Giuseppe Munda
European Commission, Joint Research Centre
(JRC), Ispra, Italy
giuseppe.munda@ec.europa.eu
orcid.org/0000-0003-3398-2971

DOI: https://doi.org/10.14195/2183-203X_56_1

Evaluation and Assessment of Sustainability Policies

Avaliação de Políticas Sustentáveis

Giuseppe Munda

Received for publication: January 10, 2022
Revision accepted for publication: March 28, 2022

ABSTRACT

Sustainability policy evaluation and assessment seeks to answer the key question, *sustainability of what and whom*? Consequently, sustainability issues are multidimensional in nature and feature a high degree of conflict, uncertainty and complexity. Social multi-criteria evaluation (SMCE) has been explicitly designed for public policies; it builds on formal modelling techniques whose main achievement is the fact that the use of different evaluation criteria translates directly into plurality of values and dimensions underpinning a policy process. SMCE aims at being inter/multi-disciplinary (with respect to the technical team), participatory (with respect to the community) and transparent. SMCE can help deal with three different types of sustainability-related policy issues: 1) epistemological uncertainty (human representation of a given policy problem necessarily reflects perceptions, values and interests of those structuring the problem); 2) complexity (the existence of different levels and scales at which a hierarchical system can be analyzed implies the unavoidable existence of non-equivalent descriptions of it both in space and time); and 3) mathematical manipulation rules of relevant information (compensability versus non-compensability, preference modelling of intensities of preference, mixed information on criterion scores, weights as trade-offs versus weights as importance coefficients, choice of a proper ranking algorithm). This paper focuses on the these three issues and provides an overview of the SMCE approaches to them.

Keywords: Complexity theory; social multi-criteria evaluation; history of economic thought: social choice; SOCRATES software.

JEL Classification: B16; B23; C44; D04; D71; D81; H4; Q01.

Acknowledgement: This article has been developed in the context of the activities of the JRC Competence Centre on Modelling and Decision Analysis. The views expressed are purely those of the writer and may not in any circumstances be regarded as stating an official position of the European Commission.

1. Complexity, incommensurability and sustainability policies

> "… there is such a long tradition in parts of economics and political philosophy of treating one allegedly homogeneous feature (such as income or utility) as the sole 'good thing' that could be effortlessly maximized (the more the merrier), that there is some nervousness in facing a problem of valuation involving heterogeneous objects,… And yet any serious problem of social judgement can hardly escape accommodating pluralities of values,… We cannot reduce all the things we have reason to value into one homogeneous magnitude." (Sen, 2009, p. 239)

Since its origins, the concept of sustainable development is often considered a policy framework for win-win strategies (e.g. Barbier, 1987), allowing the full achievement of a plurality of goals in a variety of domains; but is this possible? A legitimate question is: sustainable development *of what and whom?* (Allen et al., 2002). Norgaard (1994, p.11) writes: "*consumers want consumption sustained, workers want jobs sustained. Capitalists and socialists have their "isms", while aristocrats and technocrats have their "cracies*".

Complexity arises when something is difficult to understand and impossible to analyse by using simple frameworks. However, when dealing with sustainability policy problems, there is a natural temptation to try to reduce them to simpler, more manageable elements. Although many definitions of complexity exist, a key common characteristic of complex systems is that the information space required to represent *relevant* aspects of a complex system cannot be compressed without losing relevant information (Gell-Mann, 1994; Prigogine and Stengers, 1981; Rosen, 1985, 1991; Simon, 1962).

To make things more difficult, systems involving humans are *reflexively* complex. Reflexive systems display two peculiar characteristics: "*awareness*" and "*purpose*", both requiring an additional "jump" in describing complexity. The presence of self-consciousness and purpose (*reflexivity*) means that these systems can continuously add new relevant qualities/attributes to be considered when explaining, describing or forecasting their behaviour (i.e. human systems are learning systems); this implies that complex adaptive systems become something else over time (Funtowicz et al., 1999).

Moreover, the existence of *different levels and scales* on which a hierarchical system can be analysed implies the unavoidable existence of non-equivalent descriptions of it. Even a simple "objective" description of a geographical orientation is impossible without taking an arbitrary subjective decision on the relevant system scale. In fact, the same geographical place, for example in Europe, may be considered to be in the north, south, east or west according to the scale chosen as a reference point (the whole Europe, a single State, a region, a specific place, etc.). Therefore, the problem of *multiple identities* in complex systems cannot be interpreted solely in terms of *epistemological plurality* (non-equivalent observers), but also necessarily in terms of *ontological characteristics* of the observed system (non-equivalent observations) (Giampietro et al., 2006, 2012; Giampietro and Mayumi, 2000, 2018). There is an unavoidable political dimension in any scientific description in as much as some decision is required regarding how to frame a policy problem. Therefore, to reach a ranking of policy options, there is a prior need to decide *what is important* for different social actors as well as *what is relevant* for the representation of the real-world entity described in the model.

No mathematical model, even if legitimate in its own terms, can be sufficient for a complete analysis of the reflexive properties of a real-world problem. These reflexive properties include the human dimensions of e.g. the ecological change and the transformations of human perceptions along the way. The learning process that takes place while analyzing the issue and defining policies will itself influence perceptions and alter significantly the decisional space in which alternative strategies are chosen. At the other end, institutional and cultural representations of the same system, while also legitimate, are on their own insufficient to define what should be done in any particular case.

In general, these concerns were not considered very relevant by scientific research as long as time was considered an infinite resource. On the other hand, the new nature of the problems faced in this third millennium implies that, when dealing with problems that may have long term consequences, we are confronting issues "*where facts are uncertain, values in dispute, stakes high and decisions urgent*" (Funtowicz and Ravetz, 1991).

Scientists cannot therefore provide any useful input without interacting with the rest of society while the rest of the society cannot make any sound decision without interacting with scientists. That is, the question of "how to improve the quality of a policy process" must be put, rather quickly, on the agenda of "scientists", "policy-makers" and indeed of society as a whole. This extension of the "peer community" is essential for maintaining the quality of the process of decision-making when dealing with reflexive complex systems. In relation to this objective Funtowicz and Ravetz have developed a new epistemological framework called *"Post-Normal Science"*, with which it is possible to deal better with two crucial aspects of science in the policy domain: *uncertainty* and *value conflict*. The term "post-normal" signals a divergence from the puzzle-solving exercises of normal science, in the Kuhnian sense.

In operational terms, one should admit that there is no optimal solution to the management of complex systems. If we want to avoid reductionism, it will be necessary to take incommensurable dimensions into account and to use different scientific languages describing disparate but legitimate representations of the same system. Accepting the complexity of natural and social systems is the first step in understanding how policy problems should be structured. A second step is to choose appropriate management and policy tools: *those that address rather than ignore complexity.*

Multi-criteria decision analysis is becoming more and more popular both in the private and public sectors (see e.g. Figueira et al., 2016). Arrow and Raynaud (1986) considered the so-called *"industrial outranking problem"*, where a typical business-person is the reference decision-maker, who wishes *"to make safer the equilibrium of the productions of the firm"* (Arrow and Raynaud, 1986, p. 9). Typical business criteria may be market standing, innovation level, productivity, profitability, physical and financial resources, etc. In empirical evaluations of public projects and public provided goods, multi-criteria decision analysis seems to be an adequate policy tool as well, since it allows taking into account a wide variety of evaluation criteria (e.g. environmental impact, employment, distributional equity, and so on) which can measure the effects on the social welfare.

Social *multi-criteria evaluation (SMCE) techniques* have the potential to take into account conflictual, multidimensional and uncertain properties of policy decisions (Munda, 2004, 2008). SMCE can therefore provide insights into the nature of conflicts and complexity and facilitate the process of reaching political compromises by explaining divergent values and increasing the transparency of the decision process.

SMCE proceeds on the basis of following main concepts: dimensions, objectives, criteria, weights, criterion scores, impact matrix and compromise solution. Dimension is the highest hierarchical level of analysis and indicates the scope of objectives, criteria and criterion scores. The general categories of economic, social and environmental impacts are dimensions. Objectives indicate the direction of change desired, e.g. growth has to be maximised, social exclusion has to be minimised, carbon dioxide emissions have to be reduced. A criterion is a function that associates each alternative action with a variable indicating its performance from a specific point of view. Weights are often used to represent the relative importance attached to dimensions, objectives and criteria. The idea behind this practice is very intuitive and easy, that is, to place the greatest number in the position corresponding to the most important factor.

