A flexible multicriteria decision-making methodology to support the strategic management of Science, Technology and Innovation research funding programs

R.O. Parreiras, I. Kokshenev, M.O.M. Carvalho, A.C.M. Willer, C.F. DellezzopollesJr., D.B. NacifJr., J.A. Santana

PII: S0377-2217(18)30602-7
DOI: 10.1016/j.ejor.2018.06.050
Reference: EOR 15235

To appear in: *European Journal of Operational Research*

Received date: 12 April 2017
Revised date: 11 June 2018
Accepted date: 28 June 2018

Highlights

- Funding decisions can benefit from multi-attribute utility theory.
- Value Focused Thinking can reinforce the strategic alignment of funding programs.
- We suggest that agency personnel should be in the role of decision-makers.
- We suggest that attributes of value functions should be based on factual information.
- Factual assessments provided by reviewers serve as data for program evaluation.

A flexible multicriteria decision-making methodology to support the strategic management of Science, Technology and Innovation research funding programs☆

R. O. Parreiras[a,b,c], I. Kokshenev[a,b,*], M. O. M. Carvalho[d], A. C. M. Willer[d], C. F. Dellezzopolles Jr.[d], D. B. Nacif Jr.[d], J. A. Santana[d]

[a]*Typi Ltda., Av. Rio Branco 404/307, Florianópolis/SC, Brazil 88015–200*
[b]*Advanced System Optimization Technologies Ltda., Rua Dom Lucio Antunes, 855/302, Belo Horizonte/MG, Brazil 30535-630*
[c]*Post-Graduate Program in Electrical Engineering, Pontifical Catholic University of Minas Gerais, Av. Dom José Gaspar, 500, Belo Horizonte/MG, Brazil 30535-610*
[d]*National Council of Technological and Scientific Development (CNPq), SHIS QI 1 Conjunto B, Bloco C, sl. 103, Brasília/DF 71605–001*

Abstract

Research funding programs are a policy instrument utilized by governments to influence the innovation process. They are usually elaborated, launched and managed by research funding agencies. In order to select the most adequate research projects, agencies often rely on the peer review process.

This paper introduces a methodology to support funding decisions based on the peer review process. The methodology involves the use of a multicriteria decision model to support the assessment, evaluation, prioritization and selection of applications, under a multi-step decision-making process, which fits into a strategic management cycle within the agency. The Multiattribute Value Theory, being considered under a Value Focused Thinking approach, provides a basis for the construction of the multicriteria decision model. The good practices in peer review and also a logical framework for program management are considered by the methodology.

A pilot study, presented in the paper, involved a retrospective implementation of a peer review process in the context of a program launched by the Ministry for Science, Technology, Innovations and Communications and the National Council of Technological and Scientific Development, in Brazil. The methodology allowed a clear distinction of roles. The agency staff in the role of decision-makers was responsible for making value judgments on behalf of the agency. The experts, in the role of committee members and ad hoc reviewers, contributed with their expertise by providing objective assessments. Such assessments served as a basis for evaluating the applications, characterizing the possible portfolios, and can be considered as data in future program evaluation studies.

Keywords: Decision processes, OR in government, Multiple criteria analysis, Peer review process, Strategic management

∗Corresponding author: **illya.kokshenev@gmail.com** (I. Kokshenev)

1. Introduction

Policies for Science, Technology and Innovation can complement and reinforce the long-term strategies for social and economic development of a country (OECD, 2012). Research funding agencies contribute to the implementation of the government policies by promoting and supporting research activities. As part of their efforts, agencies usually launch research funding programs (hereinafter, simply referred to as programs) (European Science Foundation, 2011; Guthrie et al., 2013). These programs can pursue one or more objectives, which need to be aligned with the public policies, as well as with the current strategy and mission of the agency (Edquist & Zabala-Iturriagagoitia, 2012; Joyce, 2015).

In order to evaluate research projects, agencies often rely on the peer review process (PRP). It can be implemented in many different ways (August & Muraskin, 1999; Abdoul et al., 2012). However, it is commonly implemented in such a way that every application is evaluated by one or more ad hoc reviewers and, subsequently, by the members of a committee. Despite being the current dominant practice, PRPs have been widely criticized (Marsh et al., 2008; Abramo et al., 2009; Abramo & D'Angelo, 2011; Ebadi & Schiffauerova, 2016). The criticisms are mainly related to the dependency of the PRP outcomes on the selection of the individuals to be involved in the review process. It is a well known fact that the decision quality may degrade because of conflicts of interest in the committee or due to knowledge and experience limitations of the reviewers. Another typical weakness in most PRPs is associated with the fact that each reviewer often handles only a few submissions. This leads each of them to use different references for evaluating the applications (Marsh et al., 2008), which may introduce more bias into the process, instead of the promised impartiality.

Among the possible measures to overcome such problems, it is a common practice to consider multiple reviews for each application and to rely on the final consensual recommendation provided by the committee. However, these measures may not completely prevent the influence of a biased committee or of the cognitive biases, which are inherent in human mind (Kahneman, 2011). In particular, they are not suitable to deal with some collateral effects of the use of highly subjective criteria to evaluate the applications. The excessive subjectivity allows interferences of individual or group interests in the judgments. It also may hinder the ability of the agency staff to discern whether it is necessary to intervene in the review process to prevent deviations from the program objectives. Such occurrences have negative impact for the researchers

☆This work was supported by UNESCO Project 914BRZ2002 [contract number SA-83/2015] and the National Council of Technological and Scientific Development (CNPq) of Brazil [grants number 311473/2014-8 and 300707/2016-9].

*Corresponding author
Email address: illya.kokshenev@gmail.com (I. Kokshenev)

whose projects are evaluated unfairly, as well as for the reputation of the agency itself. Beyond that, these occurrences can jeopardize an effective implementation of the program (Durlak & DuPre, 2008).

In this paper, we propose a methodology, named DEMUCTI (which stands for *Decisão Multicritério em Ciência, Tecnologia e Inovação*), to support the funding decision. It aids the assessment, evaluation, prioritization and selection of the most adequate eligible applications, under a multi-step decision-making process, which fits into a strategic management process (SMP) within the agency (Costa, 2014).

DEMUCTI was developed to meet a demand of the Brazilian agency — the National Council for Scientific and Technological Development (CNPq). The main requirements underlying the methodology scope were that it should be flexible enough to be applied in the context of diverse programs; fit the needs of different PRP implementations; incorporate multiple criteria of a quantitative or a qualitative nature; and enable the agency-wide institutionalization.

In order to meet those requirements, the methodology makes use of a multicriteria decision model based on the Multiattribute Value Theory (MAVT) (Keeney & Raiffa, 1976), being considered under a Value Focused Thinking (VFT) approach (Keeney, 1992). As discussed in the paper, this endows the funding decision with more transparency and traceability, among with other important properties. DEMUCTI also considers a logical framework for program management and is compatible with the guidelines for good practices of peer review, as recommended in (European Science Foundation, 2011).

