $\tau_i$  to be defined on  $X_i^-$ : for  $x_i \in X_i^-$ , we define  $\tau_i(x_i)$  as being  $\tau_i^{-1}(x_i)$  which makes  $\tau_i$  an involution on  $X_i^- \cup X_i^+$ . The fixed point of this involution is  $\mathbf{0}_i$ . In other words  $\tau_i$  is the formalization of a *symmetry* of  $X_i$  w.r.t.  $\mathbf{0}_i$  and we shall refer to it as such; the symmetric  $\tau_i(x_i)$  of a level  $x_i$  on scale  $X_i$  will be denoted by  $x_i^T$ , for short.

**Remark** It can happen in practice that  $|X_i^+| \neq |X_i^-|$ . Such a case is a priori not compatible with the existence of a symmetry  $\tau_i$  unless  $X_i^+$  and  $X_i^-$  can be augmented by adding the “mirror levels” of  $X_i^-$  and  $X_i^+$ , respectively; this requires that the DM must be able to tell where he would locate the “mirror level”  $\tau_i(x_i^+)$  corresponding to any

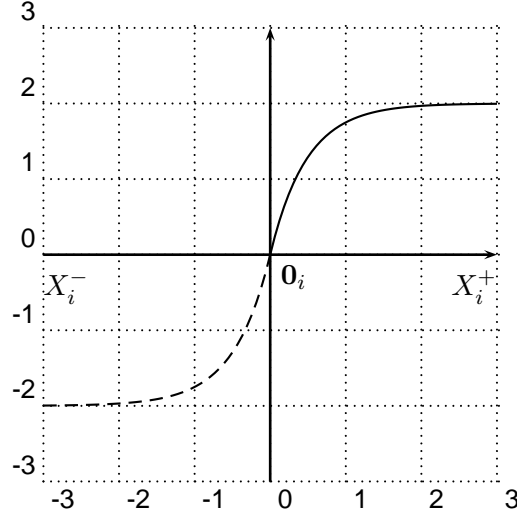


Figure 6: Bipolar univariate scale without distortion

$x_i^+ \in X_i^+$  in between levels of the scale  $X_i^+$ , and conversely for the mirror levels  $\tau_i(x_i^-)$ . This leads finally to working with the augmented level sets  $\tilde{X}_i^+ = X_i^+ \cup \tau_i(X_i^-)$  and  $\tilde{X}_i^- = X_i^- \cup \tau_i(X_i^+)$ , on which the involution, or symmetry,  $\tau_i$  is well-defined.

### 3.2.2 Using bipolarity for elicitation purposes

If we know (or assume) that the symmetry of the scale is linked with the marginal preferences, this can help in the elicitation of  $u_i$ . The most favorable case is of course when function  $u_i$  can be assumed to be antisymmetric w.r.t. the zero of the scale  $\mathbf{0}_i$ , i.e. when

$$u_i(x_i^T) = -u_i(x_i). \quad (6)$$

In the antisymmetric case, it suffices to elicit  $u_i$ , for instance on the positive part of the scale,  $X_i^+$ , and to extend  $u_i$  to the negative part,  $X_i^-$ , using (6) (the dashed curve on figure 6). Note that, in this case, there is no “distortion of the preference” around 0. The shape of the preference on the “negative consequences” is the mirror image of that on the “positive consequences” (see figure 6); distortion is not necessary for bipolarity.

In case the antisymmetry of  $u_i$  cannot be postulated, the benefits are more tiny yet they exist. The gain of complexity in the elicitation process is reduced since  $u_i$  has to be built on both parts of the scale. However, knowing in advance whether  $u_i(x_i^T) \geq -u_i(x_i)$  or, on the contrary,  $u_i(x_i^T) < -u_i(x_i)$  reduces the cognitive effort of the DM and reduces the risk of errors in his answers. This case is illustrated on figure 7; having elicited  $u_i$  on the positive part  $X_i^+$ , we use the symmetric image of the curve (the dashed line) to serve as a reference curve in the elicitation of  $u_i$  on the negative part  $X_i^-$ .

From the descriptive point of view adopted in this section, there is nevertheless a theoretical difficulty: the antisymmetry of  $u_i$  must

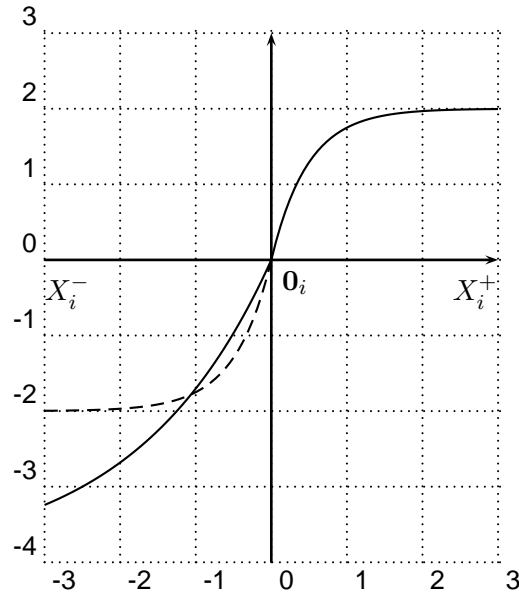


Figure 7: Eliciting the negative part using the symmetric image of the positive one as a reference curve

be expressed in terms of preference relations. This is not possible in terms of marginal preferences, but requires the definition of relations on *marginal preference differences*. Let us denote by  $\tilde{\succ}_i^*$  (see [45, 5]) the relation induced by  $\tilde{\succ}$  on  $X_i^2$ , which allows us to compare preference differences between pairs of levels of  $X_i$ ; its formal definition is as follows:

$$(x_i, y_i) \tilde{\succ}_i^* (z_i, w_i) \Leftrightarrow \begin{cases} \forall a_{-i}, b_{-i} \in X_{-i}, \\ (z_i, a_{-i}) \tilde{\succ} (w_i, b_{-i}) \Rightarrow (x_i, a_{-i}) \tilde{\succ} (y_i, b_{-i}) \end{cases} \quad (7)$$

From this relation, we define another one  $\tilde{\succ}_i^{**}$  that establishes a connection between a difference of preference  $(x_i, y_i)$  and the “opposite” difference,  $(y_i, x_i)$ :

$$(x_i, y_i) \tilde{\succ}_i^{**} (z_i, w_i) \Leftrightarrow (x_i, y_i) \tilde{\succ}_i^* (z_i, w_i) \text{ and } (w_i, z_i) \tilde{\succ}_i^* (y_i, x_i). \quad (8)$$

The antisymmetric condition (6) could be expressed by: for all  $x_i \in X_i$ ,

$$(\mathbf{0}_i, x_i) \sim_i^{**} (x_i^T, \mathbf{0}_i) \quad (9)$$

where  $\sim_i^{**}$  denotes the symmetric part of the relation  $\tilde{\succ}_i^{**}$ . In words, (9) expresses that the differences  $(\mathbf{0}_i, x_i)$  and  $(x_i^T, \mathbf{0}_i)$  are equal in terms of preference.

This formalism can of course be also used for expressing that differences of preference are not equal. However, there may be a way—provided  $X_i$  has sufficient richness—to turn  $u_i$  into an antisymmetric function. Provided a form of solvability<sup>3</sup> is satisfied, for all  $x_i$ , we can

<sup>3</sup>Roughly speaking, solvability conditions state that some “equations”, expressed in

find a level  $x'_i$  in  $X_i$  such that the difference of preference between  $\mathbf{0}_i$  and  $x_i$  is perceived as equal to that between  $x'_i$  and  $\mathbf{0}_i$ . Then, defining (or re-defining)  $\tau_i$  by putting  $x_i^\tau = x'_i$  yields a bipolar scale on  $X_i$  that satisfies (9). If this type of “pre-processing” can be performed (in a reliable manner), we only need to elicit marginal preferences (or the functions  $u_i$ ) on either the positive or the negative part of the corresponding scale.

### 3.3 The non-strict decomposable model

Clearly, the type of bipolarity considered in the framework of the additive value model corresponds to the *bipolar univariate model* (see section 2.1). In order to deal with other forms of bipolarity and to take into account other models, we introduce in this section a wider framework, the *(non-strict) decomposable value function model*. Certain preferences that will be described in section 4 do not admit a representation in the additive value model. Consider for instance the following procedure for comparing vectors of  $\mathbb{R}^n$  using the “min” operator, i.e.

$$x \succsim y \iff \min_{i \in N} x_i \geq \min_{i \in N} y_i. \quad (10)$$

The resulting preference is not independent. Indeed, let  $x = (5, 6, 7)$  and  $y = (4, 6, 7)$ ; we have  $x \succsim y$  and  $\text{Not}[y \succsim x]$ ; if we change the levels in common between  $x$  and  $y$ , namely 6 and 7, respectively into 3 and 7, then  $x' = (5, 3, 7) \sim y' = (4, 3, 7)$  since  $\min_{i \in N} x'_i = 3 = \min_{i \in N} y'_i$ . Condition (2) is not satisfied since we have  $y' \succsim x'$  but not  $y \succsim x$ ; ceteris paribus reasoning is not permitted. In reality, the “min” satisfies a weaker form of independence: a strict preference cannot be transformed into the opposite strict preference only by changing common levels; strict preference can only be changed into indifference. This weak form of independence is called *separability* (see [3], [6], and also [21]). A preference  $\succsim$  is separable if for all  $x, y, a, b$  in  $X$  and for all set of attributes  $I$ :

$$(x_I, a_{-I}) \succ (y_I, a_{-I}) \implies \text{Not}[(y_I, b_{-I}) \succ (x_I, b_{-I})]. \quad (11)$$

To deal with such preferences we have to widen the range of models that we consider. A rather general family is the non-strict decomposable model, introduced (in its strict form) in [24]. A preference  $\succsim$  is representable in the *non-strict decomposable model* (NSDM) if there are functions  $u_i : X_i \rightarrow \mathbb{R}$  and a non-decreasing function  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  such that:

$$x \succsim y \iff F(u_1(x_1), \dots, u_n(x_n)) \geq F(u_1(y_1), \dots, u_n(y_n)). \quad (12)$$

One verifies directly that such a preference is separable; it is independent as soon as  $F$  is an increasing function (instead of a non-decreasing one); the case in which  $F$  is increasing corresponds to the decomposable model introduced in [24]; we call it here the *strict decomposable*

---

terms of preferences, have a solution. Another form of solvability was required in theorem 1 as a condition for a preference being representable in the additive value model (1).



model. Obviously, the additive value model is a particular case of (strict) decomposable preference in which  $F$  is the sum of its arguments.

In the non-strict decomposable model, the  $u_i$ 's do not represent marginal preferences but more refined relations that are called marginal traces ([6]). The marginal trace  $\succsim_i^\pm$  on  $X_i$  is defined by:

$$x_i \succsim_i^\pm y_i \Leftrightarrow \begin{cases} \forall a_{-i} \in X_{-i}, b \in X, \\ (y_i, a_{-i}) \succ b \Rightarrow (x_i, a_{-i}) \succ b \\ \text{and} \\ b \succ (x_i, a_{-i}) \Rightarrow b \succ (y_i, a_{-i}). \end{cases} \quad (13)$$

In the (non-strict) decomposable model, marginal traces are weak orders that refine the marginal preferences, i.e.  $x_i \succsim_i^\pm y_i$  implies  $x_i \succsim_i y_i$ ; the  $u_i$  functions can be chosen to be numerical representations of the marginal traces  $\succsim_i^\pm$ .

The NSDM is general enough to encompass the models that will be analyzed in section 4 since it can be proved that any separable preference that is a weak order can be represented in model (12) (see [6, proposition 8], for more detail)

Note that there is considerably more freedom in the choice of  $F$  and  $u_i$  than it was the case with the additive value model. If  $F$  and  $u_i$  can be used for representing a preference  $\succsim$  in the NSDM, one can apply increasing transformations to each of the  $u_i$ 's and to  $F$  independently and use the resulting functions to represent the same preference  $\succsim$  in the NSDM. In more specific models, within the framework of NSDM, one may however hope for stronger uniqueness properties of the representation.