In operational terms, the application of a SMCE framework involves the following seven main steps (Munda, 2008):

i. Description of the relevant social actors. For example, institutional analysis may be performed on historical, legislative and administrative documents to provide a map of the relevant social actors.
ii. Definition of social actors' values, desires and preferences by using focus groups or other participatory techniques such as anonymous questionnaires and personal interviews.
iii. Generation of policy options and selection of evaluation criteria is a process of co-creation resulting from a dialogue between analysts and social actors.
iv. Construction of the multi- criteria impact matrix synthesising the scores of all criteria for all alternatives, i.e. the performance of each alternative according to each criterion.
v. Construction of a social impact matrix (i.e. a matrix showing the impacts of the alternatives on the various social actors).
vi. Application of a mathematical procedure to aggregate criterion scores and obtain a final ranking of the available alternatives. The importance of mathematical approaches is their ability to allow a consistent aggregation of the diverse information.
vii. Sensitivity analyses help elucidating conflicts among alternatives and objectives and testing the robustness of the model. Expressing results in terms of sensitivities, both to uncertainties in the model as well as divergent values, reveals model biases as rank orders of alternatives potentially change (Saltelli et al., 2004, 2013).

These seven steps are not rigid. On the contrary, flexibility and adaptability to actual situations are among the main advantages of SMCE. As a tool for policy evaluation and conflict management, SMCE has demonstrated its applicability to problems in various geographical and cultural contexts. A recent and exhaustive overview of world-wide SMCE applications can be found in Etxano and Villalba-Eguiluz (2021).

In my experience, the empirical argument that SMCE deals with complex issues in an effective way is accepted in policy contexts, but often it is not a sufficient one for scholars.

Therefore, more formal arguments have to be developed; in this context analytical philosophy is very useful. The starting point is the relationship between comparability and commensurability (Chang, 1997; O'Neill, 1993). From a philosophical perspective, it is possible to distinguish between the concepts of a) *strong comparability* (there exists a single comparative term by which all different actions can be ranked), implying *strong commensurability* (a common measure of the various consequences of an action based on an interval or ratio scale of measurement, such as money or energy), or *weak commensurability* (a common measure based on an ordinal scale of measurement, such as consumer's utility); and b) *weak comparability,* which implies incommensurability (Martinez-Alier et al., 1998). Incommensurability can be further distinguished into technical and social ones (Munda 2004). *Technical incommensurability* refers to the impossibility of compressing different dimensions into a single metric consistent will all the original dimensions and *social incommensurability* refers to the existence of an irreducible value conflict among social actors, when deciding what common comparative term should be used to rank alternative options.

Two other useful concepts are *set and rod commensurability* (Munda, 2016). Commensurability, a necessary condition for strong comparability, can be implemented in two different ways:

1. By looking for a more general category (set) that can contain <u>all</u> the characteristics of the objects we wish to compare; these characteristics are described by using adjectives. This can be defined as "set commensurability" (e.g. apples and oranges are legitimately lumped together as fruit, along with grapes, bananas, etc.).
2. By finding one property common to all objects to be compared and measurable by using one measurement unit, obviously comparison of objects is possible according to the characteristics of this property only. This can be defined as "rod commensurability".

Of course, when possible, set commensurability is the most attractive one since apparently no information is lost in the comparison process, while rod commensurability always requires a kind of reductionism. Here the question is: *when set commensurability is possible and correct?* Geach's (1956) distinction between *attributive and predicative adjectives* can help us in answering this question. In Geach's own words: *"There are familiar examples of what I call attributive adjectives. Big and small are attributive;* x *is a big flea does not split up into* x *is a flea and* x *is big, nor* x *is a small elephant into* x *is an elephant and* x *is small; for if these analyses were legitimate, a simple argument would show that a big flea is a big animal and a small elephant is a small animal. Again, the sort of adjective that the mediaevals called alienans is attributive;* x *is a forged banknote does not split up into* x *is a banknote and* x *is forged, nor* x *is the putative father of* y *into* x *is the father of* y *and* x *is putative. On the other hand, in the phrase a red book, red is a predicative adjective in my sense, although not grammatically so, for is a red book logically splits up into is a book and is red. I can now state my first thesis about good and evil: good and bad are always attributive, not predicative, adjectives"* (Geach, 1956, p. 32).

Although Geach's arguments were developed in the context of moral philosophy, they have an extraordinary explicative power for evaluation problems too. In fact, evaluation is all about an option *a* being declared better, worse or equal than another option *b*. However, although Geach saw the clear difference between predicative and attributive adjectives, he only gave examples of them but no general definition was provided, the new concepts of absolute and relative predicative adjectives were then recently introduced (Munda, 2016). An adjective is *absolute predicative* if its meaning does not change in relation to the subsets considered. It is an intrinsic characteristic of the object considered. The characteristic of being a red-headed person does not change if we consider subsets such as police officers, politicians, scientists or basketball players. In terms of measurement theory, an absolute predicative adjective is always measured on a nominal scale i.e. individual characteristics are grouped into a set of equivalence classes.

An adjective is *relative predicative* if it does not hold its meaning once one switches to a larger or different set of objects. It describes a characteristic that is dependent on the relative comparisons among the objects considered. In terms of measurement theory, a relative predicative adjective is always measured on an ordinal scale. An adjective is *attributive* if it does not have any meaning when referred to a different set or problem framework. A good person can be a bad basketball player and a good economist can be a bad person.

At this stage, the following conclusion can be derived: *when considering adjectives,* set commensurability is correct only if the adjectives considered are *absolute predicative* ones. An adjective *Z* is absolute predicative if it passes the *ontological* check of the two following logical tests: test (1) implies statements such as *"if x_1 is red and it is a car then x_1 is a red car*" and test (2) *"if x_1 is a red car and all cars are a mean of transport then x_1 is a red mean of transport"*. Adjectives that fail such tests are relative predicative or attributive adjectives, which always imply weak comparability based on incommensurability. For example, the adjective "good" clearly fails (2), statements such as *"x_1 is a good car, all cars are a mean of transport, and therefore x_1 is a good mean of transport"* or *"x_1 is a good scientist, all scientists are human beings, and therefore x_1 is a good human being"* are invalid arguments on the light of a real-world corroboration.

In summary, the point is that different metrics are also linked to different social objectives and values; in this context, the statement *"x is better than y"* implies an answer to two questions*: 1) according to what? 2) According to whom?* To use only one measurement unit for incorporating a plurality of dimensions, objectives and values, implies reductionism necessarily. If evaluative adjectives like "good" and "valuable" are attributive in standard uses, this does not however preclude the possibility of rational choices between objects, which do not fall into the range of a single comparative. *Weak comparability* based on *incommensurability* is compatible with the existence of such limited ranges; for example, regional sustainability is not evaluated as good or bad as such, but rather in relation to different descriptions or indicators. It can be at one and the same time a *"good income per capita"* and a *"bad social inclusion"*, a *"beautiful landscape"* and a *"heavy pollution"*. The use of these value terms in such contexts is attributive clearly.

In summary, we can conclude that incommensurability does not imply incomparability; on the contrary, it is in terms of weak comparability that evaluation has to take place in practice. This is exactly the basic idea of social multi-criteria evaluation.

2. Tackling the Discrete Multi-Criterion Problem in a SMCE Framework

> "Non zeli ad zelum, nec meriti ad meritum, sed solum numeri ad numerum fiat collatio" (Gregorius X (1210-1276, Papa, 1271), VI Decretalium, lib. I, tit. VI, cap. 9)

Results of a real-world policy exercise depend strongly on the way a given problem is structured during the evaluation process obviously, but mathematical models play a very important role: the one of guaranteeing consistency between assumptions used and results obtained. This implies to take into account the technical *uncertainties* properly, such as:

i. *Compensability versus non-compensability.*
ii. *A relevant preference modelling of intensities of preference.*
iii. *Mixed information on criterion scores (i.e. various measurement scales and related uncertainty).*
iv. *Weights as trade-offs versus weights as importance coefficients.*
v. *A proper ranking algorithm.*

Here, I will make an overview of the main solutions proposed inside the SMCE framework to deal with these issues and that have been implemented in a software tool called SOCRATES (SOcial multi-CRiteria AssessmenT of European policieS) (all methodological and mathematical details behind the SOCRATES software can be found in Azzini and Munda, 2020; Munda, 2004, 2009, 2012, 2022)[1].