The remainder of the paper is organized as follows: Section 2 provides an overview of the methodology and its scope. In Section 3, the main steps of the DEMUCTI are described. Section 4 presents the case study on which six steps of the methodology were pilot-tested. Section 5 discusses the results and draws some conclusions.

2. Overview of the methodology

2.1. Background

As mentioned before, programs represent one of the policy instruments that a government can use in order to influence the development and diffusion of innovations. Within the typology of instruments presented in (Borrás & Edquist, 2013), which includes the regulatory, financial and non-coercive[1] types of instruments, programs can be classified as a financial instrument with a supply-push[2] dynamic to strengthen innovation.

Usually, governments delegate the implementation of programs to agencies, which may handle one or more simultaneous programs, each one having its own objectives,

[1] based on the establishment of voluntary partnerships among organizations
[2] provides incentives to the supply-side

target applicants, scope and funding modes (Papon, 1999). Besides satisfying certain legal constraints, the funding decision needs to be coherent with the program objectives (in order to ensure a program implementation aligned with these objectives), and with the strategic objectives of the agency. In order to more effectively promote change toward some desired directions, the governmental organizations can implement the processes of strategic planning and strategic management (Poister, 2010; Joyce, 2015). Together, such processes allow organizations to plan and implement their actions aimed at achieving strategic objectives. Although these processes have not yet become a well-established practice in the public sector worldwide, several governmental agencies and enterprises have proven to be capable of implementing them (Joyce, 2015).

Public bodies are required to be not only accountable, but also transparent about their actions. Thus, funding decisions are usually made according to some standard procedures that are previously divulged. In order to ensure that the decisions are fair, credible and aim at the most meritorious research projects, agencies count on the peer review. Such factors as the program budget, the research areas, the levels of interaction among disciplines being required or encouraged, the efficiency of the end-to-end evaluation process need to be considered in the calibration of a fit-for-purpose PRP (European Science Foundation, 2011).

2.2. Main contributions of the methodology

Given the background outlined above, we indicate the main contributions brought by DEMUCTI:

- DEMUCTI supports the most common configurations of calls for proposals and PRPs. Both the one-round and the two-round modes of calls are supported. PRPs can have one or more stages provided that, if more than one review is considered per application, review panels are implemented for reaching a consensus on the assessments, following the good practice recommendations (European Science Foundation, 2011). The "Right to Reply" of applicants can be implemented. Confidentiality can be managed in accordance with the double-blind, the single-blind or the open review systems (European Science Foundation, 2011).

- The methodology recommends a clear separation of tasks among the decision-makers, who make value judgments on behalf of the agency, and the experts (reviewers and committee members), who provide objective (factual) judgments based on their expertise. This allows tasks to be distributed in such a way that individuals with technical knowledge are not required to make value judgments on the applications, unless this is desired (Keeney, 1992).

- DEMUCTI improves the strategic alignment of the programs by orienting the definition of the objectives of the funding decision so that they are aligned with the program objectives, which, in turn, are aligned with the strategic objectives of the agency. Besides, DEMUCTI is adherent to the SMP cycle: although SMP is out of the DEMUCTI scope, the DEMUCTI cycle fits into the SMP cycle.

2.3. Participants and roles

Within DEMUCTI, the funding decision is handled as a cooperative multi-step group decision-making process (Pedrycz et al., 2011). Three main roles are distinguished: the decision-makers, the committee members, and the ad hoc reviewers. Under a hierarchical structured environment (Pedrycz et al., 2011), more authority is delegated to the decision-makers, who have responsibility to provide value judgments and make the final decision on behalf of the agency. They count on some assistance from other members of the group, the experts, who are supposed to contribute to the overall understanding of the decision problem with their expertise. The experts can be in the role of committee members and, when the review requires special technical expertise, the committee can be supplemented by ad hoc reviewers.

The decision-maker role needs to be assigned to one or more agency staff members. Representatives of partner institutions and/or other stakeholders can also play the decision-maker role, jointly with agency staff members.

For a results-based program management (Pazvakavambwa & Steyn, 2014), all decision-makers are encouraged to consider the available knowledge and evidences on the effectiveness of previous implementations of the same or similar programs, in order to form value judgments based on sound reasoning. This information and knowledge being shared among the decision-makers creates a common ground for discussion, which makes it feasible to achieve a consensus in a cooperative fashion.

2.4. General workflow

DEMUCTI steps take place within a SMP cycle (Costa, 2014). As it can be seen in Figure 1, the DEMUCTI steps are preceded by the definition of strategic objectives[3] and are interspersed with the plan implementation and also with the management of the plan execution. The resulting workflow makes good use of the current strategic objectives to guide the design of programs and also the assessment, evaluation, ranking and selection of applications.

Step 7 feeds the DEMUCTI cycle by forming feedback loops to Step 1 and also to Step 2. In this way, it enables one to improve the subsequent funding decisions of the ongoing program. Step 7 also feeds into the SMP cycle, by forming multiple

[3]It is suggested (not mandatory) to employ VFT for the definition of strategic objectives. Alternatively, the approaches commonly utilized with this purpose in the organizations can be applied (Costa, 2014; Steiss & Dekker, 2003).