### 3.3.1 Bipolarity and the NSDM

What has been said about bipolarity (more precisely about bipolar univariate scales) in the additive value model remains essentially valid with the NSDM. Marginal preferences have to be substituted by marginal traces and the  $u_i$  functions represent the latter. The presence of bipolarity on (some of) the scales  $X_i$  might help in the elicitation of the  $u_i$ 's provided we can assume some relationship between the pre-existing bipolar structure on  $X_i$  and the marginal traces  $\succsim_i^\pm$ .

There is however an important difference that is linked with the (lack of) uniqueness properties of the representation in NSDM. In the additive value model, the  $u_i$ 's are essentially unique (determined up to a positive affine transformation, see theorem 1), while in NSDM, any representation of the weak order  $\succsim_i^\pm$  can do and these are determined up to an *increasing* transformation. This has the following consequence. Suppose that a preference on a product of bipolar scales has a representation in the NSDM with  $u_i$  satisfying (6); this is by no means the unique one and in particular, for any increasing transformation applied to  $u_i$  and yielding  $u_i'$ , there is a representation using  $u_i'$ ; in general,  $u_i'$  does not satisfy (6). So, in view of exploiting the bipolarity, one may be interested in representations within a model—for

instance, the NSDM, or a family that particularizes the NSDM—that satisfy additional constraints like (6).

We emphasize that fulfilling (6) is not a necessary condition for a representation of  $\succsim$  in NSDM, even when  $\succsim$  satisfies (9). Formally, from a descriptive point of view, respecting the bipolarity information could be translated into the following conditions expressed in terms of  $\succsim$  and its derived relations  $\succsim_i^\pm$  and  $\succsim_i^{**}$ :

$$x_i \succeq_i y_i \quad \Rightarrow \quad x_i \succsim_i^\pm y_i \quad (14)$$

and

$$(\mathbf{0}_i, x_i) \sim_i^{**} (x_i^\top, \mathbf{0}_i) \quad (15)$$

For satisfying these requirements, it is not necessary, in the NSDM, to impose that

$$u_i(\mathbf{0}_i) - u_i(x_i) = u_i(x_i^\top) - u_i(\mathbf{0}_i). \quad (16)$$

Indeed, if there is a representation of a preference  $\succsim$  within the NSDM that respects the bipolar information (14, 15) and, in addition, satisfies (16), then there are also other representations of the same preference that do not satisfy (16): namely, most of those obtained through applying an increasing transformation to the  $u_i$ 's; since the preference represented remains unchanged, property (15) remains satisfied. The explanation for this is that we have enough degree of freedom in the determination of the function  $F$  so that we can also take care of the fulfilment of (15).

This being said, in view of facilitating the elicitation process or diminishing the cognitive burden of the DM, one may think of forcing the  $u_i$ 's to satisfy (16) and determine  $F$  on the basis of that version of the  $u_i$ 's considered as fixed. This is only possible if that additional constraint is compatible with the other ones. Essentially, imposing (16) will be compatible with representing  $\succsim$  in a parameterized family of models if, among the representations of  $\succsim$  in that family, there is one in which differences of values of the function  $u_i$  can be used to represent the relation  $\sim_i^{**}$  i.e. if

$$(x_i, y_i) \sim_i^{**} (z_i, w_i) \quad \Leftrightarrow \quad u_i(x_i) - u_i(y_i) = u_i(z_i) - u_i(w_i) \quad (17)$$

In NSDM, this raises no difficulty, since increasing transformations leave enough freedom to take this constraint into account; the case of the “min” is similar. The same is true for the additive value model, in which all determinations of the  $u_i$ 's fulfill condition (17).

### 3.3.2 Bivariate scales (unipolar and bipolar)

The non-strict decomposable model offers good potential for the analysis of bivariate scales (unipolar and bipolar; see section 2.1 and other papers in this issue). This track has not been explored in detail so far; we limit ourselves to indicate two research directions in our descriptive framework; we borrow their denomination to P. Perny [29]. Section 5 presents specific models that fall into one or the other of the approaches described below.

**Compare then aggregate** A straightforward approach could consist in duplicating each dimension, yielding a model on  $2n$  components. Instead of  $X_i$ , we consider two sub-dimensions<sup>4</sup>, the positive part of dimension  $i$ ,  $X_i^+$  and the negative one,  $X_i^-$ . An alternative  $x$  is represented by a  $n$ -pairs-of-components vector  $x = ((x_1^+, x_1^-), \dots, (x_n^+, x_n^-))$ . As soon as a preference  $\succsim$  is a weak order that is separable w.r.t. the  $n$  dimensions, there is a representation in the NSDM, i.e. there are  $n$  functions  $u_i : (X_i^+, X_i^-) \rightarrow \mathbb{R}$  (of the pair of variables  $(x_i^+, x_i^-)$ ) and a non-decreasing function  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  such that:

$$x \succsim y \Leftrightarrow F(u_1(x_1^+, x_1^-), \dots, u_n(x_n^+, x_n^-)) \geq F(u_1(y_1^+, y_1^-), \dots, u_n(y_n^+, y_n^-)). \tag{18}$$

The (bivariate) marginal traces, represented by the functions  $u_i(x_i^+, x_i^-)$ , offer much potential to model various types of connections between the positive and negative part of attribute  $i$ . Note that  $u_i$  makes a synthesis, for dimension  $i$ , of the positive and negative aspects of an alternative along that dimension; it induces a complete ranking of the alternatives along that dimension.

A simple example of such a model is the case in which  $u_i(x_i^+, x_i^-)$  obtains as a difference of “utilities”:

$$u_i(x_i^+, x_i^-) = v_i^+(x_i^+) - v_i^-(x_i^-),$$

where  $v_i^+$  and  $v_i^-$  are numerical representations of the positive and negative levels, respectively, on a numerical scale. In a more general model, we have:

$$u_i(x_i^+, x_i^-) = \varphi_i(v_i^+(x_i^+), v_i^-(x_i^-)),$$

where  $\varphi_i$  is a real-valued function defined on  $\mathbb{R}^2$  (or a subset of it) that is nondecreasing in its first argument and nonincreasing in the second; it models the intuitive idea that an alternative is all the better as its positive aspects are stronger and its negative aspects are weaker. If there is a “zero” level on both scales (bivariate bipolar scale),  $\mathbf{0}_i^+$  on  $X_i^+$  and  $\mathbf{0}_i^-$  on  $X_i^-$ , one could set, without being restrictive,  $u_i(\mathbf{0}_i^+, \mathbf{0}_i^-) = 0 = \varphi_i(0, 0)$ . Some forms of symmetry of  $\varphi_i$  around  $(0, 0)$  could facilitate the elicitation of the marginal trace on dimension  $i$ .

**Aggregate then compare** Another line of research could focus on another decomposition of the numerical representation of the preference relation  $\succsim$ . Let us consider the case in which each dimension  $i$  has a positive set of levels  $X_i^+$  and a negative one  $X_i^-$  but the synthesis of the positive aspects and the negative ones is not done for each dimension; instead, a synthesis of the positive aspects is done through the various dimensions and similarly for the negative aspects; finally, the preference results from comparing the synthesis of the positive aspects to the synthesis of the negative ones. Formally, this is expressed

---

<sup>4</sup>These two sets cannot be considered being complementary subsets of a scale  $X_i$ , like in univariate bipolar scales; in bivariate bipolar scales, there is no such scale  $X_i$ .

by the following model:

$$x \succsim y \Leftrightarrow F(u^+(x_1^+, \dots, x_n^+), u^-(x_1^-, \dots, x_n^-)) \geq F(u^+(y_1^+, \dots, y_n^+), u^-(y_1^-, \dots, y_n^-)). \quad (19)$$

In this definition,  $u^+$  (resp.  $u^-$ ) represents the marginal trace of the preference on the  $n$ -dimensional space of positive aspects (resp. negative aspects) and  $F$  is a function of two variables that is nondecreasing in its first argument and nonincreasing in its second. The CPT model introduced in section 2.3 clearly pertains to this framework; in the CPT,  $F$  is the difference of its arguments and  $u^+$  (resp.  $u^-$ ) represents the intensity of the positive (resp. negative) affect.

### 3.4 Models based on preference differences

Although sections 4 and 5 will need only the NSDM model, we shall briefly describe, to conclude the present section, another important class of models in which bipolarity receives a very natural interpretation. There are conjoint measurement models that are based on a balance of differences of preference observed on each dimension. In these models, the main scale attached to each dimension  $i$  is the *univariate* scale of differences of preference  $X_i^2$ ; the levels on this scale are (directed) pairs  $(x_i, y_i)$  of levels belonging to another scale, that is used for assessing the alternatives on dimension  $i$ . The scale of differences  $X_i^2$  is naturally bipolar, because the difference  $(x_i, x_i) = (y_i, y_i)$ , for all  $x_i, y_i$ , is a natural zero of this scale. There is also a natural symmetry  $\tau_i$  on this scale, since the difference  $(x_i, y_i)$  is associated an “opposite” difference  $(y_i, x_i)$ ; one could thus define  $\tau_i(x_i, y_i)$  as being  $(y_i, x_i)$ . This, however, does not mean that the difference  $(y_i, x_i)$  is necessarily *perceived*, in terms of preference, as “ $-(x_i, y_i)$ ”;  $\tau_i$  is just an a priori correspondence between elements of a scale (see the discussion on the elicitation of a univariate bipolar scale in section 3.2.2).

A general model of preference based on preference differences is as follows:

$$x \succsim y \Leftrightarrow G(p_1(x_1, y_1), \dots, p_n(x_n, y_n)) \geq 0. \quad (20)$$

Conditions are known ([5], see also [7]) under which the functions  $G$  and  $p_i$  in the model have a number of natural properties. In particular, if the marginal traces on differences  $\succsim_i^*$  defined by (7) are representable weak orders, one can assume that function  $G$  is nondecreasing in its  $n$  arguments and that functions  $p_i$  are numerical representations of  $\succsim_i^*$ . We present three examples of well-known procedures that fit into such a model and see how bipolarity appears in the corresponding instantiations of the model. Note by the way that there is in general no guarantee of transitivity for preference relations described by models based on the aggregation of preference differences, as is well-known for majority rules.

**Majority rule or concordance relation** Let  $X_i$  be the set of real numbers  $\mathbb{R}$ . The weighted majority preference rule can be defined

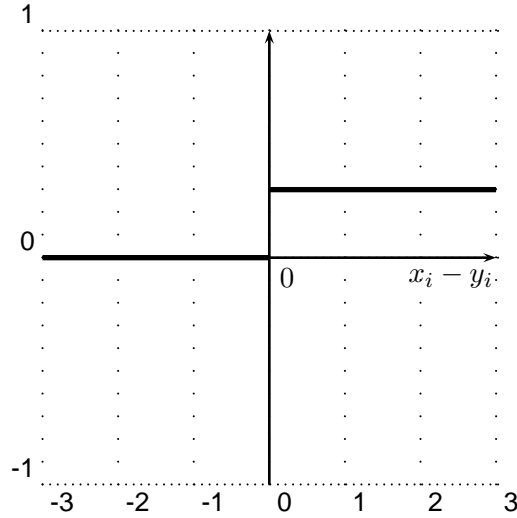


Figure 8: Univariate bipolar scale on differences for the majority rule

by means of a “concordance” index  $c(x, y)$  ([32, 42]):

$$c(x, y) = \sum_{i=1, \dots, n} p_i(x_i, y_i), \quad (21)$$

where

$$p_i(x_i, y_i) = \begin{cases} w_i & \text{if } x_i \geq y_i \\ 0 & \text{if } x_i < y_i; \end{cases}$$

the numbers  $w_i$  are the weights that are positive and sum up to 1.