The *discrete multi-criterion problem* can be described in the following way: A is a finite set of N feasible options (or alternatives); M is the number of different points of view or evaluation criteria g_m $m=1, 2, \dots, M$ considered relevant in a policy problem, where the option a is evaluated to be better than option b (both belonging to the set A) according to the *m-th* point of view if $g_m(a)>g_m(b)$. In synthesis, the information contained in the impact matrix useful for solving the so-called multi-criterion problem is:

i. *Intensity of preference* (when quantitative criterion scores are present).
ii. *Number* of criteria in favour of a given alternative.
iii. *Weight* attached to each single criterion.
iv. *Relationship* of each single alternative with all the other alternatives.

Combinations of this information generate different aggregation conventions, i.e. manipulation rules of the available information to arrive at a preference structure. The aggregation of several criteria implies taking a position on the fundamental issue of compensability. *Compensability* refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some criteria by a sufficiently large advantage on another criterion, whereas smaller advantages would not do the same. Thus, a preference relation is non-compensatory if no trade-off occurs and is compensatory otherwise. The use of weights with intensity

[1] See also https://knowledge4policy.ec.europa.eu/modelling/topic/social-multi-criteria-evaluation-policy-options_en/socrates_en

of preference originates compensatory multi-criteria methods and gives the meaning of trade-offs to the weights. On the contrary, the use of weights with ordinal criterion scores originates non-compensatory aggregation procedures and gives the weights the meaning of importance coefficients (Bouyssou, 1986; Bouyssou and Vansnick, 1986; Keeney and Raiffa, 1976; Podinovskii, 1994; Roberts, 1979; Vansnick, 1986).

The concept of importance I am using along this paper can be classified as symmetrical importance, that is "if we have two non-equal numbers to construct a vector in R^2, then it is preferable to place the greatest number in the position corresponding to the most important criterion" (Podinovskii, 1994, p. 241).

A common practice is the pragmatic solution of no criterion weighting. However, the fact that all criteria have the same weight does not guarantee at all that dimensions have the same weight. This would be guaranteed only under the condition that all the dimensions have the same number of criteria; this of course is quite unnatural and artificial, and even dangerous. On the contrary, different criterion weights can guarantee that all the dimensions are considered equal. A reasonable practice can be to start by giving the same weight to each dimension and then splitting each weight among the criteria of any dimension proportionally. Figures 1 and 2, obtained by means of the SOCRATES software, represent these situations in a graphical way. As one can see in this case the relation dimensions/criteria is a very peculiar one. In fact, most of criteria belong to the economic dimension, while other dimensions are much less populated. This implies that the starting weighting assumption can be only equal dimension weights because otherwise (under the equal criterion weighting assumption) the economic dimension would dominate since its weights would be higher than 50% of all dimensions considered (in technical terms it would become a *dictator*).

Figure 1: Equal criterion weighting (economic dimension receives 61.54%)

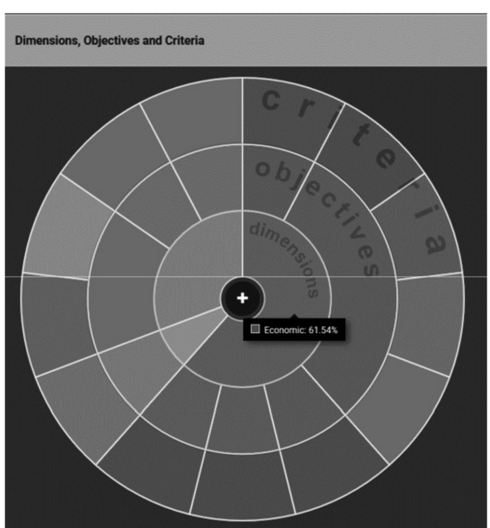
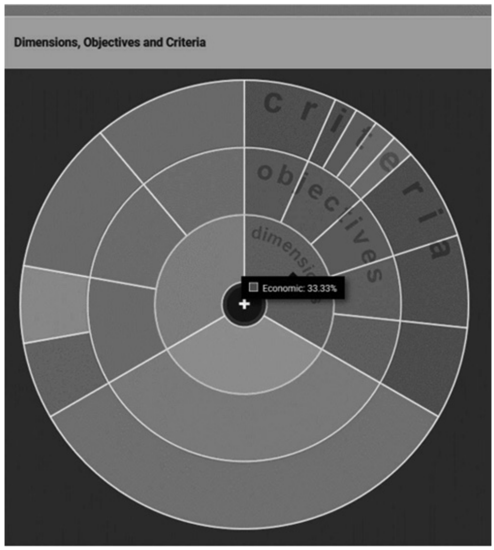


Figure 2: Equal dimension weighting (economic dimension receives 33.33%)



Of course, one could assume that some dimensions are more important than other ones, and thus their weight should be higher, but this should be justified. Finally, one should note that weights can be used in the way described here, only if they have the *meaning of importance*, which depends on the fact that they are combined with *non-compensatory aggregation mathematical rules*.

2.1. Pair-wise comparison of alternatives

The famous bald paradox in Greek philosophy (how many hairs one has to cut off to transform a person with hairs to a bald one?), later on Poincaré (1935, p. 69) and finally Luce (1956) made the point that the transitivity of indifference relation is incompatible with the existence of a sensibility threshold below which an agent either does not sense the difference between two elements, or refuses to declare a preference for one or the other. Luce was the first one to discuss this issue formally in the framework of preference modelling. Mathematical characterisations of preference modelling with thresholds can be found in Roubens and Vincke (1985).

By introducing a positive constant indifference threshold q the resulting preference model is the *threshold model*:

$$\begin{cases} a_j P a_k \Leftrightarrow g_m(a_j) > g_m(a_k) + q \\ a_j I a_k \Leftrightarrow | \, g_m(a_j) - g_m(a_k) \, | \leq q \end{cases} \tag{1}$$

where a_j and a_k belong to the set A of alternatives and g_m to the set G of evaluation criteria.

Real life experiments show that often there is an intermediary zone inside which an agent hesitates between indifference and preference. This observation led to the so-called *double threshold model* where variable indifference and preference thresholds are introduced, that is:

$$\left\{\begin{array}{l} a_j P a_k \Leftrightarrow g_m(a_j) > g_m(a_k) + p(g_m(a_k)) \\ a_j Q a_k \Leftrightarrow g_m(a_k) + p(g_m(a_k)) \geq g_m(a_j) > g_m(a_k) + q(g_m(a_k)) \\ a_j I a_k \Leftrightarrow \left\{\begin{array}{l} g_m(a_k) + q(g_m(a_k)) \geq g_m(a_j) \\ g_m(a_j) + q(g_m(a_j)) \geq g_m(a_k) \end{array}\right. \end{array}\right\} \tag{2}$$

For any $m = 1, 2, \dots, M$, being p a positive preference threshold. Relation Q has been called “weak preference” by Roy (1985, 1996). It translates the decision-maker’s hesitation between indifference and preference and not “less strong” preference as its name might lead to believe. A criterion with both preference and indifference thresholds is called a pseudo-criterion. A *pseudo-order structure* is a double threshold model upon which the following consistency condition is imposed:

$$g_m(a_j) > g_m(a_k) \Leftrightarrow \left\{\begin{array}{l} g_m(a_j) + q(g_m(a_k)) > g_m(a_k) + q(g_m(a_k)) \\ g_m(a_j) + p(g_m(a_k)) > g_m(a_k) + p(g_m(a_k)) \end{array}\right. \tag{3}$$

A problem is that the modelling procedure based on the notion of a pseudo-criterion may present a serious lack of stability. Such undesirable discontinuities make a sensitivity analysis (or robustness analysis) necessary; however, this important analysis step is very complex to manage because of the combinatorial nature of the various sets of data. One should combine variations of 2 thresholds (indifference and preference) and k possible scores of the M criteria. A solution to this problem may come from the concept of *valued preference relations*, that is a preference relation where there is a need to assign to each ordered pair of alternatives (a_j, a_k) a value $v(a_j,a_k)$ representing the “strength” or the “degree of preference” (Fishburn, 1970, 1973a; Ozturk et al., 2005; Roubens and Vincke, 1985).