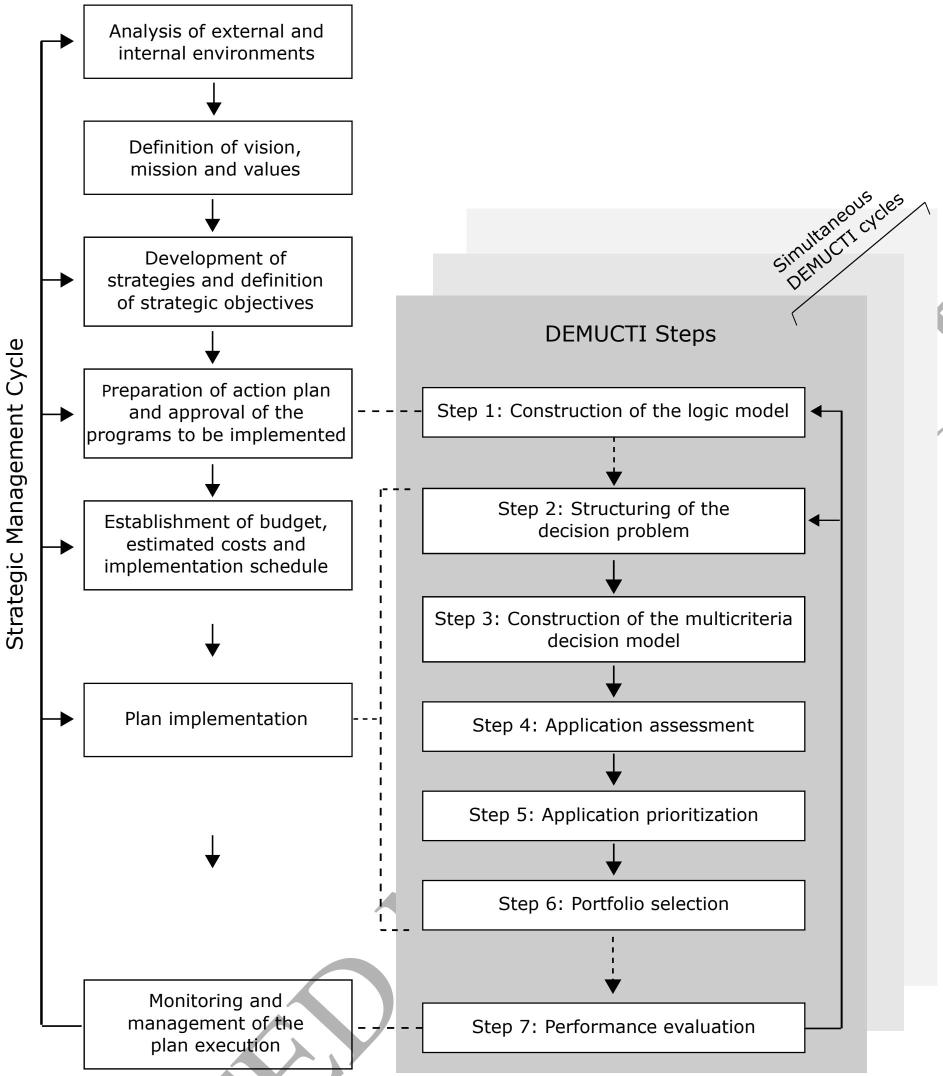

Figure 1: Steps of DEMUCTI cycle as part of a strategic management cycle

possible feedback loops. These loops are supposed to inform the allocation of budgets, the proposition and/or choice of new instruments, the prioritization of subjects in subsequent programs and also internal/external environmental scanning. Further details on the seven steps of the workflow are presented in the next section.

3. Main steps of the methodology

3.1. Step 1: Construction of the logic model

Research funding programs aim at solving certain current or future problems, which can be associated with unfulfilled human needs, an unsolved social problem or a low performance of the innovation system in a certain niche of products, processes or services (Borrás & Edquist, 2013). Because such problems may not be completely solved through this kind of programs, they need to be translated into the problems of fulfilling some specific innovation deficiencies that can be directly affected through such type of policy instrument.

This approach yields two sets of program objectives. The objectives expressed in terms of products of Science, Technology and Innovation are called direct objectives. The ultimate objectives, which can be achieved (at least partially) through fulfilling the direct objectives, are called indirect objectives. The former are associated with the mid-term effects of program execution and, the latter, with the long term effects of program execution (Schalock & Thornton, 1988; Cohen & Franco, 2011).

The causal relations between direct and indirect objectives give rise to the logic model (Royse et al., 2009). The logic model of a program is a conceptual model that outlines the connections among its inputs, activities, outputs (products and services), outcomes (i.e., mid-term effects of the output availability) and impacts (i.e., long-term effects of the output availability). Within DEMUCTI, it is intended to capture information to frame the decision context. The logic model should be developed in Step 1, together with the program elaboration and, under a feedback loop from Step 7, it should be revised, together with the program rationale. Figure 2 shows a logic model template which was adapted from the one proposed in (Guinea et al., 2015) to be utilized within DEMUCTI. The six questions below serve as guidelines to build such a logic model:

- Q1: Who are the target applicants?
- Q2: What are the needed resources to implement the program?
- Q3: What are the activities which can be financially supported by the program?
- Q4: What are the outputs (products or services) expected from those activities?
- Q5: Who are the stakeholders to be benefited by the outputs?
- Q6: What are the intended outcomes and impacts and how are they related to the direct and indirect objectives of the program?

The answers to Q1–Q6 must be coherent with the assumptions made on how the program will succeed in delivering the intended outputs, outcomes and impacts. When Q6 is addressed, one has to ensure that the program objectives are aligned with the strategic objectives of the agency. In order to define adequate criteria for evaluating the applications and indicators for evaluating the program performance, the two questions below also need to be considered:

- Q7: What are the required features of the applications and the applicants to ensure a successful program implementation?

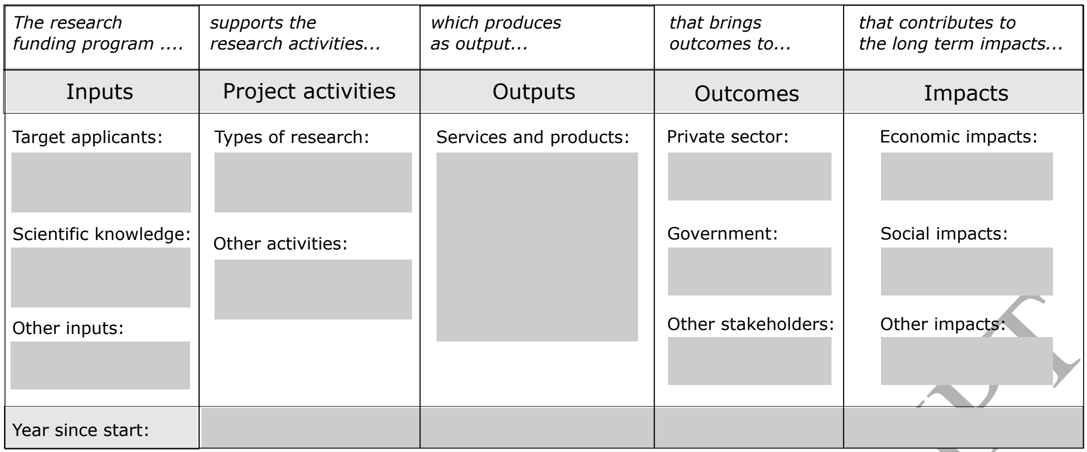

Figure 2: Logic model template for research funding programs

- Q8: What metrics should be considered to evaluate the degree of achievement of the program objectives (i.e. intended outcomes and impacts)?

The answer to Q7 provides a link between the logic model and the structure of the decision problem, which will be built in Step 2. This answer defines the features of the applications that are considered the most capable of delivering the intended outputs, outcomes and impacts. Finally, the answer to Q8 should comprise a choice of indicators to verify whether the program objectives are being achieved. The difficulty of measuring research impacts is widely recognized in the literature (Bornmann, 2012). According to (Schalock & Thornton, 1988; Cohen & Franco, 2011), if it is unfeasible to use indicators related to the ultimate impacts, some indicators intimately related to the intermediate results can be utilized instead.