The preference is defined through comparing the concordance index to a majority threshold  $K$ , a number lying between 0.5 and 1; we have:

$$x \succsim y \Leftrightarrow c(x, y) \geq K. \quad (22)$$

In this example,  $p_i$  determines a very simple univariate bipolar scale; the pairs  $(x_i, y_i)$  ordered by means of the weak order  $\succsim_i^*$  are represented on the horizontal axis; all pairs corresponding to the same difference  $x_i - y_i$  are indifferent with respect to  $\succsim_i^*$  and are thus represented by the same point labeled “ $x_i - y_i$ ” on the horizontal axis; the function  $p_i(x_i, y_i)$  is represented on figure 8. Only positive differences matter in this model and all such differences turn out to be equivalent; the negative differences are not at all taken into account.

**Concordance-discordance rule à la Electre** Building on the majority rule that has just been defined, we add *veto*s that will prevent declaring that  $x$  is preferred to  $y$  as soon as there is at least one dimension on which  $y$  is “much much better” than  $x$ . A veto is often described as the application of a *discordance rule* that operates in conjunction with a majority rule. This type of rule is at the basis of the ELECTRE methods ([32, 42]). More formally, the preference  $\succsim$  fits into

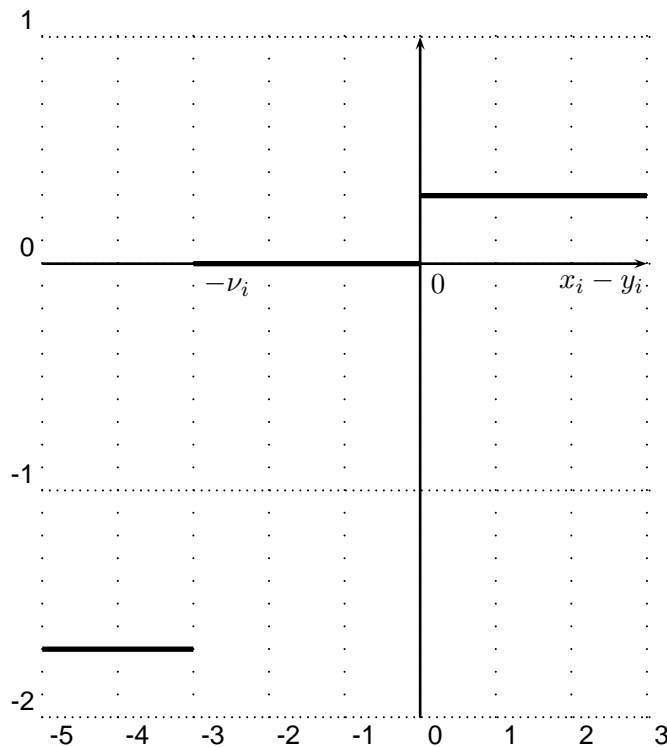


Figure 9: Univariate bipolar scale on differences for the concordance-discordance rule

the model described by equations (22) and (21) using the following definition for  $p_i$ :

$$p_i(x_i, y_i) = \begin{cases} w_i & \text{if } x_i \geq y_i \\ 0 & \text{if } y_i - \nu_i \leq x_i < y_i \\ -M & \text{if } x_i < y_i - \nu_i; \end{cases}$$

where  $M$  is a large positive number, for instance,  $M$  is larger than 1 if the weights  $w_i$  sum up to 1;  $\nu_i$  is a threshold used to determine discordance or veto: if  $x_i < y_i - \nu_i$ , this corresponds to the feeling that  $y_i$  is so much better than  $x_i$  on dimension  $i$  that it would make no sense to pretend that  $x$  could be globally preferred to  $y$ . The value of  $M$  must be chosen large enough in order to guarantee that inequality  $c(x, y) < K$  is never fulfilled as soon as at least one of the  $p_i$ 's is equal to  $-M$ . Figure 9 shows the shape of  $p_i$ .

**Tversky's model of additive differences** The preferences that are obtained using the rules described above cannot be supposed to be transitive in general. Indeed, it is well-known that performing pairwise comparisons of alternatives by means of majority rules may lead to preferences that have cycles (Condorcet paradox). This is obviously an undesirable feature if those preferences are to be used for decision purposes. However, on the basis of empirical evidence, several authors have argued that preferences cannot always be assumed

to be transitive and A. Tversky [39] was one of the first to propose a model generalising the additive value model and able to encompass preferences that lack transitivity. His model is known as the *additive difference model* in which,

$$x \succsim y \Leftrightarrow \sum_{i=1}^n \Phi_i(u_i(x_i) - u_i(y_i)) \geq 0, \quad (23)$$

where  $\Phi_i$  are increasing and odd functions.

Preferences that satisfy (23) may be intransitive but they are complete (due to the postulated oddness of  $\Phi_i$ ). This model of preferences, as compared to those based on majority rules, allows for taking preference differences into consideration more gradually. These are represented as algebraic differences of marginal utility functions  $u_i$ . Contrary to what happens with majority rules, the influence of  $(y_i, x_i)$  is exactly the opposite of that of  $(x_i, y_i)$  due to the oddness of the functions  $\Phi_i$ . Figure 6 can be used to illustrate this model, if we view it as representing preference differences instead of levels.

A slightly more general model that encompasses the majority rule is the so-called non-transitive additive model ([4, 11, 12, 13, 43]) in which:

$$x \succsim y \Leftrightarrow \sum_{i=1}^n p_i(x_i, y_i) \geq 0, \quad (24)$$

where the  $p_i$ 's are real-valued functions on  $X_i^2$  and may have several additional properties (e.g.  $p_i(x_i, x_i) = 0$ , for all  $i \in \{1, 2, \dots, n\}$  and all  $x_i \in X_i$ ). In this model, what corresponds to the oddness of the  $\Phi_i$ 's is the skew-symmetry of the  $p_i$ 's, i.e.  $p_i(x_i, y_i) = -p_i(y_i, x_i)$ . In case the skew-symmetry of the  $p_i$ 's is not postulated, figure 7 is a good illustration of the model if we represent all pairs  $(x_i, y_i)$  on the horizontal axis; those pairs can be ordered by means of the relation  $\succsim_i^{**}$  and we can decide as a convention that the points representing  $(x_i, y_i)$  and  $y_i, x_i$  will be symmetric w.r.t. the origin. In case  $p_i$  is not skew-symmetric, we then get the typical shape illustrated on figure 7.

## 4 A constructive approach to decomposable models based on capacities and related concepts

We take as a starting point the non strict decomposable model (NSD) given in (12). In this section we show how to construct such a model, taking into account the bipolarity of the affect if any. We will show that more general models can be thought of, depending on the type of bipolarity.

To build a NSD model, we need to solve two problems:

- build scales  $u_i$ 's on criteria in a unique way
- build aggregation function  $F$ .

A cornerstone in our construction is the determination on each criterion of two particular levels having an absolute meaning (that is, which does not depend on a particular criterion). One of them is the neutral level, already introduced in Section 2.2. It has an absolute meaning independent of the criteria since it corresponds for the DM to a state of feeling which is neither good nor bad. The second absolute level is introduced hereafter.

### 4.1 The satisficing level

A natural choice for the second absolute level would be to take the upper bound  $\mathbf{1}$  of each criterion (see Section 2.2), since it corresponds to the maximum satisfaction of the DM.

However, in many cases, there is no upper (or lower) bound for a scale. For example, the binary relations “is more priority than”, “is more attractive than” have no upper bound since one can always find something more priority (attractive) than a given object, so we have to find a substitute. This is given by the theory of *satisficing bounded rationality* of H. Simon [33, 34]. The theory asserts that, faced to a decision problem (e.g., which move to do in a chess game (chess player), where to go to find food (animal), etc.), due to the limited amount of information and limited resources (time, memory, intelligence, etc.), we are not able to *optimize* the decision w.r.t. some criterion, but only to reach a *satisficing* solution, i.e., which meets our level of satisfaction or aspiration. The search for a solution is stopped once such a satisficing solution has been found. The decision maker is aware that better solutions may exist, but limitations of resources prevent the DM to verify their existence. Coming back to unbounded scales, we denote by  $\mathbf{1}$  the satisficing level on a (unipolar or bipolar) scale which is not bounded above. By a kind of symmetry, we assume the existence of an *inacceptable level*, denoted by  $-\mathbf{1}$ , which is such that the decision maker rejects all objects  $a$  such that  $a \lesssim -\mathbf{1}$ . Clearly,  $-\mathbf{1} < \mathbf{0}$ .



## 4.2 Construction of scales

Let us consider a given criterion  $i \in N$  and try to build a scale for  $i$  reflecting the preference of the DM. We choose to take as underlying scale for each criterion a bipolar univariate scale, or simply a unipolar scale if it happens that the criterion is clearly of unipolar nature. However, in the rest of the section, we consider that all criteria are bipolar since this is mathematically more general and can encompass as particular cases unipolar criteria as well. The choice between type I and type II is dictated by the fact that most of the time, a mixed feeling of good and bad affect for a given object on a given criterion comes from the presence of several subcriteria hidden in the criterion (see the chocolate/brocoli example in Section 2.1). For those cases where this decomposition into subcriteria is not possible or difficult, we refer the reader to Section 5 where the case of mixed affect on a single criterion is treated. Hence, we conclude that under our assumption, a scale for criterion  $i$  can be represented by a function  $u_i : X_i \rightarrow \mathbb{R}$ .

The problem is now to build this function  $u_i$  in a unique way. In order to achieve this, we make the assumption that the scales are interval scales, i.e. they are uniquely defined up to a positive affine transformation (two degrees of freedom). Since we made the assumption that on each scale  $u_i$  there exist two particular levels  $\mathbf{0}_i, \mathbf{1}_i$  having an absolute meaning, hence independent of  $i$ , we fix in a unique way all scales  $u_i$  by putting  $u_i(\mathbf{0}_i) = u_j(\mathbf{0}_j)$  and  $u_i(\mathbf{1}_i) = u_j(\mathbf{1}_j)$  for all  $i \neq j$ . Although this is arbitrary, it seems natural to put  $u_i(\mathbf{0}_i) = 0$  and  $u_i(\mathbf{1}_i) = 1$  for all  $i$ . Scales obtained in such a way are said to be *commensurate*.

Since this is not the central topic of this paper, we do not detail the practical construction of each scale  $u_i$ , which can be done using the MACBETH methodology [1, 2], and refer the reader to [18, 25, 17]. In short, the scale  $u_i$  is built from the preference of the decision maker over the set of fictitious alternatives  $(x_i, \mathbf{0}_{-i})$  when  $x_i$  ranges in  $X_i$ , that is,  $(x_i, \mathbf{0}_{-i}) \succsim (y_i, \mathbf{0}_{-i})$  iff  $u_i(x_i) \geq u_i(y_i)$ , and the scale is uniquely determined by putting  $u_i(\mathbf{0}_i) := 0$ , and  $u_i(\mathbf{1}_i) := 1$ .

## 4.3 Construction of the aggregation model

We try to construct an aggregation function  $F$  sophisticated enough to take into account interaction between criteria. For doing this, the usual way to put weights  $w_i$  on criteria, determined by the intensity of preference of the alternative  $(\mathbf{1}_i, \mathbf{0}_{-i})$ , is not sufficient, and we have to consider at least the set of alternatives where all possibilities of having some criteria satisfied and some other being at the neutral level [17, 18]. These are called *binary alternatives*  $(\mathbf{1}_A, \mathbf{0}_{-A})$ , for any  $A \subseteq N$ , since criteria take only two possible values.