In this framework, an interesting concept is the one of a *fuzzy preference relation* (Kacprzyk and Roubens, 1988). If A is assumed to be a finite set of N alternatives, a *fuzzy preference relation* is an element of the $N \times N$ matrix $R = (r_{jk})$, i.e.

$$r_{jk} = m_R(a_j, a_k), \text{ with } j, k = 1, 2, \dots, N \qquad \text{and} \qquad 0 \leq r_{jk} \leq 1. \tag{4}$$

$r_{jk} = 1$ indicates the maximum credibility degree of preference of a_j over a_k; each value of r_{jk} in the open interval $(0.5, 1)$ indicates a definite preference of a_j to a_k (a higher value means a stronger credibility); $r_{jk} = 0.5$ indicates the indifference between a_j and a_k. This definition implies that fuzzy preference relations can be used as mathematical models of intensity of preference.

Usually, fuzzy preference relations are assumed to satisfy two properties:

(a) reciprocity, i.e. $r_{jk} + r_{kj} = 1$;
(b) max-min transitivity, i.e. if a_i is preferred to a_j and a_j is preferred to a_k, then a_i should be preferred to a_k with at least the same credibility degree, that is

$$r_{ij} \geq 0.5, \quad r_{jk} \geq 0.5 \quad \Rightarrow r_{ik} \geq \min\left(r_{ij}, r_{jk}\right). \tag{5}$$

By using a fuzzy preference modelling since small variations of input data (scores and thresholds) modify in a continuous way; the consequential preference model can allow one to avoid the drawbacks of the pseudo-criterion model.

Let's now consider any criterion g_m belonging to the set G and any pair of alternatives a_j and a_k belonging to the set A. The criterion scores $g_m(a_j)$ and $g_m(a_k)$ are measured on an interval or ratio scale. Let p_m be a constant preference threshold and q_m a constant indifference threshold for the criterion g_m. Then the credibility degree m of preference (P) and indifference (I) relations between a_j and a_k can be computed as follows:

$$\begin{cases} \mu_{(a_jPa_k)} = \left[1 + c_{pm}\left(g_m(a_j) - g_m(a_k)\right)^{-2}\right]^{-1} \\ \mu_{(a_jIa_k)} = e^{-cqm\left|g_m(a_j) - g_m(a_k)\right|} \\ \mu_{(a_kPa_j)} = \left[1 + c_{pm}\left(g_m(a_k) - g_m(a_j)\right)^{-2}\right]^{-1} \end{cases} \tag{6}$$

where $\mu_{(a_jIa_k)} \quad \forall \; g_m(a_j)$ and $g_m(a_k)$ and

$$\mu_{(a_jPa_k)} \qquad \text{if} \quad g_m(a_j) - g_m(a_k) > 0 \tag{7}$$

$$\mu_{(a_kPa_j)} \qquad \text{if} \quad g_m(a_j) - g_m(a_k) < 0. \tag{8}$$

It has to be admitted that the shape of the function representing the credibility degrees of the preference and indifference relations is arbitrary. However, some consistency requirement such as that the functions are continuous and monotonic and that $p_m > q_m$ exist, thus reducing considerably the degree of arbitrariness.

2.2. The case of mixed information on criterion scores

Ideally the information available for a policy problem should be precise, certain, exhaustive and unequivocal. But in real-world problem, it is often necessary to use information which does not have these characteristics and thus to deal with uncertainty of a stochastic and/or fuzzy nature present in the data. Let's then introduce a more realistic assumption, i.e. that the set of evaluation criteria $G = \{g_m\}$, $m = 1, 2,..., M$, on the set $A = \{a_n\}$, $n = 1, 2,..., N$ of potential alternatives may include either crisp (that is impacts measured on interval or ratio scales), stochastic and fuzzy criterion scores. A very useful concept for quantifying vagueness on criterion scores is the one of a fuzzy number. A *fuzzy number* is simply a fuzzy set in the real line and is completely defined by its membership function such as μ_x: $R \rightarrow [0, 1]$. For computational purposes, in general this definition is restricted to those fuzzy numbers which are both normal and convex.

Normality: $\sup\{\mu(\mathrm{x})\} = 1$ *with* $x \in R$.

Convexity: $\mu\{\lambda x_1 + (1 - \lambda)x_2\} \geq \min\{\mu(\mathrm{x}_1), \mu(\mathrm{x}_2)\}$ *with* $x \in R$ *and* $\lambda \in \{0, 1\}$.

The requirement of convexity implies that the points of the real line with the highest membership values are clustered around a given interval (or point). This fact allows one to easily understand the semantics of a fuzzy number by looking at its distribution.

A general type of fuzzy number is the so-called *L-R fuzzy number*; it is defined as follows:

$$\mu(x) = \begin{cases} \dfrac{F_L(x-m)}{\alpha} & if \quad -\infty < x < m, \alpha \in R^+ \\ 1 & if \quad x = m \\ \dfrac{F_R(x-m)}{\delta} & if \quad m < x < +\infty, \delta \in R^+ \end{cases} \tag{9}$$

where *m, a, d,* are the "middle" value, the left-hand and the right-hand variation, respectively. $F_L(x)$ is a monotonically increasing membership function and $F_R(x)$, not necessarily symmetric to $F_L(x)$, is a monotonically decreasing function.

The treatment of mixed information on criterion scores proposed here is mainly based on the semantic distance originally developed in Munda (1995) and furtherly formalised in Munda (2012). This semantic distance allows dealing consistently with an impact matrix which may include crisp, stochastic or fuzzy measurements of the performance of an alternative with respect to an evaluation criterion. Therefore, the multi-criterion problem is considered in its more general form. The only restriction is that in the case of fuzzy information, continuous, convex membership functions allowing also a definite integration are required.

Let's start with the case of fuzzy criterion scores:
if $m_1(x)$ and $m_2(x)$ are two fuzzy numbers, one can write (see Ragade and Gupta, 1977, for a formal proof):

$$f(x) = k_1\mu_1(x) \quad \text{and} \quad g(y) = k_2\mu_2(x), \tag{10}$$

where $f(x)$ and $g(y)$ are two functions obtained by rescaling the ordinates of $m_1(x)$ and $m_2(x)$ through k_1 and k_2, such as

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{+\infty} g(y)\,dy = 1 \tag{11}$$

The distance between all points of the membership functions is computed as follows:

If $f(x)$ is defined on $X = [x_L, x_U]$ and $g(y)$ is defined on $Y = [y_L, y_U]$, where sets X and Y can be non-bounded from one or either sides, then

$$S_d(f(x), g(y)) = \iint_{x,y} |x - y| f(x) g(y) \, dydx \tag{12}$$

If the intersection between the 2 membership functions is empty, it is $x > y \ \forall x \in X$ and $\forall y \in Y$, it follows that a continuous function in 2 variables is defined over a rectangle. Therefore, the double integral can be calculated as iterated single integrals; the result is

$$S_d(f(x), g(y)) = |E(x) - E(y)|, \tag{13}$$

where $E(x)$ and $E(y)$ are the expected values of the 2 membership functions.

When the intersection between 2 fuzzy sets is not empty, their distance is greater than the difference between the respective expected values since $|x - y|$ is always greater than $(x - y)$. In this case one finds

$$S_d(f(x), g(y)) = \int_{-\infty}^{+\infty} \int_{x}^{+\infty} (y - x) f(x) g(y) \, dydx + \int_{-\infty}^{+\infty} \int_{-\infty}^{x} (x - y) f(x) g(y) \, dydx \tag{14}$$

This is the case of a double integral over a general region; since this is not vertically simple or horizontally simple, it is not possible its computation by means of iterated integration, but it is necessary to take the limit of the Riemann sum. This problem can be easily overcome by means of numerical analysis.

From a theoretical point of view, the following main conclusions can be drawn:

(a) the absolute value metric is a particular case of the semantic distance;
(b) the comparison between a fuzzy number and a crisp number is equal to the difference between the expected value of the fuzzy number and the value of the crisp number considered;
(c) stochastic information can be taken into account too.