3.2. Step 2: Structuring of the decision problem

The steps from 2 to 5 are associated with structuring and solving the decision problem. They can be performed in sequence, without reentering the SMP cycle (refer to Figure 1). Step 2, in particular, is dedicated to building the structure of the decision problem related to the assessment, evaluation, prioritization and choice of applications. This structure is composed of the following basic elements (Keeney, 1992):

- a set of fundamental objectives to be achieved through the decision;
- a value tree, which is a hierarchy of specific objectives unfolding from fundamental objectives, up to a set of criteria (each criterion corresponds to a leaf objective of the tree) to evaluate the applications;
- a set of measurable attributes of the applications which serves as a basis for the definition of value functions.

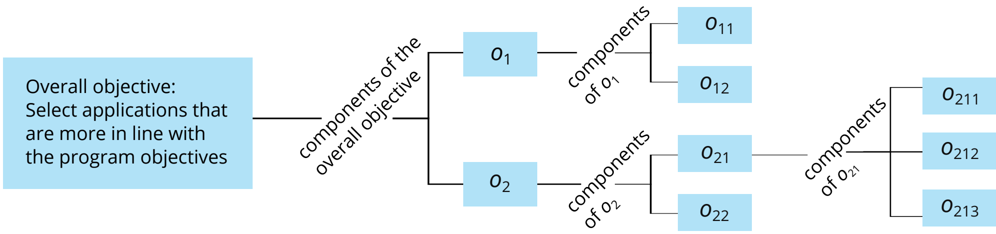

Figure 3: Example of a value tree with two fundamental objectives

According to the VFT main principles, in order to construct adequate value functions, first, it is necessary to define the fundamental objectives to be achieved through the decision (Keeney, 1992). In the context being considered, the fundamental objectives seek to prioritize the applications with more potential to bring the desired outputs, outcomes and impacts. In more general terms, they prioritize the applications that have certain fundamental characteristics (the ones specified in the answer to Q7).

Each fundamental objective can be unfolded into more specific components, in order to reduce the vagueness of the definitions of the corresponding fundamental characteristics and, thereby, avoid misinterpretations by the reviewers. A value tree similar to the one shown in Figure 3 accommodates all fundamental objectives and their respective components (Keeney, 1992). This structure, jointly with a set of measurable attributes, serves as a basis for the definition of value functions. For each objective with no descendants, a value function must be defined in terms of a selected attribute. The set of criteria needs to satisfy the following main conditions (Keeney & Raiffa, 1976; Keeney, 1992):

- they should cover just the requirements that are considered essential to ensure a successful execution of the program;
- they must avoid double counting of characteristics of the proposals;
- there must be mutual preference independence among them, which enables an independent treatment of each criterion in the analysis;
- each of them must be defined as a function of measurable attributes.

The attributes must provide means for expressing value judgments (Keeney, 1992). Additionally, DEMUCTI recommends that all attributes need to be based on information of a factual nature, being objective and as free of personal interpretations as possible (Simon, 1997). Either natural, proxy or constructed attributes (Keeney, 1992) can be based on information of this kind. For instance, the h-index of an author is a factual information that can serve as a proxy attribute (perhaps, not the most adequate one) for the citation impact of the work of each applicant.

In the absence of suitable quantitative indicators to compose the attributes (for instance, based on bibliometrics, altmetrics (Erdt et al., 2016), and others), nominal scales can be built by following different approaches. Both the goal attainment and the cumulative scale types (Colton & Covert, 2007) are applicable for this purpose. When the categories are defined, ambiguity should be kept to a minimum. This implies that one should avoid technical terms that are unknown by the evaluators and vague standards such as "low - high", "good - bad" (Colton & Covert, 2007). Designations of such kind are of little use as their meaning depends on local references (Durlak & DuPre, 2008).

Finally, the source of information for each attribute must be defined. It can be an automated data source (e.g., a data warehouse, an external index database, etc.) or a reviewer-dependent source (e.g., an application manuscript, a curriculum vitae, etc.). In the latter case, it is also necessary to elaborate the assessment questionnaire (to be addressed by the ad hoc reviewers and/or committee members) and, compatible with it, the application template for the submission of research projects or grant requests.

3.3. Step 3: Construction of the multicriteria decision model

In Step 3, the decision model needs to be built by the decision-makers. As it will be discussed next, they are asked to define single-attribute value functions for each leaf objective, as well as select and adjust aggregation operators for each intermediary objective of the value tree.

DEMUCTI considers the use of value functions from MAVT for modeling value judgments. The fact that such a framework allows considering value judgments over factual assessments helps to increase the impartiality of the decision-makers: they can be focused on valuing factual characteristics in the light of the program objectives, and not directly on the applications (Ferretti, 2016).

By assuming the representation of applications in terms of the attribute values $x \in \mathcal{X}$ in the domain $\mathcal{X}$, value judgments can be expressed by means of a value function $v : \mathcal{X} \to [0, V_{\max}]$. In order to ensure the compatibility of assessments expressed in different scales, references (anchors) are defined according to the "least preferable" and "most preferable" applications from the decision-maker standpoint. These anchors give meaning to the scores 0 and $V_{\max}$, respectively. While the typical MAVT implementations consider $V_{\max} = 1$, giving rise to the scores on the unit interval, DEMUCTI considers the scale $[0, 10]$. Hence, from now on, we assume $V_{\max} = 10$.

3.3.1. Construction of single-attribute value functions

The literature presents different techniques for the construction of single-attribute value functions (Beinat, 1997; Dyer, 2016). In order to shorten the process of preference elicitation, while keeping a satisfactory degree of versatility of the MAVT framework, DEMUCTI suggests applying the curve selection technique (Beinat, 1997) for

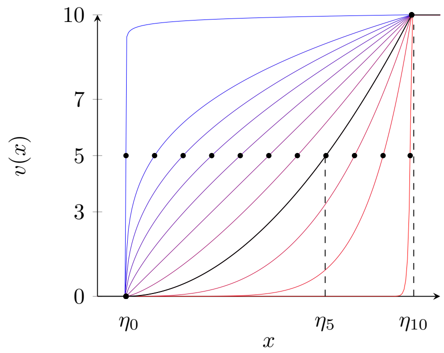

Figure 4: A family of value functions produced by different choices of the middle point anchor η_5

the attributes that are based on quantitative scales, by considering the unified value function model:

$$v(x) = 10 \cdot \rho(x)^{\gamma}, \tag{1}$$

where

$$\rho(x) = \min\left(0, \max\left(1, \frac{x - \eta_0}{\eta_{10} - \eta_0}\right)\right), \qquad \gamma = -1/\log_2 \rho(\eta_5). \tag{2}$$

In (2), η_0, η_5 and η_{10} are the anchor parameters that determine a value function with $v(\eta_0) = 0$, $v(\eta_5) = 5$ and $v(\eta_{10}) = 10$. While η_0 and η_{10} determine the limits and direction of the value scale, the choice of η_5, the middle point, controls the shape of the function. Different choices of η_5 within the limits (either $\eta_0 < \eta_5 < \eta_{10}$ or $\eta_{10} < \eta_5 < \eta_0$) result in a family of value functions that range from a threshold-like shape to a linear shape, as shown in Figure 4. In the interval between η_0 and η_{10}, this family of functions exhibits constant elasticity and generalizes log-interval transformations of the attribute scale. These properties are suitable for modeling the preferences on psychophysical (Roberts, 1985), as well as financial quantities (Ingersoll, 1987).