### 4.3.1 Determination of the model for binary and ternary alternatives

As we deal with bipolar scales, it is of course important to take into account the influence of the unacceptable levels, but let us consider

for the moment only the above set of alternatives, which we call more precisely *positive binary alternatives*, since all criteria are at least at the neutral level (unipolar case). Let us define

$$\mu_+(A) := u(\mathbf{1}_A, \mathbf{0}_{-A}), \quad \forall A \subseteq N$$

that is,  $\mu_+$  represents the overall score of binary alternatives. Using as before the MACBETH approach, we may suppose that the scale  $\mu_+$  is an interval scale, which is uniquely determined as soon as one fixes two points. It seems natural to use the two extreme binary alternatives  $(\mathbf{1}_N, \mathbf{0}_\emptyset)$  and  $(\mathbf{1}_\emptyset, \mathbf{0}_N)$ , and to assign to them respectively the values 1 (best possible binary alternative) and 0 (worst possible binary alternative). On the other hand, considering  $A \subseteq B$  we have that  $(\mathbf{1}_A, \mathbf{0}_{-A})$  is dominated by  $(\mathbf{1}_B, \mathbf{0}_{-B})$  in the sense that the latter is at least as good as the former on each criterion. Then it is natural to have  $(\mathbf{1}_A, \mathbf{0}_{-A}) \preceq (\mathbf{1}_B, \mathbf{0}_{-B})$ , which gives  $\mu_+(A) \leq \mu_+(B)$ .

**Definition 1** [9, 38] *A function  $\mu : 2^N \rightarrow \mathbb{R}_+$  is a capacity or fuzzy measure if  $\mu(\emptyset) = 0$  and it satisfies  $\mu(A) \leq \mu(B)$  whenever  $A \subseteq B$  (monotonicity). The capacity is normalized if in addition  $\mu(N) = 1$ . If  $\mu$  is not monotone, then we call it a game.*

Hence, the function  $\mu_+$  induced by the set of binary alternatives is a normalized capacity.

Similarly, we consider *negative binary alternatives* of the form  $(-\mathbf{1}_A, \mathbf{0}_{-A})$ . Putting

$$\mu_-(A) := -u(-\mathbf{1}_A, \mathbf{0}_{-A}), \quad \forall A \subseteq N$$

we clearly define another capacity, which represents the overall score of negative binary alternatives.

We turn now to the general case, considering *ternary alternatives*, where we mix unacceptable, satisficing and neutral values. Defining  $\mathcal{Q}(N) := \{(A, B) \in 2^N \times 2^N \mid A \cap B = \emptyset\}$ , the set of ternary alternatives is

$$\{(\mathbf{1}_A, -\mathbf{1}_B, \mathbf{0}_{-(A \cup B)}) \mid (A, B) \in \mathcal{Q}(N)\}.$$

Several methods are possible to define the score of ternary alternatives. The simplest idea is to compute a net predisposition between the positive part and the negative part of the ternary alternatives, namely:

$$u(\mathbf{1}_A, -\mathbf{1}_B, \mathbf{0}_{-(A \cup B)}) := \mu^+(A) - \mu^-(B), \quad \forall (A, B) \in \mathcal{Q}(N).$$

This is the solution given by Cumulative Prospect Theory (CPT) [40]. Referring to our section on bipolar scales (Sec. 2),  $u$  lies on a bipolar univariate scale, and can be seen as the (linear) net predisposition between the overall score attributed to the positive part of the ternary alternative, and the overall score of the negative part.

A more general solution would be to define a function  $v$  on  $\mathcal{Q}(N)$  by:

$$v(A, B) := u(\mathbf{1}_A, -\mathbf{1}_B, \mathbf{0}_{-(A \cup B)}), \quad \forall (A, B) \in \mathcal{Q}(N).$$

Hence, there is no longer a separation between positive and negative parts, and consequently we lose the notion of net predisposition. Using as before the MACBETH approach, it suffices to fix

two points on the scale for its unique determination. We may fix for example  $u(\mathbf{1}_N, -\mathbf{1}_\emptyset, \mathbf{0}_\emptyset) = 1$  and  $u(\mathbf{1}_\emptyset, -\mathbf{1}_\emptyset, \mathbf{0}_N) = 0$ . This gives  $v(N, \emptyset) = 1$  and  $v(\emptyset, \emptyset) = 0$ . Let us consider now  $(A, B)$  and  $(C, D)$  in  $\mathcal{Q}(N)$ , and assume that  $A \subseteq C$  and  $B \supseteq D$ . We remark that  $(\mathbf{1}_A, -\mathbf{1}_B, \mathbf{0}_{-(A \cup B)})$  is dominated by  $(\mathbf{1}_C, -\mathbf{1}_D, \mathbf{0}_{-(C \cup D)})$  since on each criterion the latter is at least as good as the former. Thus it is natural to have  $(\mathbf{1}_A, -\mathbf{1}_B, \mathbf{0}_{-(A \cup B)}) \preceq (\mathbf{1}_C, -\mathbf{1}_D, \mathbf{0}_{-(C \cup D)})$ , which gives  $v(A, B) \leq v(C, D)$ .

**Definition 2** [15, 16] *A function  $v : \mathcal{Q}(N) \rightarrow \mathbb{R}$  such that  $v(\emptyset, \emptyset) = 0$ , and  $v(A, B) \leq v(C, D)$  whenever  $(A, B), (C, D) \in \mathcal{Q}(N)$  with  $A \subseteq C$  and  $B \supseteq D$  (monotonicity) is called a bi-capacity. Moreover, a bi-capacity is normalized if in addition  $v(N, \emptyset) = 1$  and  $v(\emptyset, N) = -1$ .*

Hence the function  $v$  defined as the overall score of ternary alternatives is a bi-capacity, which is not necessarily normalized since  $v(\emptyset, N)$  may be different from  $-1$ . In this case too,  $u$  is a bipolar univariate scale.

A third solution would be to consider for  $u$  a unipolar bivariate scale, that is,  $u$  is a pair of non negative numbers:

$$u(\mathbf{1}_A, -\mathbf{1}_B, \mathbf{0}_{-(A \cup B)}) := (\zeta^+(A, B), \zeta^-(A, B)), \quad \forall (A, B) \in \mathcal{Q}(N),$$

with  $\zeta^+(A, B), \zeta^-(A, B) \geq 0$ , the intensity of positive and negative affect felt by the decision maker when faced to a ternary alternative. Again,  $\zeta^+$  and  $\zeta^-$  have natural properties. If we consider that  $\zeta^+, \zeta^-$  reflect interval scales, then they can be uniquely determined by fixing two points. It is natural to put  $\zeta^+(N, \emptyset) = 1$  and  $\zeta^+(\emptyset, \emptyset) = 0$  for  $\zeta^+$ , and  $\zeta^-(\emptyset, N) = 1$  and  $\zeta^-(\emptyset, \emptyset) = 0$  for  $\zeta^-$ . Moreover, if as before we consider  $(A, B)$  and  $(C, D)$  in  $\mathcal{Q}(N)$  such that  $A \subseteq C$  and  $B \supseteq D$ , then because of dominance we should have  $\zeta^+(A, B) \leq \zeta^+(C, D)$  and  $\zeta^-(A, B) \geq \zeta^-(C, D)$ .

**Definition 3** [22] *A bipolar capacity is a function  $\zeta : \mathcal{Q}(N) \rightarrow [0, 1]^2$ ;  $(A, B) \mapsto (\zeta^+(A, B), \zeta^-(A, B))$  such that  $\zeta^+(N, \emptyset) = \zeta^-(\emptyset, N) = 1$ ,  $\zeta^+(\emptyset, \emptyset) = \zeta^-(\emptyset, \emptyset) = 0$ , and if  $(A, B), (C, D) \in \mathcal{Q}(N)$  with  $A \subseteq C$  and  $B \supseteq D$ , then  $\zeta^+(A, B) \leq \zeta^+(C, D)$  and  $\zeta^-(A, B) \geq \zeta^-(C, D)$  (monotonicity).*

Hence we have used a bipolar capacity to represent the overall score on symmetric scales.

### 4.3.2 Determination of the model for general alternatives

**A general methodology** So far, we have determined the way to compute the overall score of ternary alternatives, not of general alternatives. A simple way to do this is to consider that the overall score of a given alternative will be obtained as an interpolation between scores of neighbouring ternary alternatives. We explain the procedure first for (positive) binary alternatives. Since  $u_i(\mathbf{1}_i) = 1$  and  $u_i(\mathbf{0}_i) = 0$  for all  $i \in N$ , vectors  $(u_1(x_1), \dots, u_n(x_n))$  describe the set of vertices

of the hypercube  $[0, 1]^n$  when only binary alternatives are considered. This means that the aggregation function  $F$  has a fixed and known value at all vertices, and we have to interpolate  $F$  inside the hypercube.

As many types of interpolation exist, we are looking here for a linear interpolation using as few points as possible, which are here vertices of  $[0, 1]^n$ . For a given  $x \in [0, 1]^n$ , let us denote by  $\mathcal{V}(x)$  the set of vertices used for the linear interpolation, which writes

$$F(x) = \sum_{A \subseteq N | (1_A, 0_{-A}) \in \mathcal{V}(x)} \left[ \alpha_0(A) + \sum_{i=1}^n \alpha_i(A)x_i \right] F(1_A, 0_{-A}), \quad (25)$$

where  $\alpha_i(A) \in \mathbb{R}$ ,  $i = 0, \dots, n$ ,  $\forall A \in \mathcal{V}(x)$ . To keep the meaning of interpolation, we force that the convex hull  $\text{conv}(\mathcal{V}(x))$  contains  $x$ , and any  $x \in [0, 1]^n$  should belong to a unique polyhedron  $\text{conv}(\mathcal{V}(x))$  (except for common facets), and continuity should be ensured. Hence, the hypercube is partitioned into  $q$  polyhedra defined by their sets of vertices  $\mathcal{V}_1, \dots, \mathcal{V}_q$ , all vertices being vertices of  $[0, 1]^n$ . Such an operation is called a *triangulation*. Note that the least possible number of vertices is  $n + 1$ , otherwise the polyhedra would not be  $n$ -dimensional, and hence a finite number of them would not cover the whole hypercube.

Many different triangulations are possible, but there is one which is of particular interest, since it leads to an interpolation where all constant terms  $\alpha_0(A)$  are null. This triangulation uses the  $n!$  *canonical polyhedra* of  $[0, 1]^n$ :

$$\text{conv}(\mathcal{V}_\sigma) = \{x \in [0, 1]^n \mid x_{\sigma(1)} \leq \dots \leq x_{\sigma(n)}\}, \text{ for some permutation } \sigma \text{ on } N.$$

**Proposition 1** *The linear interpolation (25) using the canonical polyhedra writes*

$$F(x) = \sum_{i=1}^n [x_{\sigma(i)} - x_{\sigma(i-1)}] \mu(\{x_{\sigma(i)}, \dots, x_{\sigma(n)}\}), \quad (26)$$

with  $\mu(A) := F(1_A, 0_{-A})$ . Moreover,  $F$  is continuous on  $[0, 1]^n$ .

The aggregation function  $F(x)$  defined in (26) is in fact the Choquet integral of  $x$  w.r.t. the capacity  $\mu$ , which we denote by  $\mathcal{C}_\mu(x)$ , considering  $x$  as a function over  $N$ . This way of introducing the Choquet integral as the simplest linear interpolation is given in [14]. It was also discovered by Lovász [26], considering the problem of extending the domain of pseudo-Boolean functions to  $\mathbb{R}^n$  (for this extension problem, see also Singer [35]). The fact that the so-called *Lovász extension* was the Choquet integral was remarked by Marichal [27].

It remains to apply this result to our different models established for ternary alternatives.

**The CPT model** The simplest model for ternary alternatives is to use two capacities  $\mu^+$ ,  $\mu^-$ . Applying the above methodology, we interpolate  $F$  between vertices of  $[0, 1]^n$  and  $[-1, 0]^n$ , corresponding respectively to positive and negative binary alternatives, and we are

led to what is called the *Cumulative Prospect Theory* model (CPT), proposed by Tversky and Kahnemann [40]:

$$\text{CPT}_{\mu^+, \mu^-}(x) := \mathcal{C}_{\mu^+}(x^+) - \mathcal{C}_{\mu^-}(x^-)$$

where  $x^+ := x \vee 0$  and  $x^- := (-x)^+$ .