In sum, this semantic distance allows one to deal with fuzzy numbers, probability distributions and crisp numbers with the theoretical guarantee that all these sources of information are tackled equivalently, thus solving an open problem for multi-criterion methods dealing with mixed information. Of course, this search for an equivalent treatment of available information implies a trade-off with precision. For example, if stochastic information only is available, a stochastic dominance approach is more effective (see e.g. Markowitz, 1989, Martel and Zaras, 1995), or if fuzzy numbers only have to be compared, Matarazzo and Munda (2001) present a more sophisticated approach based on area comparison. However, in the case of mixed information in a multi-criterion framework, the semantic distance illustrated here is probably the best available compromise solution between generality and precision. Moreover, the use of this semantic distance allows a homogeneous preference modelling on all the criteria, impossible otherwise. Going back to the pair-wise comparison of alternatives, let's assume $f(x) = g_m(a_j)$ *and* $g(y) = g_m(a_k)$, where g_m is any criterion belonging to the set G and a_j and a_k any pair of alternatives belonging to the set A. The criterion scores $g_m(a_j)$ and $g_m(a_k)$ are *fuzzy or stochastic* in nature. Let p_m be a preference threshold and q_m an indifference threshold for the criterion g_m. Then it is:

$$\begin{cases} \mu_{(a_j P a_k)} = \left[1 + c_{pm}\left(\iint_{x,y} (x-y)\ f(x)\, g(y)\, dydx\right)^{-2}\right]^{-1} \\ \mu_{(a_j I a_k)} = e^{-c_{qm}\left(\iint_{x,y} |x-y|\, f(x) g(y) dydx\right)} \\ \mu_{(a_k P a_j)} = \left[1 + c_{pm}\left(\iint_{y,x} (y-x)\ f(x)\, g(y)\, dydx\right)^{-2}\right]^{-1} \end{cases} \tag{15}$$

where $\mu_{(a_j I a_k)}$ $\quad \forall\ x, y$ and

$$\mu_{(a_j P a_k)} \qquad if \quad \iint_{x,y} (x-y)\ f(x)\, g(y)\, dydx > 0 \tag{16}$$

$$\mu_{(a_k P a_j)} \qquad if \quad \iint_{x,y} (x-y)\ f(x)\, g(y)\, dydx < 0. \tag{17}$$

One should note that the comparison between the criterion scores of each pair of actions is carried out by means of the semantic distance. Since the absolute value metric is a particular case of this distance, fuzzy, stochastic and crisp criterion scores are dealt with equivalently.

2.3. Introducing weights as importance coefficients

At this point, a very sensitive step has still to be tackled i.e. the exploitation of the inter-criteria information in the form of weights. Let's then assume the existence of a set of criterion weights $W = \{w_m\}$, $m = 1, 2,..., M$, with $\Sigma_{m=1}^{M} w_m = 1$ derived as importance coefficients. The problem here is the theoretical guarantee that weights are really treated as importance coefficients and not as trade-offs. The point is that no connection must be done between criterion weights and the corresponding criterion intensity of preference. Our objectives are then:

(a) to find a way to combine weights with credibility degrees without a direct interpretation of the latter as intensity of preference;
(b) to divide each criterion weight in 2 parts proportionally to the credibility degrees of the indifference and preference fuzzy relations. In doing so, the requirement that $\Sigma_{m=1}^{M} w_m = 1$ should not be lost.

Let's define μ_p as the fuzzy preference relation between a pair of alternatives and μ_I as the fuzzy indifference relation between the same pair of alternatives. Let's put

$\mu_{min} = \min(\mu_p, \mu_I)$ and $\mu_{max} = \max(\mu_p, \mu_I)$. Clearly, it is $\mu_p = \mu_{min}$ on the left of the intersection point between the indifference and the preference fuzzy relations and vice-versa on the right. A criterion weight w_m is divided proportionally to μ_p and μ_I, according to equation (18):

$$\begin{cases} W_{m1} = W_m \dfrac{\mu_{\min}}{\mu_{\max} + \mu_{\min}} \\ W_{m2} = W_m \dfrac{\mu_{\max}}{\mu_{\max} + \mu_{\min}} \end{cases} \tag{18}$$

Equation (18) presents the following properties:

$$W_{m1} + W_{m2} = W_m \tag{19}$$

$$if\ \mu_{\min} = 0 \Rightarrow W_{m2} = W_m \tag{20}$$

$$if\ \mu_{\min} = \mu_{\max} = 0 \Rightarrow W_m = 0 \tag{21}$$

$$if\ \mu_{\min} = \mu_{\max} \Rightarrow W_{m1} = W_{m2} = \frac{1}{2} W_m \tag{22}$$

As a consequence, equation (18) fits our objective that the addition of all weights should be kept equal to one perfectly. Moreover, in equation (18) no direct use of the concept of intensity of preference is done; as a result, we can be sure that criterion weights are used consistently with their nature of importance coefficients. Finally, if a criterion score is *ordinal in nature*, it can be considered a particular case where $\mu_{\min} = 0$. Again, *the treatment of crisp, fuzzy, stochastic and ordinal criterion scores is perfectly equivalent.* Moreover, when indifference and preference thresholds are not used, the corresponding criteria can be dealt with as ordinal criteria, where

$$\begin{cases} a_j P a_k \Leftrightarrow g_m(a_j) > g_m(a_k) \\ a_j I a_k \Leftrightarrow g_m(a_j) > g_m(a_k) \end{cases} \tag{23}$$

Now a $N \times N$ matrix E can be built, where any generic element e_{jk}, with $j \neq k$, is the result of the pair-wise comparison, according to all the M criteria, between alternatives j and k. Such a global pair-wise comparison is obtained by means of equation (24):

$$e_{jk} = \Sigma_{m=1}^{M} \left(w_m (P_{jk}) + \frac{1}{2} w_m (I_{jk}) \right) \tag{24}$$

where $w_m(P_{jk})$ and $w_m(I_{jk})$ are derived from μp and μI through equation (18). It is

$$e_{jk} + e_{kj} = 1. \tag{25}$$

Property (25) is very important since it allows us to consider matrix E as a *voting matrix* i.e., a matrix where instead of using criteria, alternatives are compared by means of voters' preferences (with the principle one agent, one vote). This analogy between the multi-criterion problem and the social choice one, as noted by Arrow and Raynaud (1986), is very useful for tackling the step of ranking the N alternatives in a consistent axiomatic framework.

2.4. Ranking algorithm

Vansnick (1990) showed that the two main approaches in multi-criteria decision analysis i.e., the compensatory and non-compensatory ones can be directly derived from the seminal work of Borda (1784) and Condorcet (1785). Indeed, looking at social choice literature, one can realize that various ranking procedures used in multi-criterion methods have their origins in social choice. Just to give a few examples, the weakness-strength approach, typical of outranking methods (Roy, 1985, 1996), has a clear derivation from two Condorcet consistent rules, i.e. the Copeland (1951) and Simpson (1969) rules; Arrow-Raynaud propose a sequential procedure (building on Köhler, 1978) which is also based on some principles of the Condorcet rule; the so-called frequency matrix approach (Hinloopen et al., 1983; Matarazzo, 1988) comes directly from Borda algorithm, or the permutation method (Paelinck, 1978), has a strict connection with an original Condorcet approach developed to tackle cycles, and so on.

Given that there is a consensus in the literature that the Condorcet' theory of voting is non-compensatory (Vansnick, 1986) and useful for generating a ranking of the available alternatives while Borda's one is more useful for isolating one alternative, considered the best (Moulin, 1988; Truchon, 1995; Young, 1988, 1995), here clearly it is advisable to follow the Condorcet tradition[2] (since in a SMCE framework non-compensability and a complete ranking of alternatives are considered desirable properties (Munda, 2004)).

A basic problem inherent in the Condorcet's approach is the presence of cycles, i.e. cases where *aPb, bPc* and *cPa* may be found. This problem has been studied by various scientists (e.g., Fishburn, 1973; Kemeny, 1959; Moulin, 1985; Truchon, 1995; Young and Levenglick, 1978, Vidu, 2002; Weber, 2002). Now the question is: Is it possible to tackle the cycle issue in a general way? The answer to this question is yes, and it is generally known in social choice as the Kemeny method. However, in reality other scientists, including Condorcet himself, have contributed to the development of this ranking method. The historical reconstruction of this method and a deeper methodological analysis can be found in Munda (2008, Chapter 6). Here, I just synthesise some main points.

[2] Arrow and Raynaud (1986) also arrive at the conclusion that a Condorcet aggregation algorithm has in general to be preferred in a multi-criterion framework. They show that whenever the majority rule can be operationalized, it should be applied. However, the majority rule often produces undesirable intransitivities, thus "*more limited ambitions are compulsory. The next highest ambition for an aggregation algorithm is to be Condorcet*" (Arrow and Raynaud, 1986, p. 77).