The three anchors of (1) must be defined in such a way that the decision-maker's value judgments are captured by the model. The value of η_5 can be defined by applying the Bisection Procedure (Beinat, 1997; Belton & Stewart, 2002). When it is permissible to express value judgments in terms of descriptive statistics (minimum, maximum, average, quartiles or percentiles), the value function can be constructed by calculating these quantities for a selected subset of applications. Such an approach enables one to handle the received set of proposals, when the complete evaluation model needs to be divulged prior to the proposals submission deadline. For instance, when quantity-based attributes are considered, one can achieve a suitable value function that is outliers-resistant, by setting $\eta_0 = 0$, while η_5 and η_{10} are set to the median and 95th percentile of a certain population, respectively. This approach also helps to obtain well distributed scores for highly skewed or power law distributions.

In case of attributes based on nominal scales, a discrete value function maps the categories of the scale to scores. In such a case, the direct rating of categories on a cardinal scale can be applied (Belton & Stewart, 2002). If the decision-makers face difficulties in applying this procedure, DEMUCTI suggests, in a similar way as in (Belton & Stewart, 2002), to consider the classical procedure of Analytic Hierarchy Process (AHP) (Saaty, 1980) (namely, the construction of a pairwise comparison matrix and its analysis based on the principal eigenvector) with an adequate adjustment for normalization and scaling as described in (Salo & Hämäläinen, 1997). Such adjustments allow AHP to be applied to obtain the scores for the nominal scales, and also the aggregation weights (Dyer, 2016).

The use of verbal scales as a measure for preference strengths in AHP has been questioned because their numerical counterparts may fail to reflect the verbal expressions. In order to overcome such weaknesses, DEMUCTI and also (Grafakos et al., 2010) recommend the use of a continuous numerical scale, so that the decision-makers could express the pairwise comparisons quantitatively.

3.3.2. Aggregation of value functions

Within DEMUCTI, the degree of achievement of each objective that has descendants is evaluated through an aggregation operation. In this way, scores produced by value functions of leaf objectives are aggregated bottom-up throughout the value tree, resulting in the final scores assigned to the applications.

In order to deal with some different attitudes that a decision-maker can have towards a choice problem (for instance: optimistic, pessimistic, allowing compensation among objectives, and others), two options of aggregation operations are provided: the weighted arithmetic mean (WAM) and a parameterized aggregation operator, named ordered weighted average (OWA) (Yager, 1988).

The most traditional approach corresponds to applying WAM, as given by the following expression:

$$v(x) = \sum_{i=1}^{m} w_i v_i(x), \tag{3}$$

where $v_1, \ldots, v_m$ correspond to m value functions being aggregated and the contribution degree of each value function is modulated through the weights $w_i \geq 0$, $i = 1, \ldots, m$ such that $\sum_{i=1}^{m} w_i = 1$ (Choo et al., 1999). By applying WAM, the decision-maker allows compensation among objectives (Pedrycz et al., 2011): each proposal is rewarded for their good performance in some objectives, despite of their poor performances in other objectives.

The approach based on applying OWA enables the decision-maker to control the degree of compensation among objectives, under a proper adjustment of its input parameters. The use of OWA yields the following aggregated score:

$$v(x) = \sum_{p=1}^{m} \omega_p v_{i_p}(x), \tag{4}$$

where i_p is the index of pth smallest value among $v_1, \ldots, v_m$ and ω_p, $p = 1, \ldots, m$ are the weights satisfying the conditions $\omega_p \geq 0$, and $\sum_{p=1}^{m} \omega_p = 1$.

In order to define the weights of WAM, DEMUCTI suggests applying the Swing Weights (Edwards & Barron, 1994), or the aforementioned tools from the AHP method (Saaty, 1980) in combination with the normalization and scaling adjustments described in (Salo & Hämäläinen, 1997). In the latter case, the decision-maker needs to provide the ratios of preference differences for the pairwise comparison of descendant objectives (Dyer, 2016).

DEMUCTI recommends defining the OWA weights indirectly by setting an integer pessimism-optimism index $q \in [1-m, m-1]$ as given by the following expression:

$$\omega_p = \begin{cases} \dfrac{1}{m-|q|}, \text{ if } 0 < p - q \leq m; \\ 0, \text{ otherwise;} \end{cases} \tag{5}$$

which can be seen as a simplified form and a specific case of the linguistic quantifier model, proposed in (Yager, 1995).

If $q = 1 - m$, the model reproduces the min operator. Such an aggregation is non-compensatory, in the sense that the high satisfaction of some criteria does not relieve the remaining ones from the requirement of being satisfied (Pedrycz et al., 2011). It is also considered a pessimistic or "risk-averse" approach, because it does not allow any proposal to assume a good score, if it fails in satisfying at least one criterion.

If $q = m - 1$, the model reproduces the max operator. Such an aggregation is extremely compensatory, in the sense that the high level of satisfaction of any criterion is sufficient, independently of which criterion is satisfied to a high level (Pedrycz et al., 2011). It is also considered an optimistic approach, being suitable to the cases when weaknesses of the proposals are not very critical and the decision-maker realizes ways by which their strengths contribute to a successful implementation of the program.

Intermediate pessimistic degrees can be realized by a proper setting of the OWA weights, as in the example shown in Table 1, where the OWA is applied to aggregate five criteria, that is $m = 5$.

The choice of aggregation operators plays a fundamental role in the decision, as the recommendations are derived from the mathematical models. The use of WAM may be attractive because of its intuitive appeal (Pedrycz et al., 2011). Certainly, the stakeholders have already utilized it to aggregate some kind of scores. Moreover, one could argue that the simplicity and popularity of the WAM facilitate communicating the evaluation rules in the public call. However, employing the OWA in the form of

Table 1: Example of OWA weights for different q values and $m = 5$

q	verbal rule	ω_1	ω_2	ω_3	ω_4	ω_5
-4	minimum score	1	0	0	0	0
-3	average of the two lowest scores	1/2	1/2	0	0	0
-2	average of the three lowest scores	1/3	1/3	1/3	0	0
-1	average of the four lowest scores	1/4	1/4	1/4	1/4	0
0	arithmetic average (unweighted)	1/5	1/5	1/5	1/5	1/5
1	average of the four highest scores	0	1/4	1/4	1/4	1/4
2	average of the three highest scores	0	0	1/3	1/3	1/3
3	average of the two highest scores	0	0	0	1/2	1/2
4	maximum score	0	0	0	0	1

(5) also yields simple rules, which can be communicated in a straightforward verbal manner, as demonstrated in Table 1. Hence, the communication effort should not affect the choice of an operator. Instead, the choice has to be based on the essence of the criteria and of their relations, in the light of the decision-maker attitude towards the decision.