**The Choquet integral for bi-capacities** In this case, we consider the hypercube  $[-1, 1]^n$ , and all points of the form  $(1_A, -1_B, 0_{-(A \cup B)})$ , which correspond to ternary alternatives. Let us apply again an interpolation approach, and call  $F$  the function we obtain by interpolation. To do this, we examine in details the case  $n = 2$  (Fig. 10). Let us

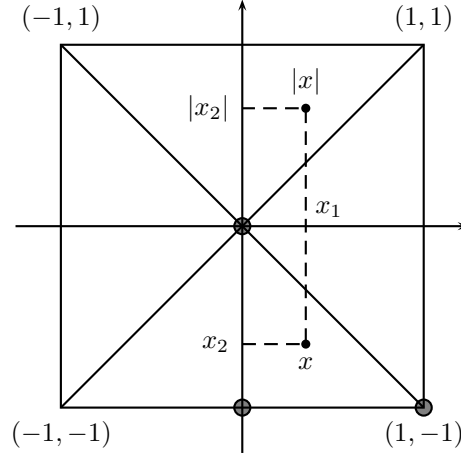


Figure 10: Interpolation for the case of bi-capacities

take any point  $x$  such that  $x_1 \geq 0$ ,  $x_2 \leq 0$  and  $|x_1| \leq |x_2|$ . Then, for  $|x|$  which is in the first (positive) quadrant, we already know that the best linear interpolation is the Choquet integral. It suffices to use the formula with the adequate vertices:

$$F(x_1, x_2) := |x_1|F(1, -1) + (|x_2| - |x_1|)F(0, -1)$$

This is a Choquet integral w.r.t a game  $\nu_1$  defined by:

$$\begin{aligned} \nu_1(\{1, 2\}) &= F(1, -1) \\ \nu_1(\{2\}) &= F(0, -1). \end{aligned}$$

Let us consider now the general case. Defining  $N_x^+ := \{i \in N \mid x_i \geq 0\}$ ,  $N_x^- = N \setminus N_x^+$ , with similar considerations of symmetry, we obtain:

$$\begin{aligned} F(x) &= |x_{\sigma(1)}|F(1_{N_x^+}, -1_{N_x^-}, 0_{-(N_x^+ \cup N_x^-)}) \\ &+ \sum_{i=2}^n (|x_{\sigma(i)}| - |x_{\sigma(i-1)}|)F(1_{\{\sigma(i), \dots, \sigma(n)\} \cap N_x^+}, -1_{\{\sigma(i), \dots, \sigma(n)\} \cap N_x^-}, 0_{-\{\sigma(i), \dots, \sigma(n)\}}) \end{aligned}$$

where  $\sigma$  is a permutation on  $N$  such that  $|x_{\sigma(1)}| \leq \dots \leq |x_{\sigma(n)}|$ . This expression is the Choquet integral of  $|x|$  w.r.t. a game  $\nu_{N_x^+}$  defined by:

$$\nu_{N_x^+}(A) := F(1_{A \cap N_x^+}, -1_{A \cap N_x^-}, 0_{-A}).$$

Recalling that  $F(1_A, -1_B, 0_{-(A \cup B)}) =: v(A, B)$ , we finally come up with the following definition.

**Definition 4** *Let  $v$  be a bi-capacity and  $x$  be a real-valued function on  $N$ . The (general) Choquet integral of  $x$  w.r.t  $v$  is given by*

$$\mathcal{C}_v(x) := \mathcal{C}_{\nu_{N_x^+}}(|f|)$$

where  $\nu_{N_x^+}$  is a game on  $N$  defined by

$$\nu_{N_x^+}(C) := v(C \cap N_x^+, C \cap N_x^-),$$

and  $N_x^+ := \{i \in N \mid x_i \geq 0\}$ ,  $N_x^- = N \setminus N_x^+$ .

**The Choquet integral for bipolar capacities** In this case the Choquet integral w.r.t. a bipolar capacity  $\zeta$  is calculated as follows [22]:

$$\mathcal{C}_\zeta(x) := \mathcal{C}_{\nu_{N^+}^+}(|x|) - \mathcal{C}_{\nu_{N^+}^-}(|x|), \quad (27)$$

with  $N^+$  and  $N^-$  as above, and  $\nu_{N^+}^+(A) := \zeta^+(A \cap N^+, A \cap N^-)$  and  $\nu_{N^+}^-(A) := \zeta^-(A \cap N^+, A \cap N^-)$ . Of course, if  $v(A, B) = \zeta^+(A, B) - \zeta^-(A, B)$  we get

$$\mathcal{C}_v(x) = \mathcal{C}_\zeta(x).$$

## 5 The bivariate setting

In this section we introduce a set of models describing multiple criteria evaluations in terms of possible negative and positive evaluations with respect to the same criterion. In psychological terms this model represents the positive, the negative and the net affect resulting from a multiplicity of stimulus, each of them characterized by a negative and a positive affect. The proposed models can be considered as a generalization to a multiplicity of dimensions (stimuli) of the bivariate evaluative space model [8]. To be more precise, while the bivariate evaluative space model analyzes the activation of positive and negative evaluations with respect to a single criterion (stimulus) (for example the simultaneous presence of sadness and happiness, hot and cold feelings and so on), we consider the cases in which from a multiplicity of criteria (stimuli) comprehensive evaluations are drawn. Both at the level of single criterion and at the level of comprehensive evaluation, positive and negative evaluations are considered independent and therefore they are represented in terms of a bivariate scale, rather than in terms of a bipolar scale. This means that positive and negative evaluations with respect to a single criterion (stimulus) and with respect to the comprehensive evaluations are present simultaneously.

The proposed model admits as specific particular cases the decomposable model and the the CPT model presented in section 2 and the Choquet integral model for bi-capacities presented in section 4. These are bipolar models at the level of a single criterion, i.e. they rely on the assumption that each criterion has only a positive or a negative evaluation and cannot simultaneously receive a positive and a negative evaluation. As explained in section 3, in one of these models, the CPT model, the simultaneous presence of a positive and a negative evaluation is considered only at the level of the comprehensive evaluation. However, the comprehensive positive evaluation and the comprehensive negative evaluation are aggregated in the simplest way, i.e. obtaining the comprehensive evaluation as a simple algebraic difference between the positive and the negative comprehensive evaluations, neither considering any form of different weighting of the positive and the negative evaluations nor any form of nonlinearity. This is very intuitive, but perhaps too simple and not general enough. Let us remark that within the bivariate evaluative space model [8], even if only at the level of a single criterion-stimulus, both the effect of the different weighting and of the non-linearity of positive and negative evaluations are investigated and quantitatively measured. More precisely, [8] proposes the following representation of the net result of the aggregation of the positive evaluation  $x_i^+$  and negative evaluation  $x_i^-$  with respect to criterion-stimulus  $i$ :

$$u_i(x_i^+, x_i^-) = w_i^+ v_i^+(x_i^+) - w_i^- v_i^-(x_i^-) + v_i^{+-}(x_i^+, x_i^-) + c_i \quad (28)$$

where  $w_i^+$  is the weight of the positive evaluation,  $w_i^-$  is the weight of the negative evaluation,  $v_i^+(x_i^+)$  is the activation function for positivity,  $v_i^-(x_i^-)$  is the activation function for negativity,  $v_i^{+-}(x_i^+, x_i^-)$

represents the non-additive effects and  $c_i$  is a constant. Starting from a general model coherent with model (28), we firstly propose a general multiple criteria model representing bivariate evaluations at the level of single criteria and at the level of the comprehensive evaluation. Then we recall a generalization of bipolar Choquet integral [20] which permits to take into account the bivariate nature of the evaluations both at the level of single criteria and of comprehensive evaluation. Finally, we briefly discuss the introduction of bivariate scales in models based on difference of preferences.

## 5.1 A general bivariate multicriteria model

As anticipated in section 3.3.2, in a bivariate setting each alternative is represented by  $n$ -pairs of components vectors  $x = ((x_1^+, x_1^-), \dots, (x_n^+, x_n^-))$  or equivalently as a  $2n$ -component vector  $x = (x_1^+, \dots, x_n^+, x_1^-, \dots, x_n^-)$  which results from putting together the two  $n$ -vectors  $x^+ = (x_1^+, \dots, x_n^+)$  and  $x^- = (x_1^-, \dots, x_n^-)$ , such that  $x = (x^+, x^-)$ . In very general terms, in this bivariate setting the comprehensive positive evaluation  $CPE(x)$ , the negative comprehensive evaluation  $CNE(x)$  and the comprehensive evaluation  $CE(x)$  can be represented as follows:

$$CPE(x) = F^+(x_1^+, \dots, x_n^+, x_1^-, \dots, x_n^-)$$

$$CNE(x) = F^-(x_1^+, \dots, x_n^+, x_1^-, \dots, x_n^-)$$

$$CE(x) = G(F^+(x_1^+, \dots, x_n^+, x_1^-, \dots, x_n^-), F^-(x_1^+, \dots, x_n^+, x_1^-, \dots, x_n^-))$$

where  $F^+ : \mathbb{R}_+^{2n} \rightarrow \mathbb{R}_+$  and  $F^- : \mathbb{R}_+^{2n} \rightarrow \mathbb{R}_+$  are functions non-decreasing with respect to their first  $n$  arguments and non-increasing with respect to their last  $n$  arguments, while  $G : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  is a function non-decreasing in its first argument and non-increasing in its second argument.

Let us observe that in this very general model, the comprehensive positive evaluation  $CPE(x)$  depends not only on the positive evaluations  $(x_i^+, i \in N)$ , but also on the negative evaluations  $(x_i^-, i \in N)$ . Coming back to the example of car evaluation outlined in the introduction, this means that the comprehensive positive evaluation depends on the positive aspects of the cars  $(x_i^+, i \in N)$  such as a good maximum speed and a good acceleration, but it depends also on the negative aspects  $(x_i^-, i \in N)$  such as a high price and an high fuel consumption. Analogously, the comprehensive negative evaluation depends on the negative aspects of the car  $(x_i^-, i \in N)$  but it depends also on the positive aspects  $(x_i^+, i \in N)$ . A special case of this very general model is the following one, where the comprehensive positive evaluation depends only on the positive evaluations and the comprehensive negative evaluation depends only on the negative evaluations:

$$CPE(x) = F^+(x_1^+, \dots, x_n^+)$$

$$CNE(x) = F^-(x_1^-, \dots, x_n^-)$$

$$CE(x) = G(F^+(x_1^+, \dots, x_n^+), F^-(x_1^-, \dots, x_n^-))$$



## 5.2 The Choquet integral in a bivariate setting

For the sake of simplicity in this subsection each alternative  $x$  is represented by a  $2n$ -vector  $(x_1^+, \dots, x_n^+, x_1^-, \dots, x_n^-) \in [0, 1]^{2n}$ , such that

- $x_i^+ = 0$  means that the positive evaluation with respect to criterion  $i$  is null,
- $x_i^+ = 1$  means that the positive evaluation with respect to criterion  $i$  is maximal,
- $x_i^- = 0$  means that the negative evaluation with respect to criterion  $i$  is null,
- $x_i^- = 1$  means that the negative evaluation with respect to criterion  $i$  is maximal.