Condorcet himself was aware of the problem of cycles in his approach; he built examples to explain it and he was even close to find a consistent rule able to rank any number of alternatives when cycles are present. Attempts of clarifying, fully understanding and axiomatizing Condorcet's approach for solving cycles have been mainly done by Kemeny (1959) who made the first intelligible description of the Condorcet approach, and by Young and Levenglick (1978) who made its clearest exposition and complete axiomatization. For this reason, I call this approach the Condorcet-Kemeny-Young-Levenglick ranking procedure, or the C-K-Y-L ranking procedure. Its main methodological foundation is the maximum likelihood concept. The maximum likelihood principle selects as a final ranking the one with the maximum pair-wise support. This selected ranking is also the one which involves the minimum number of pair-wise inversions. The selected ranking is also a median ranking for those composing the profile (in multi-criteria terminology it is the "compromise ranking" among the various conflicting points of view), for this reason the corresponding ranking procedure is often known as the Kemeny median order.

A problem of the C-K-Y-L ranking procedure is that it does not respect the axiom of independence of irrelevant alternatives (Arrow, 1963). However, two considerations have to be made on this subject.

1. A Condorcet consistent rule always presents smaller probabilities of the occurrence of a rank reversal in comparison with any Borda consistent rule (Moulin, 1988; Young, 1995). This is a strong argument in favor of a Condorcet consistent rule.
2. Young (1988, p. 1241) claims that the C-K-Y-L ranking procedure is the *"only plausible ranking procedure that is locally stable"*. Where *local stability* means that the ranking of alternatives does not change if only an interval of the full ranking is considered. It is interesting to note that this property was also studied by Jacquet-Lagrèze (1969), one of the first researchers in multi-criteria analysis, who called it the median procedure.

Other properties of the C-K-Y-L ranking procedure are the following (Young and Levenglick, 1978).

i. *Neutrality:* it does not depend on the name of any alternative, all alternatives are equally treated.
ii. *Unanimity* (sometimes called *Pareto Optimality*): if all criteria prefer alternative *a* to alternative *b* than *b* should not be chosen.
iii. *Monotonicity:* if alternative *a* is chosen in any pair-wise comparison and only the criterion scores of *a* are improved, then *a* should be still the winning alternative. Monotonicity is an essential property in a SMCE framework since dominated alternatives are not advised to be deleted from the analysis.
iv. *Reinforcement:* if the set *A* of alternatives is ranked by 2 subsets G_1 and G_2 of the criteria set *G*, such that the ranking is the same for both G_1 and G_2, then $G_1 \cup G_2 = G$ should still supply the same ranking. This general consistency requirement is very important in a multi-criterion framework where to test robustness

of results, one may wish to apply the criteria belonging to each single dimension first and then pool them in the general model.

It has to be noted that the C-K-Y-L ranking procedure is the only Condorcet consistent rule which holds the reinforcement property and as noted by Arrow and Raynaud, reinforcement *"… has definite ethical content and is therefore relevant to welfare economics and political science."* (Arrow and Raynaud, 1986, p. 96). Given that Arrow and Raynaud deal with the *"industrial outranking problem"* relevant for business people they do think that in this framework, reinforcement is less important that independence of irrelevant alternatives. On the contrary, in the framework of public policy, dealt with here, reinforcement becomes a decisive argument in favour of the C-K-Y-L ranking procedure

Although as one can see, the theoretical characterization of the C-K-Y-L ranking procedure is not that easy, the algorithm *per se* indeed is very simple. The maximum likelihood ranking of alternatives, in a multi-criterion framework, is the ranking supported by the maximum number of criteria for each pair-wise comparison, summed over all pairs of alternatives. More formally, the C-K-Y-L ranking procedure can be adapted to a multi-criterion framework as follows.

All the $N(N-1)$ pair-wise comparisons compose the matrix E, where let's remember that $e_{jk} + e_{kj} = 1$, with $j \neq k$. Let's call R the set of all the $N!$ possible complete rankings, of alternatives, $R=\{r_s\}$, $s=1, 2,..., N!$. For each r_s, let's compute the corresponding score φ_s as the sum of all e_{jk} over all the $\binom{N}{2}$ pairs jk of alternatives, such that a_j is preferred to a_k in the ranking r_s. More formally, let us denote $a_j >^s a_k$ the fact that a_j is preferred to a_k in the ranking r_s, then

$$\varphi_s = \Sigma_{j=1}^{N} \Sigma_{k=1}^{N} e_{jk}^{s}, \qquad (26)$$

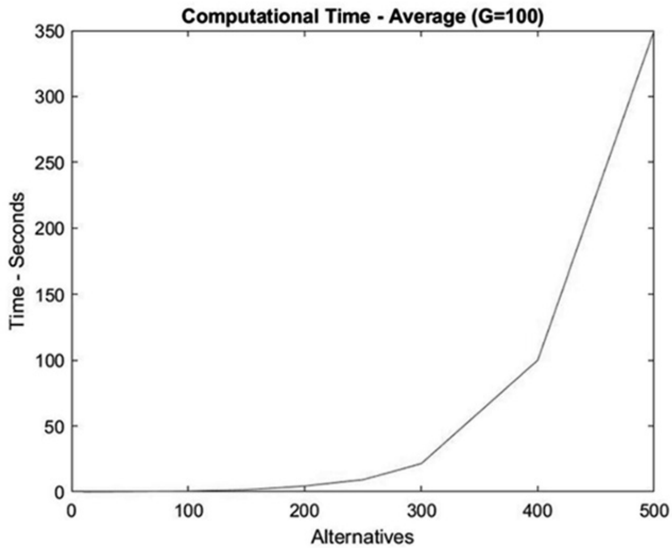
where $j \neq k$ and $e_{jk}^{s} = \begin{cases} e_{jk} & \text{if } a_j >^s a_k \\ 0 & \text{otherwise} \end{cases}$.

The final ranking is r_t, $t \in \{1, 2,..., N!\}$, such that

$$\varphi_t = \max \varphi_s,\ s = 1, 2,\ldots, N! \qquad (27)$$

The computational problem is a clear drawback of this approach. One should note that the number of permutations can easily become unmanageable; for example, when 10 alternatives are present, it is 10! = 3,628,800. A numerical algorithm solving this computational drawback in an efficient way has been developed recently (Azzini and Munda, 2020). In 1000 simulations, keeping constant the number of criteria (G=100), the average computational time is about 1 second till 100 alternatives and it reaches a maximum of 350 seconds for 500 alternatives (see Figure 3).

Figure 3: Computational time needed by increasing the number of alternatives



2.5. Introducing the minority principle: a borda approach

At this point, we have to refer to the normative tradition in political philosophy, which has also an influence in modern social choice (Moulin, 1981) and public policy (Mueller, 1978). The basic idea is that any coalition controlling more than 50% of votes may be converted in an actual dictator. As a consequence, the *"remedy to the tyranny of the majority is the minority principle, requiring that all coalitions, however small, should be given some fraction of the decision power. One measure of this power is the ability to veto certain subsets of outcomes..."* (Moulin, 1988, p. 272). The introduction of a veto power in a multi-criterion framework can be further justified in the light of the so-called *"prudence"* axiom (Arrow and Raynaud, 1986, p. 95), whose main idea is that it is not prudent to accept alternatives whose degree of conflictuality is too high (and thus the decision taken might be very vulnerable).[3] The point is then how can we implement this idea of veto power in a multi-criterion framework?

Historically, the first attempt was done by Roy (1985, 1996) in the so-called ELECTRE methods. Basically, Roy proposed that for any pair of alternatives one should look at the majority principle expressed as a concordance index and to the minority one in the form of the discordance index. The discordance index is calculated according to the intensity of

[3] It has to be noted that to mitigate the vulnerability of the C-K-Y-L ranking procedure is very important since this is one of the main criticism against this method.

preference any single criterion has against the concordance coalition. This means that on each single criterion a veto threshold needs to be defined.

In my opinion, the implementation of the veto power in a SMCE framework needs three desirable properties:

1. To be independent of arbitrary ad hoc thresholds.
2. To consider the global opposition against the final ranking and not against a pair of alternatives, or any specific possible ranking.
3. No specific intensity of preference should be considered (if one combines a weight with a veto threshold on each single criterion, the resulting concept of criterion importance depends on the intensity of preference too, this means that probably weights cannot anymore considered as importance coefficients).