3.4. Step 4: Application assessment

This step begins with the administration of the assessment questionnaire to the ad hoc reviewers. More than one review can be considered per application and, among them, there may be conflicting assessments. In the literature, it has been recognized that aggregation procedures, which in practice are commonly applied to combine assessments provided by different experts (e.g., the weighted sum of scores provided by reviewers) may be an inadequate and simplistic way to deal with discordances among knowledge based assessments (rather than preference based) (Dubois & Prade, 2012). According to recommendations of the (European Science Foundation, 2011), in order to deal with these disagreements, a committee has to consolidate the application assessments by providing their collective opinion under the rubric of consensus.

3.5. Step 5: Application prioritization

In Step 5, the aggregated scores are assigned to the applications by applying the multicriteria decision model. These scores are utilized to produce a ranking of applications. Depending on when the value judgments are elicited, Step 5 can be performed in accordance with either the *a priori* or the *a posteriori* approach for decision-making:

- an *a priori* decision situation is set up, if the decision-maker is required to construct the multicriteria decision model and to set all of its parameters before receiving the applications (Pedrycz et al., 2011). In this case, it is possible to increase the transparency of the funding decision by disclosing the model before the applicants prepare and submit their proposals. The value functions, defined a priori in terms of descriptive statistics of future submissions, can be used to ensure a more predictable behavior of the decision model, when little is known about distributions of attributes.

- an *a posteriori* decision situation is established, if the decision-maker is allowed to adjust the weights and/or other parameters of the model after receiving the applications (Pedrycz et al., 2011). This implies that some details of the evaluation rules should be disclosed only after Step 5 is performed. In such circumstances, prior to assigning scores to the applications, the decision-maker has an opportunity to simulate several decisions, while observing the properties of the resulting rankings.

The main differences between both approaches are associated with the implementation of steps from 3 to 5, as highlighted in Figure 3.5. The *a priori* decision allows a more transparent, but less flexible process, when it is compared with the *a posteriori* decision. Indeed, the latter allows the model parameters to be readjusted, for instance, to guarantee a more diversified portfolio, if the set of eligible applications are less diverse than it was expected. However, the *a posteriori* approach may be unacceptable for some programs, due to certain policy restrictions or the risk of legal implications.

The decision-maker is encouraged to perform a sensitivity analysis to check the robustness of results and the validity of generalizations that have been made in the elicitation process. In the *a priori* decision, the results of a sensitivity analysis may provide hints for improving the funding decision of a subsequent public call. In the *a posteriori* decision, however, the results can be applied to improve the evaluation model of the ongoing funding decision.

3.6. Step 6: Portfolio selection

The portfolio selection is based on the availability of budgetary resources, the amount to be allocated to each application, and the ranking of applications. In the most basic setup, the applications can be selected through the definition of a cutoff position in the ranking. Being a common practice in organizations, such procedure ensures Pareto efficient portfolios only if the funding amounts are equal per application. Despite the sub-optimal performance of this procedure for un-equal funding amounts, it still may be an admissible choice when the differences are negligibly small, compared with the average funding amount.

When the values of funding per application differ significantly from one another, it might be tempting to rank applications on the basis of their respective score/amount ratios, instead of their pure scores. That, in fact, would turn the aforementioned cutoff procedure into a greedy solution to the 0-1 knapsack problem. By treating this problem with proper optimization techniques, one can obtain Pareto efficient portfolios (from a theoretical standpoint). However, in practice, the results can be misleading,

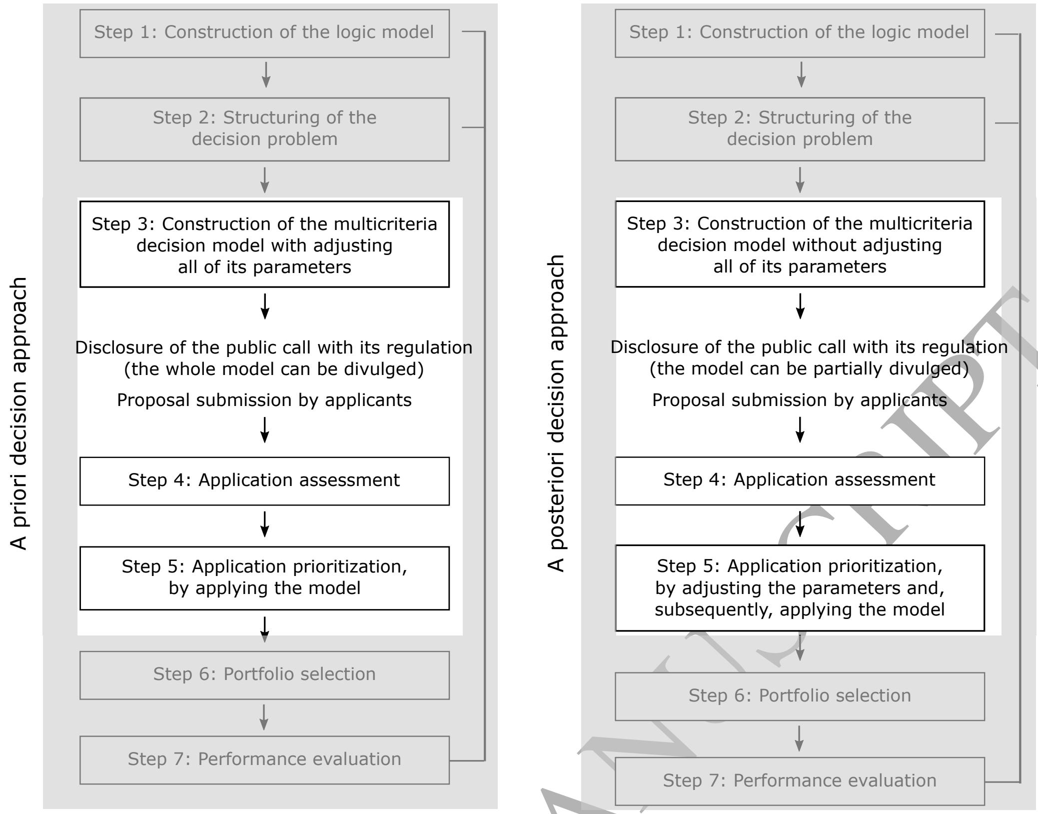

Figure 5: Differences between the approaches for *a priori* decision and *a posteriori* decision

because the score/amount ratio does not necessarily quantify the net program impacts, as in a benefit-cost relation. Indeed, these scores capture the decision-maker value judgments and, in general, cannot be taken as measures of program impacts.

Instead of considering the score/amount ratios to address the different funding amounts per application, DEMUCTI recommends applying a simple procedure for portfolio composition. It requires the definition of groups of application with roughly the same funding amounts. In this way, resources can be allocated, firstly, across the groups and, subsequently, across the applications belonging to each group. If necessary, each group can have its own set of objectives, criteria, and decision model. In such a case, steps from 1 to 5 need to be performed by considering each group separately.