In this bivariate setting, as an interesting specific case, one can consider a very simple model formulated as

$$CPE(x) = \sum_{i=1}^n w_i^{++} x_i^+ - \sum_{i=1}^n w_i^{-+} x_i^-,$$

$$CNE(x) = \sum_{i=1}^n w_i^{--} x_i^- - \sum_{i=1}^n w_i^{+-} x_i^+,$$

$$CE(x) = CPE(x) - CNE(x) = \sum_{i=1}^n [w_i^{++} + w_i^{+-}] x_i^+ - \sum_{i=1}^n [w_i^{-+} + w_i^{--}] x_i^-,$$

where  $w_i^{++}, w_i^{-+}, w_i^{+-}, w_i^{--} \in \mathbb{R}_+$  are non negative weights. In the above representation, the possibility of some interactions between different criteria is not considered at all. A possible specific model representing these types of interaction can be formulated as follows:

$$\begin{aligned} CPE(x) &= \sum_{i=1}^n w_i^{++} x_i^+ - \sum_{i=1}^n w_i^{-+} x_i^- + \\ &+ \sum_{i,j \in N, i < j} v_{i,j}^{++,+}(x_i^+, x_j^+) + \sum_{i,j \in N} v_{i,j}^{+-,+}(x_i^+, x_j^-) + \sum_{i,j \in N, i < j} v_{i,j}^{-+,-}(x_i^-, x_j^-) \\ CNE(x) &= \sum_{i=1}^n w_i^{--} x_i^- - \sum_{i=1}^n w_i^{+-} x_i^+ + \\ &+ \sum_{i,j \in N, i < j} v_{i,j}^{++,-}(x_i^+, x_j^+) + \sum_{i,j \in N} v_{i,j}^{+-,-}(x_i^+, x_j^-) + \sum_{i,j \in N, i < j} v_{i,j}^{-+,-}(x_i^-, x_j^-) \\ CE(x) &= CPE(x) - CNE(x) \end{aligned}$$

where

- $v_{i,j}^{++,+}(x_i^+, x_j^+) [v_{i,j}^{++,-}(x_i^+, x_j^+)]$  measures the interaction of  $x_i^+$  and  $x_j^+$  with respect to the comprehensive positive [resp. negative] evaluation  $CPE(x)$  [resp.  $CNE(x)$ ],

- $v_{i,j}^{+-,+}(x_i^+, x_j^-)$  [ $v_{i,j}^{+-,-}(x_i^+, x_j^-)$ ] measures the interaction of  $x_i^+$  and  $x_j^-$  with respect to the comprehensive positive [resp. negative] evaluation  $CPE(x)$  [resp.  $CNE(x)$ ],
- $v_{i,j}^{-,-,+}(x_i^-, x_j^-)$  [ $v_{i,j}^{-,-,-}(x_i^-, x_j^-)$ ] measures the interaction of  $x_i^-$  and  $x_j^-$  with respect to the comprehensive positive [resp. negative] evaluation  $CPE(x)$  [resp.  $CNE(x)$ ].

Let us observe that the proposed model can be generalized, considering not only interactions between pairs of signed evaluations of the type  $(x_i^+, x_j^+)$ ,  $(x_i^+, x_j^-)$  or  $(x_i^-, x_j^-)$ , but in general pairs of sets of signed evaluations of the type  $(x_i^+, i \in A, x_j^-, j \in B)$ , with  $A, B \subseteq N$  such that  $A \cup B \neq \emptyset$ . Taking into account this aspect we get the following model:

$$\begin{aligned}
 CPE(x) &= \sum_{i=1}^n w_i^{++} x_i^+ - \sum_{i=1}^n w_i^{-+} x_i^- + \\
 &+ \sum_{A, B \subseteq N, |A \cup B| \geq 2} v_{A, B}^+(x_i^+, i \in A, x_j^-, j \in B), \\
 CNE(x) &= \sum_{i=1}^n w_i^{--} x_i^- - \sum_{i=1}^n w_i^{+-} x_i^+ + \\
 &+ \sum_{A, B \subseteq N, |A \cup B| \geq 2} v_{A, B}^-(x_i^+, i \in A, x_j^-, j \in B), \\
 CE(x) &= CPE(x) - CNE(x),
 \end{aligned}$$

where  $v_{A, B}^+(x_i^+, i \in A, x_j^-, j \in B)$  [resp.  $v_{A, B}^-(x_i^+, i \in A, x_j^-, j \in B)$ ] measures the interaction of  $x_i^+, i \in A$  and  $x_j^-, j \in B$ ,  $A, B \subseteq N$  with respect to the comprehensive positive [resp. negative] evaluation  $CPE(x)$  [resp.  $CNE(x)$ ]. The following properties can be required from the interaction functions  $v_{A, B}^+$  and  $v_{A, B}^-$ :

- monotonicity:  $v_{A, B}^+$  and  $v_{A, B}^-$  are functions non-decreasing in all their arguments;
- lower boundary condition:  $v_{A, B}^+(x_i^+, i \in A, x_j^-, j \in B) = 0$  and  $v_{A, B}^-(x_i^+, i \in A, x_j^-, j \in B) = 0$  if there exists  $i \in A$  such that  $x_i^+ = 0$  or if there exists  $j \in B$  such that  $x_j^- = 0$ ;
- upper boundary condition:  $v_{A, B}^+(x_i^+, i \in A, x_j^-, j \in B) = 1$  and  $v_{A, B}^-(x_i^+, i \in A, x_j^-, j \in B) = 1$  if, for all  $i \in A$ ,  $x_i^+ = 1$  and, for all  $j \in B$ ,  $x_j^- = 1$ .

The lower boundary condition says that in order to have an interaction effect with respect to positive evaluations on criteria from  $A$  and negative evaluations on criteria from  $B$ , none of these evaluations must be null; in other words, if even only one among positive evaluations on criteria from  $A$  or negative evaluations on criteria from  $B$  is null, then there is no interaction effect. On the contrary, the upper boundary condition says that if each one of the positive evaluations on criteria from  $A$  and the negative evaluation on criteria from  $B$  is maximum,

then the interaction also reaches its maximum value. Two simple formulations of functions  $v_{A,B}^+$  and  $v_{A,B}^-$  are the following:

- interaction based on product:

$$v_{A,B}^+(x_i^+, i \in A, x_j^-, j \in B) = w_{A,B}^+ \prod_{i \in A} x_i^+ \prod_{j \in B} x_j^-$$

and

$$v_{A,B}^-(x_i^+, i \in A, x_j^-, j \in B) = w_{A,B}^- \prod_{i \in A} x_i^+ \prod_{j \in B} x_j^-,$$

- interaction based on minimum:

$$v_{A,B}^+(x_i^+, i \in A, x_j^-, j \in B) = w_{A,B}^+ \min[\min_{i \in A} x_i^+, \min_{j \in B} x_j^-]$$

and

$$v_{A,B}^-(x_i^+, i \in A, x_j^-, j \in B) = w_{A,B}^- \min[\min_{i \in A} x_i^+, \min_{j \in B} x_j^-],$$

where  $w_{A,B}^+, w_{A,B}^- \in \mathbb{R}$  are weights related to the sign of the interaction ( $w_{A,B}^+ > 0, w_{A,B}^- > 0$ : positive interaction;  $w_{A,B}^+ < 0, w_{A,B}^- < 0$ : negative interaction) and to the strength of the effect of interaction with respect to the comprehensive evaluation.

An interesting property one could expect from a multicriteria evaluation model is the following.

**Definition 5** A bivariate multicriteria evaluation model  $\langle CPE, CNE, CE \rangle$ , is Stable for the same Positive Linear transformations (**SPL**) if

$$\begin{aligned} & CPE((rx_1^+ + s, \dots, rx_n^+ + s), (rx_1^- + s, \dots, rx_n^- + s)) \\ &= rCPE((x_1^+, \dots, x_n^+), (x_1^-, \dots, x_n^-)) + s \\ & CNE((rx_1^+ + s, \dots, rx_n^+ + s), (rx_1^- + s, \dots, rx_n^- + s)) \\ &= rCNE((x_1^+, \dots, x_n^+), (x_1^-, \dots, x_n^-)) + s \\ & CE((rx_1^+ + s, \dots, rx_n^+ + s), (rx_1^- + s, \dots, rx_n^- + s)) \\ &= rCE((x_1^+, \dots, x_n^+), (x_1^-, \dots, x_n^-)) + s \end{aligned}$$

for all  $x \in \mathbb{R}^{2n}$ ,  $r > 0$ ,  $s \in \mathbb{R}$ .

Let us observe that the bivariate model with interactions based on minimum satisfies the SPL property, while the interaction based on multiplication does not. Thus, it is interesting to note that the bivariate model with interactions based on minimum corresponds to a specific generalization of the Choquet integral [20] which is based on the concept of generalized bipolar capacity.

Let  $\mathcal{Q}^*(N)$  denote a set of pairs of subsets of  $N$ , defined as follows,

$$\mathcal{Q}^*(N) = \{ (C, D) : C \subseteq N, D \subseteq N \}.$$

We define the generalized bipolar capacity  $\zeta^*$  on  $N$  as any function of the form,

$$\zeta^* : \mathcal{Q}^*(N) \rightarrow [0, 1] \times [0, 1]; (A, B) \mapsto (\zeta^{*+}(A, B), \zeta^{*-}(A, B))$$

such that:

1.  $\zeta^*(C, \emptyset) = (c, 0)$  and  $\zeta^*(\emptyset, D) = (0, d)$ , with  $c, d \in [0, 1]$ ;
2.  $\zeta^*(N, \emptyset) = (1, 0)$  and  $\zeta^*(\emptyset, N) = (0, 1)$ ;
3. for each  $(C, D), (E, F) \in \mathcal{Q}^*(N)$ , such that  $C \supseteq E$  and  $D \subseteq F$ , we have  $\zeta^*(C, D) = (c, d)$  and  $\zeta^*(E, F) = (e, f)$ ,  $c, d, e, f \in [0, 1]$ , with  $c \geq e$  and  $d \leq f$ .

The generalized bipolar capacity,  $\zeta^*$  is related to multicriteria bivariate evaluations more or less as the bi-capacity is related to bipolar evaluations. Let us consider the set of alternatives on which the positive as well as the negative criterion evaluations can only be maximal or null. These alternatives are of the form  $x^+ = (\mathbf{1}_A, \mathbf{0}_{A^c})$ ,  $x^- = (\mathbf{1}_B, \mathbf{0}_{B^c})$ ,  $(A, B) \in \mathcal{Q}^*(N)$  ( $A^c$  and  $B^c$  respectively denote the complement of  $A$  and  $B$  in  $N$ ). Thus alternative  $x = (x^+, x^-) = ((\mathbf{1}_A, \mathbf{0}_{A^c}), (\mathbf{1}_B, \mathbf{0}_{B^c}))$  has the maximal positive evaluation on criteria from  $A$ , the maximal negative evaluation on criteria from  $B$ , while positive evaluations on criteria not from  $A$  and negative evaluations on criteria not from  $B$  are null. These alternatives are called *bivariate binary alternatives*. Now let us consider the function  $\phi : \mathcal{Q}^*(N) \rightarrow [0, 1]^2$  such that  $\phi = (\phi^+, \phi^-)$  where for each  $(A, B) \in \mathcal{Q}^*(N)$  we have:

$$\phi^+(A, B) = CPE[(\mathbf{1}_A, \mathbf{0}_{A^c}), (\mathbf{1}_B, \mathbf{0}_{B^c})],$$

$$\phi^-(A, B) = CNE[(\mathbf{1}_A, \mathbf{0}_{A^c}), (\mathbf{1}_B, \mathbf{0}_{B^c})].$$

It is natural to suppose that

- $\phi(\emptyset, N) = (0, b)$ , with  $b \in [0, 1]$ ,
- $\phi(A, \emptyset) = (a, 0)$ , with  $a \in [0, 1]$ ,
- $\phi(N, B) = (1, b)$ , with  $b \in [0, 1]$ ,
- $\phi(A, N) = (a, 1)$ , with  $a \in [0, 1]$ .

Let us also remark that for  $(A, B), (C, D) \in \mathcal{Q}^*(N)$  such that  $A \supseteq C$  and  $B \subseteq D$  and  $x = ((\mathbf{1}_A, \mathbf{0}_{A^c}), (\mathbf{1}_B, \mathbf{0}_{B^c}))$  and  $y = ((\mathbf{1}_C, \mathbf{0}_{C^c}), (\mathbf{1}_D, \mathbf{0}_{D^c}))$ , we have that  $x$  is at least as good as  $y$  with respect to both positive evaluations and negative evaluations on each criterion. Then it is natural to have that  $x$  is at least as good as  $y$  with respect to the comprehensive positive evaluation ( $CPE(x) \geq CPE(y)$ ) and the comprehensive negative evaluation ( $CNE(x) \leq CNE(y)$ ), so that:

$$\phi^+(A, B) \geq \phi^+(C, D) \text{ and } \phi^-(A, B) \leq \phi^-(C, D).$$

Hence the function  $\phi$  is a generalized bipolar capacity.