It is interesting to note that the approach fitting these requirements can again be found in classical social choice and in particular, this time in the Borda's approach. The Borda rule is normally used to find a Borda winner, where the winner is the alternative which receives the highest score in favour (an alternative receives *N-1* points if ranks first and so on till *0* score if it ranks last on a given criterion). In the same way, a *Borda loser* can be defined as the alternative which receives the highest score against (where *N-1* points are assigned to the last alternative in the ranking and so on till *0* points are given to the alternative which ranks first).

Formally the procedure I am proposing can be described as follows by taking inspiration from the concept of frequency matrices (Hinloopen et al., 1983, Matarazzo, 1988). Let's call F the matrix where any element f_{ij} means that a given criterion g_m scores alternative a_j in the *i-th* ordinal position. Now it is possible to define the $N \times N$ matrix $\mathbf{\Phi}$ where any element φ_{ij} represents the summation of the weights of criteria which score alternative j at the *i-th* position; that is

$$\varphi_{ij} = \Sigma_{m \in G_i} w_m \tag{28}$$

$$\textit{where } G_i = \{g_m : g_m(a_j) = f_{ij}\} \quad \textit{with} \quad G_i \subset G \tag{29}$$

I = 1, 2,…, N and j = 1, 2,…, N.

It is obviously:

$$\Sigma_{i=1}^{N} \varphi_{ij} = 1 \quad \forall\, a_j \in A \qquad \text{and} \tag{30}$$

$$\Sigma_{j=1}^{N} \varphi_{ij} = 1 \quad \textit{with } j = 1, 2, ..., N \tag{31}$$

Now for any alternative a_j let's apply the Borda rule in search of the Borda looser, that is

$$B(a_j) = \Sigma_{i=1}^{N} (\varphi_{ij} \times b_i), \;\; b_i = N-1, N-2, ..., 0 \quad \textit{with } i = N, N-1, ..., 1. \tag{32}$$

The vetoed alternative $\bar{a}_j$ is the Borda looser, i.e. the a_j for which $B(a_j)$ = max.

One should note that by means of this procedure, weights are never combined with intensities of preference and no *ad hoc* parameter is needed. Consistently with the Borda approach only one alternative, considered the one with the highest opposition, is selected as alternative to be vetoed. It has to be remembered that the Borda procedure respects all the properties of the C-K-Y-L one, except local stability. This is the main reason why Borda consistent rules are more adequate for the selection of one alternative only and not for the generation of rankings.

Finally, a question to be answered is: do Borda and Condorcet rules normally lead to different solutions? One can in fact think that the divergence of solutions is a very special case and thus the value added of introducing the Borda looser is very limited. This question has been answered by Fishburn (1973b) and Moulin (1988), who proved that Condorcet consistent rules and Borda voting rules are deeply different in nature and consequently it is useful to combine them in a complementary way.

3. Conclusion

> "We live in a world of contradiction and paradox, a fact of which perhaps the most fundamental illustration is this: that the existence of a problem of knowledge depends on the future being different from the past, while the possibility of the solution of the problem depends on the future being like the past." (Knight, 1921, p. 313)

This article has illustrated how SMCE can help in dealing with three different types of sustainability related policy issues: 1) epistemological uncertainty 2) complexity 3) mathematical manipulation rules of relevant information. In summary, we can conclude that:

In sustainability policies evaluation and assessment, key questions to be answered are sustainability of *what and whom*? Consequently, sustainability problems are multidimensional in nature and characterised by a high degree of conflict, uncertainty and complexity.

Complexity arises when something is difficult to understand and impossible to analyse by using simple frameworks.

To reach a ranking of policy options, there is a prior need to decide *what is important* for different social actors as well as *what is relevant* for the representation of the real-world entity described in the model.

In operational terms, social *multi-criteria evaluation (SMCE) techniques* have the potential to take into account conflictual, multidimensional and uncertain properties of policy decisions. SMCE can therefore provide insights into the nature of conflicts and complexity and facilitate the process of reaching political compromises by explaining divergent values and increasing the transparency of the decision process.

From a theoretical perspective, we can conclude that commensurability, a necessary condition for strong comparability, can be implemented by means of "set commensurability" and "rod commensurability"; both of them are not of a general applicability. Different metrics are linked to different objectives and values. To use only one measurement unit for incorporating a plurality of objectives and values, implies reductionism necessarily, therefore weak comparability grounded on incommensurability, can be implemented by using social multi-criteria evaluation.

No mathematical model, even if legitimate in its own terms, can be sufficient for a complete analysis of the reflexive properties of a real-world problem. Results of a real-world policy exercise depend strongly on the way a given problem is structured during the evaluation process obviously, but mathematical models play a very important role: the one of guaranteeing consistency between assumptions used and results obtained. This implies to take into account the technical *uncertainties properly*; consequently, this article has presented a mathematical aggregation convention useful for the solution of the so-called discrete multi-criterion problem in a SMCE context. This mathematical aggregation procedure is a "reasonable" approach based on theoretical and empirical grounds, all of them made explicit and thus easy to evaluate in relation with a particular use.

Throughout the whole pair-wise comparison step, it is guaranteed that ordinal, crisp, stochastic and fuzzy criterion scores are tackled equivalently. To deal with the lack of stability of the pseudo-order structure, valued preference relations modelled by means of fuzzy preference relations are introduced. Weights are never combined with intensities of preference, as a consequence the theoretical guarantee they are importance coefficients exists. The pair-wise comparisons can be synthesised in an outranking matrix, which can be interpreted as a voting matrix. The information contained in the voting matrix is exploited to rank all alternatives in a complete pre-order by using a Condorcet consistent rule. Consistently with the normative tradition in political philosophy and following the prudence axiom, the minority principle is introduced by means of a veto power, grounded on the original Borda approach implemented by using a frequency matrix.

References

Arrow, K.J. (1963) *Social Choice and Individual Values*, 2d edition. New York, Wiley.

Arrow, K.J.; Raynaud H. (1986) *Social Choice and Multicriterion Decision Making*. Cambridge, M.I.T. Press.

Azzini, I; and Munda, G. (2020) A new approach for identifying the Kemeny Median Ranking. *European Journal of Operational Research, 281*, 388-401.

Barbier, E.B. (1987) The concept of sustainable economic development. *Environmental Conservation*, *14*(2), 1-110.

Borda, J.C. de (1784) Mémoire sur les élections au scrutiny. In Histoire de l' Académie Royale des Sciences, Paris.

D. (1986) Some remarks on the notion of compensation in MCDM. *European Journal of Operational Research*, *26*, 150-160.

Bouyssou, D.; Vansnick, J.C. (1986) Noncompensatory and generalized noncompensatory preference structures. *Theory and Decision*, *21*, 251-266. Chang, R. (ed.) (1997) *Incommensurability, Incomparability, and Practical Reason*. Cambridge, Harvard University Press.

Condorcet, M. de (1785) Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Probabilité des Voix. Paris, De l' Imprimerie Royale.

Copeland, A. H. (1951) *A reasonable social welfare function*, University of Michigan, mimeo.

Etxano, I.; Villalba-Eguiluz, U. (2021) Twenty-five years of social multi-criteria evaluation (SMCE) in the search for sustainability: Analysis of case studies. *Ecological Economics, 188*, 107-131.

Figueira, J.; Greco, S.; Ehrgott, M. (eds.) (2016) *Multiple-Criteria Decision Analysis. State of the Art Surveys*. New York, Springer International Series in Operations Research and Management Science.

Fishburn, P.C. (1970) Utility theory with inexact preferences and degrees of preference. *Synthese*, *21*, 204-222.

Fishburn, P.C. (1973a) Binary choice probabilities: on the varieties of stochastic transitivity. *Journal of Mathematical psychology*, *10*, 327-352.

Fishburn, P.C. (1973b) *The Theory of Social Choice*. Princeton, University Press.

Funtowicz, S.; Martinez-Alier, J.; Munda, G.; Ravetz, J. (1999) *Information tools for environmental policy under conditions of complexity (Report* Environmental Issues Series, No. 9), Luxembourg, Office for Official Publications of the European Communities.

Funtowicz, S.O.; Ravetz, J. R. (1991) A new scientific methodology for global environmental issues. In R. Costanza (ed.) *Ecological Economics*, New York, Columbia, pp. 137-152.

Geach, P. T. (1956) Good and Evil. In Philippa (ed.) *Theories of Ethics*. Oxford, University Press, pp. 64-73.

Gell-Mann, M. (1994) The Quark and the Jaguar: Adventures in the Simple and the Complex. New York, W.H. Freeman.