In this regard, it is also worth mentioning that more sophisticated techniques could also be applied to ensure that the portfolio meets criteria of synergy, completeness and/or risk aversion, but at the expense of a decreased transparency of the funding decision. Such techniques usually employ optimization procedures which do not comply with the transparency requirements of the public call regulation. In these cases, it may not be possible to justify the decision on each application on an individual basis, but only in terms of the portfolio performance as a whole.

Finally, having selected a portfolio, its output goals[4] can be defined by considering the expected outputs of each selected application. For more reasoned estimates of feasible goals, it is helpful to include, in the application template, instructions for the applicants to specify the outputs their projects intend to deliver.

3.7. Step 7: Performance evaluation

The performance evaluation, in Step 7, is part of a continuous monitoring process of the program execution. Within DEMUCTI, the evaluations focus on both the output and the outcome assessments.

Upon completion of the projects, an ex-post evaluation is carried out to verify if the output goals were met. This information can be used in order to improve the subsequent funding decisions of the ongoing program, through the feedback loops from Step 7 to Step 1 and to Step 2 of the DEMUCTI cycle. The outcome assessments need to be conducted later, so that they capture the mid-term and long-term effects. They inform the allocation of budgets, the proposition and/or choice of new instruments, as well as the prioritization of subjects in subsequent programs.

More in-depth program evaluation studies, impact analysis and benefit-cost analysis are not within the primary scope of DEMUCTI. They usually follow experimental or quasi-experimental approaches and need to be specifically designed for each program (Cohen & Franco, 2011; Langbein & Felbinger, 2006). However, SMP practitioners are encouraged to consider implementing them in parallel with DEMUCTI for more informed decisions on which course to follow in the strategic management of the action plan execution. DEMUCTI might obtain valuable information for implementing those studies, through the application assessments performed by ad hoc reviewers and committee members.

4. Case study

In a pilot testing, DEMUCTI was applied to support the reevaluation of the applications which were submitted to the call for proposals of a program named Universal. This program is implemented by CNPq and the Ministry of Science, Technology, Innovations and Communications of Brazil.

The Universal program offers financial support for scientific and technological research projects in multiple areas of knowledge. It has a fixed duration, a specified opening date, a non-thematic scope and involves a one-round call, with a two-staged PRP (the assessments of ad hoc reviewers are consolidated by a scientific committee). Applicants are required to have doctoral degree and the curriculum vitae registered at a digital platform of CNPq, named *Plataforma Lattes*. They also must be employees

[4]Here, a goal is a target level which corresponds to a specific degree of objective achievement (Pedrycz et al., 2011).

of a non-profit higher education, R&D, or public Science, Technology and Innovation institutions headquartered in Brazil.

The pilot testing consisted in performing the first six steps of DEMUCTI, by considering the research projects which were submitted to the Chemical Engineering area. The public call distinguished among three groups of funding ranges: group A for requests of up to a maximum of R$30,000.00; group B for requests of up to a maximum of R$60,000.00; and group C for requests of up to a maximum of R$120,000.00. Because of space limitations, only the results related to the group B are presented. It accommodates 81 eligible applications, which are mainly from intermediate candidates, who are neither junior nor very experienced senior researchers. As the effective funding decision was not made when the pilot was performed, the official cutoff position was not defined yet. Hence, a portfolio of 21 applications was considered as a reference for the analysis. It was estimated from the amount of resources available for the Universal Call.

A group of individuals belonging to the CNPq staff worked on the implementation of Steps 1 to 6 of DEMUCTI, while a subgroup of four individuals were in the decision-maker role (smaller groups are expected to suffice for real applications in CNPq). In the first stage of the PRP, the same ad hoc reviewers that contributed to the official review process addressed the assessment questionnaire remotely. Subsequently, in a face-to-face panel meeting, the same members of the scientific committee that contributed to the official review process examined all applications and provided their collective answers to the questions of the assessment questionnaire.

4.1. Results of Step 1

In Step 1, the decision-makers gave their collective answers to questions Q1-Q8. The logic model shown in Figure 6 was elaborated by addressing Q1-Q6.

In the answer to Q7 (regarding the features of applicants and applications), the following features of the applicants were specified: experience in managing research projects, training of human resources, and producing relevant scientific publications. The fundamental characteristics of the research projects comprised the following features: methodological consistency; access to the needed infrastructure and human resources; potential to promote scientific and/or technological impact; alignment with the national Science, Technology and Innovation strategy.

In the answer to Q8 (regarding the outcome and impact metrics), the following outcome metrics were suggested to be considered in the evaluation of program performance: number of patents filed and granted; number of licenses granted; number of software products being developed; number of journal publications having the applicant among its authors; number of published text books, aimed at graduate or advanced research audience, having the applicant among its authors; number of master dissertations having the applicant as supervisor; number of doctoral theses having

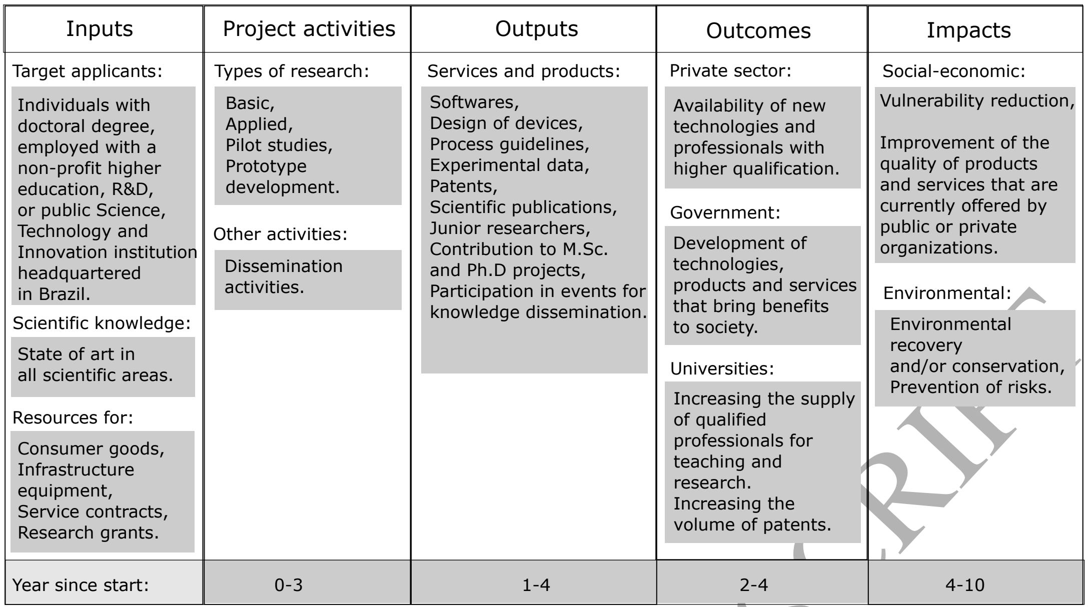

Figure 6: Logic model constructed for the Universal program

the applicant as supervisor.