It is interesting to observe the relation between the bipolar capacity  $\phi$  and the weights  $w_i^{++}$ ,  $w_i^{+-}$ ,  $w_i^{-+}$ ,  $w_i^{--}$ ,  $w_{A,B}^+$  and  $w_{A,B}^-$ :

$$\phi^+(A, B) = \sum_{i \in A} w_i^{++} + \sum_{i \in B} w_i^{-+} + \sum_{C, D \subseteq N, A \cup B \neq \emptyset, C \subseteq A, D \subseteq B} w_{A,B}^+$$

$$\phi^-(A, B) = \sum_{i \in A} w_i^{+-} + \sum_{i \in B} w_i^{--} + \sum_{C, D \subseteq N, A \cup B \neq \emptyset, C \subseteq A, D \subseteq B} w_{A,B}^-$$

As with the CPT model and the Choquet integral for bi-capacities it is possible to pass from scores of ternary alternatives to scores on general alternatives as showed in section 4, also in this bivariate setting it is possible to pass from evaluations on bivariate binary alternatives to general alternatives by means of a proper interpolation between scores of neighboring bivariate binary alternatives. In this case, we consider the hypercube  $[0, 1]^{2n}$  and all points of the form  $((\mathbf{1}_A, \mathbf{0}_{A^c}), (\mathbf{1}_B, \mathbf{0}_{B^c}))$ . In this way, we obtain the bipolar Choquet bi-integral [20] which can be analytically defined as follows. For each  $(x^+, x^-) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ ,  $x^+ = (x_1^+, \dots, x_n^+) \in \mathbb{R}_+^n$  and  $x^- = (x_1^-, \dots, x_n^-) \in \mathbb{R}_+^n$ , we call  $x^+$  the positive part of  $(x^+, x^-)$  while  $x^-$  is its negative part.

For each  $(x^+, x^-) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ , let us consider the following one-to-one correspondence,

$$\{1, \dots, 2n\} \rightarrow \mathcal{N}^* = \{1^+, \dots, n^+, 1^-, \dots, n^-\}$$

such that:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(j)} \leq \dots \leq x_{(n)} \leq \dots \leq x_{(2n)},$$

where

$$x_{(j)} = \begin{cases} x_i^+ & \text{if } (j) = i^+ \\ x_i^- & \text{if } (j) = i^- \end{cases}$$

Intuitively the values  $x_{(j)}$  represent the positive and the negative evaluations reordered according to their intensity.

For each  $j \in N$  let us consider also the following two subsets of  $N$ :

1.  $C_{(j)} = \{i \in N : x_i^+ \geq x_{(j)}\}$ ;
2.  $D_{(j)} = \{i \in N : x_i^- \geq x_{(j)}\}$ .

Given a generalized bipolar capacity,  $\zeta^*$  on  $N$ , and  $(x^+, x^-) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ , we can define the positive part of the bipolar Choquet bi-integral in the following way:

$$\begin{aligned} \mathcal{C}^{B^{**}}((x^+, x^-), \zeta^*) &= \sum_{j=1}^{2n} (x_{(j)} - x_{(j-1)}) \zeta^{**}(C_{(j)}, D_{(j)}) = \\ &= \sum_{j=1}^{2n} x_{(j)} (\zeta^{**}(C_{(j)}, D_{(j)}) - \zeta^{**}(C_{(j+1)}, D_{(j+1)})) \end{aligned}$$

where  $x_{(0)} = 0$  and  $\zeta^{**}(C_{(n+1)}, D_{(n+1)}) = 0$ .

Analogously, the negative part of the bipolar Choquet bi-integral can be defined as follows:

$$\begin{aligned}\mathcal{C}^{B*-}((x^+, x^-), \zeta^*) &= \sum_{j=1}^{2n} (x_{(j)} - x_{(j-1)}) \zeta^{*-}(C_{(j)}, D_{(j)}) = \\ &= \sum_{j=1}^{2n} x_{(j)} \left( \zeta^{*-}(C_{(j)}, D_{(j)}) - \zeta^{*-}(C_{(j+1)}, D_{(j+1)}) \right)\end{aligned}$$

where  $\zeta^{*-}(C_{(n+1)}, D_{(n+1)}) = 0$ .

Finally, the bipolar Choquet bi-integral of  $(x^+, x^-) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ , with respect to the generalized capacity  $\zeta^*$  is defined as follows:

$$\mathcal{C}^{B*}((x^+, x^-), \zeta^*) = \mathcal{C}^{B*+}((x^+, x^-), \zeta^*) - \mathcal{C}^{B*-}((x^+, x^-), \zeta^*).$$

Summarizing, we can say that  $\mathcal{C}^{B*+}((x^+, x^-), \zeta^*)$  represents the comprehensive positive evaluation,  $\mathcal{C}^{B*-}((x^+, x^-), \zeta^*)$  represents the comprehensive negative evaluation and finally  $\mathcal{C}^{B*}((x^+, x^-), \zeta^*)$  represents the final net comprehensive evaluation, i.e.  $CPE(x) = \mathcal{C}^{B*+}(x)$ ,  $CNE(x) = \mathcal{C}^{B*-}(x)$  and  $CE(x) = \mathcal{C}^{B*}(x)$ .

Now to conclude our discussion about coincidence between  $\mathcal{C}^{B*+}(x)$ ,  $\mathcal{C}^{B*-}(x)$  and  $\mathcal{C}^{B*}(x)$  and the above general bivariate model with interactions represented by minimum, after having already showed how to pass from weights  $w_{A,B}^+$ ,  $w_{A,B}^-$  to a generalized bipolar capacity  $\zeta^*$  considering bivariate binary alternatives, we discuss how to pass from generalized bipolar capacity  $\zeta^*$  to weights  $w_i^{++}$ ,  $w_i^{-+}$ ,  $w_i^{+-}$  and  $w_i^{--}$ ,  $i = 1, \dots, n$ , and  $w_{A,B}^+$  and  $w_{A,B}^-$ ,  $A, B \subseteq N$ . With this aim, putting  $w_i^{++} = w_{\{i\}, \emptyset}^+$ ,  $w_i^{-+} = w_{\emptyset, \{i\}}^+$ ,  $w_i^{+-} = w_{\{i\}, \emptyset}^-$  and  $w_i^{--} = w_{\emptyset, \{i\}}^-$ ,  $i = 1, \dots, n$ , we have to consider the following generalized Möbius transform [20]:

$$\begin{aligned}w_{A,B}^+ &= \sum_{\{(C,D) \in \mathcal{Q}^*(N: C \subseteq A \text{ and } D \subseteq B)\}} (-1)^{(|(A \cup B)| - |(C \cup D)|)} \zeta^+(C, D) \\ w_{A,B}^- &= \sum_{\{(C,D) \in \mathcal{Q}^*(N: C \subseteq A \text{ and } D \subseteq B)\}} (-1)^{(|(A \cup B)| - |(C \cup D)|)} \zeta^-(C, D)\end{aligned}$$

A characterization of the generalized bipolar Choquet bi-integral has been presented in [20].

Let us now investigate the relation between the bipolar Choquet bi-integral and some interesting specific multiple criteria models which can be derived from it as particular cases. A specific case of the bipolar Choquet bi-integral considers bivariate scales with respect to criteria, but gives only a bipolar comprehensive final net evaluation without distinguishing between the positive comprehensive evaluation and the negative comprehensive evaluation, that is, in this model, for each alternative  $x \in \mathbb{R}_+^{2n}$ , the model gives  $CE(x)$  but it does not give  $CPE(x)$  and  $CNE(x)$ .

In this model we consider a generalized bi-capacity  $v^*$  on  $N$ , being a function

$$v^* : \mathcal{Q}^*(N) \rightarrow [-1, 1]$$

such that:

1.  $v^*(\emptyset, \emptyset) = 0$ ;
2.  $v^*(N, \emptyset) = 1$ , and  $v^*(\emptyset, N) = -1$ ;
3. If  $C \supseteq E$  and  $D \subseteq F$ , then  $v^*(C, D) \geq v^*(E, F)$ .

Properties 1 and 2 are the boundary conditions, while property 3 is the monotonicity condition.

Let us observe that using the generalized bipolar capacity  $\zeta^*$  on  $N$ , we can obtain a corresponding generalized bi-capacity  $v^*$  on  $N$  as follows:

$$v^*(C, D) = \zeta^{*+}(C, D) - \zeta^{*-}(C, D), \quad \forall (C, D) \in \mathcal{Q}^*(N).$$

Therefore, between the generalized bipolar capacity and the generalized bi-capacity, there is the same difference as between the bipolar capacity and the bi-capacity, that is the generalized bipolar capacity permits to distinguish between the negative and the positive predisposition, while the generalized bi-capacity considers only the net predisposition.

The Choquet integral corresponding to the generalized bi-capacity is the following:

$$\begin{aligned} \mathcal{C}^B((x^+, x^-), v^*) &= \sum_{j=1}^{2n} (x_{(j)} - x_{(j-1)}) v^*(C_{(j)}, D_{(j)}) = \\ &= \sum_{j=1}^{2n} x_{(j)} (v^*(C_{(j)}, D_{(j)}) - v^*(C_{(j+1)}, D_{(j+1)})) \end{aligned}$$

where  $v^*(C_{(n+1)}, D_{(n+1)}) = 0$ .

Another model which can be seen as a special case of the generalized bipolar Choquet bi-integral is the following extension of the *CPT* model. Let us consider a capacity  $\mu_+$  for the positive evaluations  $x^+ = (x_1^+, \dots, x_n^+)$  and a capacity  $\mu_-$  for the negative evaluations  $x^- = (x_1^-, \dots, x_n^-)$  and the corresponding Choquet integral given by:

$$\begin{aligned} CPE(x) &= \mathcal{C}_{\mu^+}(x^+), \\ CNE(x) &= \mathcal{C}_{\mu^-}(x^-), \\ CE(x) &= \mathcal{C}_{\mu^+}(x^+) - \mathcal{C}_{\mu^-}(x^-). \end{aligned}$$

In this model, in contrast with the *CPT* model, the scales of the evaluations on single criteria are bivariate and therefore we do not necessarily have  $x^+ = -x^-$ ; on the contrary, in general, we have  $x^+ \neq -x^-$ .