Giampietro, M.; Mayumi K.; Munda G. (2006) Integrated assessment and energy analysis: quality assurance in multi-criteria analyses of sustainability. *Energy, 31*(1), 59-86.

Giampietro, M.; Mayumi, K. (2018) Unraveling the Complexity of the Jevons Paradox: The Link Between Innovation, Efficiency, and Sustainability. *Frontiers in Energy Research*, *6*. https://doi.org/10.3389/fenrg.2018.00026

Giampietro, M.; Mayumi, K. (2000) Multiple-scale integrated assessment of societal metabolism: integrating biophysical and economic representations across scales. *Population and Environment*, *22* (2), 155-210.

Giampietro, M.; Mayumi, K. (2000) Multiple-scale integrated assessment of societal metabolism: Introducing the approach. *Population and Environment*, *22*(2), 109-154.

Giampietro, M.; Mayumi, K.; Sorman, A.H. (2012) *The Metabolic Pattern of Societies: Where Economists Fall Short.* Routledge.

Hinloopen, E.; Nijkamp, P.; Rietveld, P. (1983) Qualitative discrete multiple criteria choice models in regional planning. *Regional Science and Urban Economics*, 77-102.

Jacquet-Lagrèze, E. (1969) L'agrégation des opinions individuelles. Informatiques et Sciences Humaines, 4.

Kacprzyk, J.; Roubens, M. (Eds.). (1988) *Non-Conventional Preference Relations in Decision Making.* Heidelberg, Springer-Verlag. https://doi.org/10.1007/978-3-642-51711-2

Keeney, R.; Raiffa, H. (1976) Decision with Multiple Objectives: Preferences and Value Trade-offs. New York, Wiley.

Kemeny, J. G. (1959) Mathematics without Numbers. *Daedalus*, *88*(4), 577-591.

Knight, F. (1921) Risk, Uncertainty and Profit. Cambridge, Che Riverside Press.

Köhler, G. (1978) *Choix multicritère et analyse algébrique des données ordinales*, Ph.D. Thesis, Université Scientifique et Médicale de Grenoble, France.

Luce, R.D. (1956) Semiorders and a theory of utility discrimination. *Econometrica*, *24*, 178-191.Markowitz; H.M. (1989). Mean-variance Analysis in Portfolio Choice and Capital Markets. Oxford, Basil-Blackwell.

Martel, J.M.; Zaras, K. (1995) Stochastic dominance in multicriteria analysis under risk. *Theory and Decision*, *35*(1), 31-49.

Martinez-Alier, J.; Munda, G.; O'Neill, J. (1998) Weak comparability of values as a foundation for ecological economics. *Ecological Economics*, *26*, 277-286.

Matarazzo, B. (1988) Preference ranking global frequencies in multicriteria analysis (PRAGMA). *European Journal of Operational Research*, *36*, 36-49.

Matarazzo, B.; Munda, G. (2001) New approaches for the comparison of L-R fuzzy numbers: A theoretical and operational analysis. *Fuzzy Sets and Systems*, *118* (3*)*, 407-418.

Moulin, H. (1981) The proportional veto principle. *Review of Economic Studies*, *48*, 407-416.

Moulin, H. (1985) From social welfare orderings to acyclic aggregation of preferences, *Mathematical Social Sciences*, *9*, 1-17.

Moulin, H. (1988) *Axioms of Co-operative Decision Making. Econometric Society Monographs*, Cambridge, Cambridge University Press.

Mueller, D. (1978) Voting by veto. *Journal of Public Economics*, 10, 57-75.

Munda, G (2009) A conflict analysis approach for illuminating sustainability distributional issues. *European Journal of Operational Research*, *194*, 307-322

Munda, G (2022) Qualitative reasoning or quantitative aggregation rules for impact assessment of policy options? A multiple criteria framework. *Quality & Quantity. International Journal of Methodology,* 56, 3259-3277.

Munda, G. (1995) Multicriteria Evaluation in a Fuzzy Environment. In *Contributions to Economics Series,* Heidelberg, Physica-Verlag, pp. 131-148.

Munda, G. (2004) Social multi-criteria evaluation (SMCE): Methodological foundations and operational consequences. *European Journal of Operational Research*, *158* (3), 662- 677.

Munda, G. (2008) *Social Multi-criteria Evaluation for a Sustainable Economy*. Heidelberg, Springer-Verlag,

Munda, G. (2012). Intensity of preference and related uncertainty in non-compensatory aggregation rules. *Theory and Decision*, *73*(4), 649-669. https://doi.org/10.1007/s11238-012-9317-4

Munda, G. (2016) Beyond welfare economics: Some methodological issues. *Journal of Economic Methodology*, *23* (2), 185-202Norgaard, R. B. (1994) *Development Betrayed*. London, Routledge.

O'Neill, J. (1993) *Ecology, Policy and Politics*. London, Routledge.

Ozturk, M.; Tsoukias, A.; Vincke, Ph. (2005) Preference modelling. In Figueira, J.; Greco, S. and Ehrgott, M. (eds.) *Multiple-criteria Decision Analysis. State of the Art Surveys*. New York, Springer International Series in Operations Research and Management Science, pp. 27-71.

Paelinck J.H.P. (1978) Qualiflex, a flexible multiple criteria method. *Economic Letters*, 3, 193-197.

Podinovskii, V.V. (1994) Criteria importance theory. *Mathematical Social Sciences*, *27*, 237- 252.

Poincaré, H. (1935) *La Valeur de la Science*. Flammarion.

Prigogine, I.; Stengers, I. (1981) *Order Out of Chaos*. New York, Bantam Books.

Ragade, R.K. and Gupta, M.M. (1977) Fuzzy set theory: Introduction. In M.M. Gupta, G.N; Saridis and B.R. Gaines, *Fuzzy Automata and Decision Processes*. Amsterdam, North-Holland, pp. 105-131

Roberts, F. S. (1979) Measurement theory with applications to decision making, utility and the social sciences, London, Addison-Wesley.

Rosen, R. (1985) Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations. New York, Pergamon Press.

Rosen, R. (1991) Life Itself: A Comprehensive Inquiry into the Nature, Origin and Fabrication of Life. New York, Columbia University Press.

Roubens; M.; Vincke, Ph. (1985) *Preference Modelling*. Heidelberg, Springer-Verlag,

Roy, B. (1996) Multicriteria Methodology for Decision Analysis. Kluwer, Dordrecht.Roy; B. (1985) Méthodologie Multicritere d' Aide à la Decision. *Economica*, Paris.

Saltelli, A.; Tarantola S.; Campolongo, F. and Ratto, M. (2004) *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. New York, John Wiley & Sons publishers,

Saltelli, A.; Van der Sluijs, J.; Guimarães Pereira Â.; Funtowiz, S.O. (2013) What do I make of your Latinorum? Sensitivity auditing of mathematical modelling. *International Journal Foresight and Innovation Policy*, *9*(2/3/4), 213-234.

Sen, Amartya. (2009) *The Idea of Justice*. The Belknap Press of Harvard University Press, Cambridge, Massachusetts.

Simon, H. A. (1962) *The architecture of complexity*. Proc. Amer. Philos. Soc. *106*, 467-482.

Simpson, P. (1969) On defining areas of voter choice: Professor Tullock on stable voting. *Quarterly Journal of Economics*, 83, 478-490.

Truchon, M. (1995) Voting games and acyclic collective choice rules. *Mathematical Social Scien*ces, *25*, 165-179.

Vansnick, J. C. (1990) Measurement theory and decision aid. In Bana e Costa C.A. (ed.), *Readings in Multiple Criteria Decision Aid*. Springer-Verlag, Berlin, pp. 81-100.

Vansnick, J.C. (1986) On the problem of weights in multiple criteria decision making (the non-compensatory approach). *European Journal of Operational Research*, *24*, 288-294.

Vidu, L. (2002) Majority cycles in a multi-dimensional setting. *Economic Theory*, *20*, 373-386.

Weber, J. (2002) How many voters are needed for paradoxes? *Economic Theory*, *20*, 341-355.

Young, H.P. (1988) Condorcet's theory of voting. *American Political Science Review*, *82* (4), 1231-1244.

Young, H.P. (1995) Optimal voting rules. *Journal of Economic Perspectives*, *9*, 51-64.

Young, H.P.; Levenglick, A. (1978) A consistent extension of Condorcet's election principle, *SIAM Journal of Applied Mathematics*, 35, 285-300.