### 5.3 Models based on preference differences in a bivariate setting

The idea of the bivariate evaluation space model can be extended also to models based on difference of preferences [20]. In this case it is interesting to take into consideration the bipolar or bivariate nature of the difference of preferences and the bipolar or bivariate nature of the evaluation of comprehensive differences of preference [23]. The

following preference model based on a bipolar representation both at the level of each criterion and of the comprehensive evaluation has been considered in section 3.4

$$x \succsim y \Leftrightarrow G(p_1(x_1, y_1), \dots, p_n(x_n, y_n)) \geq 0,$$

where  $G$  is a function nondecreasing in its arguments. A model considering a bivariate evaluation both at the level of single criterion and comprehensive evaluation ([23], [20]) can be written as

$$x \succsim^+ y \Leftrightarrow CPE(x, y) = G^+(p_1^+(x_1, y_1), \dots, p_n^+(x_n, y_n), p_1^-(x_1, y_1), \dots, p_1^-(x_n, y_n)) \geq 0$$

$$x \succsim^- y \Leftrightarrow CNE(x, y) = G^-(p_1^+(x_1, y_1), \dots, p_n^+(x_n, y_n), p_1^-(x_1, y_1), \dots, p_1^-(x_n, y_n)) \geq 0$$

$$x \succsim y \Leftrightarrow CE(x, y) = V(CPE(x, y), CNE(x, y)) \geq 0,$$

where in a given product space  $X = \prod_{i=1}^n X_i$

- $\succsim^+$  is a preference relation on  $X$  such that, for all  $x, y \in X$ , the semantics of  $x \succsim^+ y$  is “there are reasons in favor of the conclusion that  $x$  is at least as good as  $y$ ”,
- $\succsim^-$  is a preference relation on  $X$  such that, for all  $x, y \in X$ , the semantics of  $x \succsim^- y$  is “there are reasons against the conclusion that  $x$  is at least as good as  $y$ ”,
- $p_i^+ : X_i \times X_i \rightarrow \mathbb{R}_+$ ,  $i = 1, \dots, n$ , is a function such that, for all  $x_i, y_i \in X_i$ ,  $p_i^+(x_i, y_i)$  measures the positive part of the difference of preference of  $x_i$  over  $y_i$ ,
- $p_i^- : X_i \times X_i \rightarrow \mathbb{R}_+$ ,  $i = 1, \dots, n$ , is a function such that, for all  $x_i, y_i \in X_i$ ,  $p_i^-(x_i, y_i)$  measures the negative part of the difference of preference of  $x_i$  over  $y_i$ ,
- $G^+ : \mathbb{R}_+^{2n} \rightarrow \mathbb{R}$  is a function non decreasing in its first  $n$  arguments and non increasing in its last  $n$  arguments,
- $G^- : \mathbb{R}_+^{2n} \rightarrow \mathbb{R}$  is a function non increasing in its first  $n$  arguments and non decreasing in its last  $n$  arguments,
- $V : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  is a function increasing in its first argument and decreasing in its second argument.

This model is related to the four valued logic approach [41].

A specific formulation of this bivariate model of difference of preference based on the bipolar Choquet bi-integral has been proposed in [20].

## 5.4 Summary

Table 1 summarizes the different aspects and types of bipolarity of the main models presented in this paper; the associated type of Choquet integral is mentioned.



model	bipolarity of criteria	bipolarity of the comprehensive evaluation	Aggregation function $F$
capacity $\mu$	unipolar	unipolar	Choquet integral
pair of capacities $(\mu^+, \mu^-)$	bipolar univariate	bipolar univariate with linear net predisposition	CPT
bi-capacity $v$	bipolar univariate	bipolar univariate	Choquet integral w.r.t. a bi-capacity
bipolar capacity $\zeta$	bipolar univariate	unipolar bivariate	Choquet integral w.r.t. a bipolar capacity
generalized bipolar capacity	unipolar bivariate	unipolar bivariate	Choquet bi-integral

Table 1: Main models and variants of Choquet integral.

## 6 Concluding remarks

This paper has tried to account for several aspects of bipolarity in MCDA. Bipolarity is understood in a broad sense: the scales representing bipolar criteria (or overall evaluations) are either bipolar univariate or unipolar bivariate. In section 3, adopting a behaviorist or descriptive perspective, we have considered various models of preference and looked for those in which bipolar scales receive their most natural representation. Section 4 has adopted a constructive approach, i.e. it focuses on techniques for building bipolar scales and aggregating them within a suitable model in order to elicit a preference. Section 5 sketches a very broad panorama of models in which the bipolarity of the criteria is represented by bivariate scales; the overall evaluation itself can be bivariate. As compared to the classical additive value function model, the methods described allow for modeling criteria that interact differently in the positive and negative zones of the bipolar scales or with respect to the two variables of a bivariate scale.

Clearly, the expressive power of such models is superior to that of the usual additive value function model, i.e. more preferences can find a representation within the former. Also, eliciting some of these models in practice seems not to be overly complicated (see section 4). Since its intuitive content is easily perceived, it thus appears that bipolarity offers good perspectives for the development of new models in the field of MCDA, both on practical and theoretical grounds.

In particular, more effort is needed in order to better understand the role of bipolarity in models of preference. The conjoint measurement approach outlined in section 3 provides a general framework for analyzing preferences. Questions like producing a formal definition of bipolarity in such a framework or characterizing the aggregation methods based on bipolar scales deserve further investigation. The methods described in sections 4 and 5 rely on the Choquet integral and some variants thereof; getting, for example, a clearer understanding of which preferences can be represented by means of a Choquet integral but not by an additive value function surely is a question that is worth the attention of researchers.

## References

- [1] C.A. Bana e Costa, and J.C. Vansnick. The MACBETH approach: basic ideas, software and an application. In N. Meskens and M. Roubens, editors, *Advances in Decision Analysis*, pages 131–157, 1999, Kluwer Academic Publishers.
- [2] C.A. Bana e Costa, and J.C. Vansnick. Applications of the MACBETH approach in the framework of an additive aggregation model. *J. of Multicriteria Decision Analysis*, 6:107–114, 1997.
- [3] C. Blackorby, D. Primont, and R. Russell. *Duality, separability, and functional structure: Theory and economic applications*. North-Holland, New-York, 1978.
- [4] D. Bouyssou. Some remarks on the notion of compensation in MCDM. *European Journal of Operational Research*, 26:150–160, 1986.
- [5] D. Bouyssou and M. Pirlot. Nontransitive decomposable conjoint measurement. *Journal of Mathematical Psychology*, 46:677–703, 2002.
- [6] D. Bouyssou and M. Pirlot. Preferences for multi-attributed alternatives: Traces, dominance and numerical representations. *Journal of Mathematical Psychology*, 48:167–185, 2004.
- [7] D. Bouyssou and M. Pirlot. ‘Additive difference’ models without additivity and subtractivity. *Journal of Mathematical Psychology*, 48:263–291, 2004.
- [8] J.T. Cacioppo, W.L. Gardner, and G.G. Berntson. Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Personality and Social Psychology Review*, 1(1):3–25, 1997.
- [9] G. Choquet. Theory of capacities. *Annales de l’Institut Fourier*, 5:131–295, 1953.
- [10] G. Debreu. Representation of a preference ordering by a numerical function. In R. Thrall, C. H. Coombs, and R. Davies, editors, *Decision Processes*, pages 159–175, New York, 1954. Wiley.
- [11] P. C. Fishburn. Additive non-transitive preferences. *Economic Letters*, 34:317–321, 1990.
- [12] P. C. Fishburn. Continuous nontransitive additive conjoint measurement. *Mathematical Social Sciences*, 20:165–193, 1990.
- [13] P. C. Fishburn. Nontransitive additive conjoint measurement. *Journal of Mathematical Psychology*, 35:1–40, 1991.
- [14] M. Grabisch. The Choquet integral as a linear interpolator. In *10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004)*, pages 373–378, Perugia, Italy, July 2004.
- [15] M. Grabisch and Ch. Labreuche. Bi-capacities for decision making on bipolar scales. In *EUROFUSE Workshop on Informations Systems*, pages 185–190, Varenna, Italy, September 2002.

- [16] M. Grabisch and Ch. Labreuche. Bi-capacities. Part I: definition, Möbius transform and interaction. *Fuzzy Sets and Systems*, 151:211–236, 2005.
- [17] M. Grabisch and Ch. Labreuche. Fuzzy measures and integrals in MCDA. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis*. Kluwer Academic Publishers, 2005.
- [18] M. Grabisch, Ch. Labreuche, and J.C. Vansnick. On the extension of pseudo-Boolean functions for the aggregation of interacting bipolar criteria. *Eur. J. of Operational Research*, 148:28–47, 2003.
- [19] M. Grabisch, S. Greco, Ch. Labreuche, and R. Słowiński. The bipolar Sugeno integral. Manuscript, 2003.
- [20] S. Greco and J. Figueira. Dealing with interaction between bipolar multiple criteria preferences in outranking methods. Research Report 11-2003, INESC-Coimbra, Portugal, 2003.
- [21] S. Greco, B. Matarazzo, and R. Słowiński. Conjoint measurement and rough set approach for multicriteria sorting problems in presence of ordinal criteria. In A. Colorni, M. Paruccini, B. Roy (Eds.), *A-MCD-A Aide multicritre la dcision/Multicriteria decision aid*, pp. 117–144, European Commission, Joint Research Centre, Luxembourg, 2001.
- [22] S. Greco, B. Matarazzo, and R. Słowiński. Bipolar Sugeno and Choquet integrals. In *EUROFUSE Workshop on Informations Systems*, pp. 191–196, Varenna, Italy, September 2002.
- [23] S. Greco, B. Matarazzo, and R. Słowiński. Preference representation by means of conjoint measurement and decision rule approach. In D. Bouyssou, E. Jacquet-Lagrèze, P. Perny, R. Słowiński, D. Vanderpooten (Eds.), *Aiding Decisions with Multiple Criteria: Essays in Honour of Bernard Roy*, pp. 263–313, Kluwer Academic Publishers, Dordrecht, 2002.
- [24] D.H. Krantz, R.D. Luce, P. Suppes, and A. Tversky. *Foundations of measurement*, vol. 1: *Additive and polynomial representations*. Academic Press, New-York, 1971.
- [25] Ch. Labreuche and M. Grabisch. The Choquet integral for the aggregation of interval scales in multicriteria decision making. *Fuzzy Sets and Systems*, 137:11–26, 2003.
- [26] L. Lovász. Submodular function and convexity. In A. Bachem, M. Grötschel and B. Korte, editors, *Mathematical programming. The state of the art*, pp. 235-257, Springer Verlag, 1983.
- [27] J.-L. Marichal. Aggregation of interacting criteria by means of the discrete Choquet integral. In T. Calvo, G. Mayor and R. Mesiar, editors, *Aggregation operators: new trends and applications*, pp.224-244, Studies in Fuzzyness and Soft Computing, Vol. 97, Physica Verlag, 2002.
- [28] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum. *The measurement of meaning*. University of Illinois Press, Urbana, IL, 1957.

- [29] P. Perny. *Modélisation des préférences, agrégation multicritère et systèmes d'aide à la décision*. Thèse d'habilitation, Université Pierre et Marie Curie, Paris, 2000.
- [30] H. Raiffa. *Decision Analysis*. Addison-Wesley, Reading, Mass. 1968.
- [31] F.S. Roberts. *Measurement Theory*. Addison-Wesley, 1979.
- [32] B. Roy, D. Bouyssou. *Aide multicritère à la décision: Méthodes et cas*. Economica, Paris, 1993.
- [33] H. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138, 1956.
- [34] H.A. Simon. Theories of bounded rationality. In Peter Earl, editor, *The Legacy of Herbert Simon in Economic Analysis*, volume 1. Edward Elgar Publishing Ltd, 2001.
- [35] I. Singer. Extensions of functions of 0-1 variables and applications to combinatorial optimization. *Numerical Functional Analysis and Optimization*, 7(1):23-62, 1984.
- [36] J. Šipoš. Integral with respect to a pre-measure. *Mathematica Slovaca*, 29:141–155, 1979.
- [37] P. Slovic, M. Finucane, E. Peters, and D.G. MacGregor. The affect heuristic. In T. Gilovitch, D. Griffin, and D. Kahneman, editors, *Heuristics and biases: the psychology of intuitive judgment*, pages 397–420. Cambridge University Press, 2002.
- [38] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.
- [39] A. Tversky. Intransitivity of preferences. *Psychological Review*, 76:31–48, 1969.
- [40] A. Tversky and D. Kahneman. Advances in prospect theory: cumulative representation of uncertainty. *J. of Risk and Uncertainty*, 5:297–323, 1992.
- [41] A. Tsoukias and Ph. Vincke. A new axiomatic foundation of partial comparability. *Theory and Decision*, 39:79–114, 1995.
- [42] Ph. Vincke. *Multi-criteria decision aid*. Wiley, New York, 1992.
- [43] K. Vind. Independent preferences. *Journal of Mathematical Economics*, 20:119–135, 1991.
- [44] D. von Winterfeldt and W. Edwards. *Decision analysis and behavioral research*. Cambridge University Press, Cambridge, 1986.
- [45] P.P. Wakker. *Additive representations of preferences: A new foundation of decision analysis*. Kluwer, Dordrecht, 1989.