4.2. Results of Steps 2 – 3

In Step 2, a value tree was constructed by the decision-makers by considering the desirable characteristics for the applicants and also for the research projects, which were defined in the answer to Q7. The value tree, Figure 7, consisted of two fundamental objectives, unfolding into four levels of specific objectives and ending with 26 leaves (the complete value tree comprised five levels).

In Step 3, all 26 value functions were defined by the decision-makers, who were advised by the scientific committee on the selection of adequate attributes. The committee was also consulted on how some of the 9 quantitative and 17 qualitative scales should be interpreted in the context of the Chemical Engineering field (for instance the *h*-index, journal impact factors, etc.).

Certain measures were taken to minimize errors and biases in the elicitation of value judgments. The decision-makers were previously trained to build the models and use the techniques considered by DEMUCTI; were informed on the common biases and errors in decision-making; could simulate the effects of different shapes of value functions on the results, while observing the resulting rankings and portfolios. The decision-makers articulated and tested their value judgments individually and in group, under a process in which they had several opportunities to discuss and revise the model, observing the effects on results, until consensus was achieved. Although these measures ought to avoid the major biases, we recognize that there is room for improvement in future implementations of DEMUCTI. For instance, consider (Montibeller & von Winterfeldt, 2015).

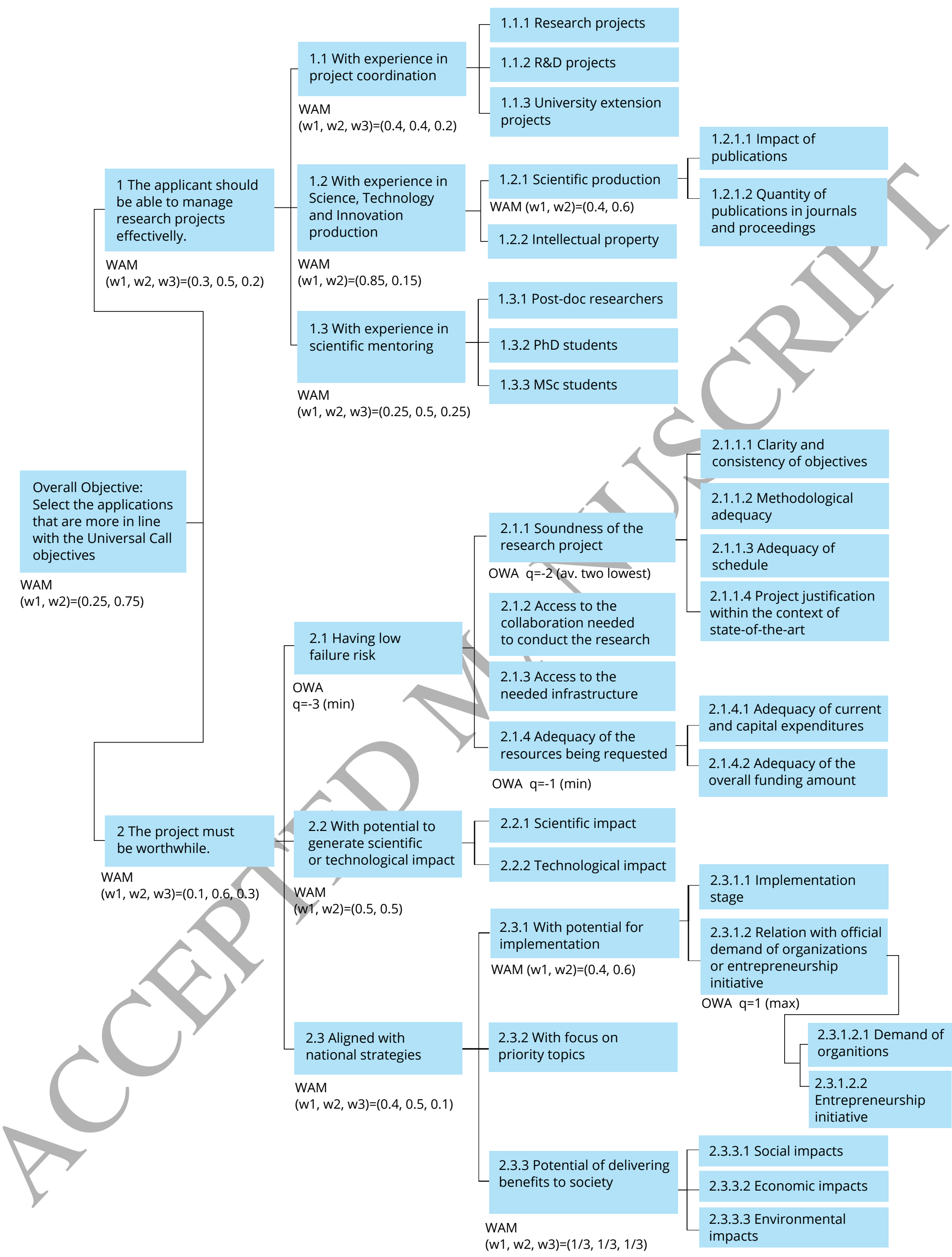

Figure 7: The value tree obtained in Step 2 and the aggregation operations defined in Step 3 (here WAM stands for weighted arithmetic mean and OWA stands for ordered weighted average).

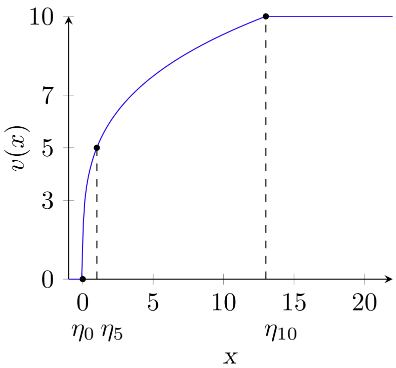

Figure 8: The value function for "number of master dissertation thesis supervised by the applicant"

Most of the value functions for quantitative attributes were defined by the decision-makers in terms of descriptive statistics, with η_5 and η_{10} set as the median and 95th percentile statistics, respectively. In some cases, the Bisection Procedure was applied with the purpose of identifying the middle point, η_5. For instance, the anchors for the attribute "number of master dissertations having the applicant as supervisor" were set to $\eta_0 = 0$, $\eta_5 = 1$ and $\eta_{10} = 13$, as shown in Figure 8. In order to define the discrete value functions for the qualitative attributes, the tools from AHP were applied, as recommended in Section 3.3.

Regarding the aggregation of value functions, except for objectives 2.1 ("Having low failure risk") and 2.3.1.2 ("Relation with official demand of organizations or entrepreneurship initiative"), the decision-makers chose to implement WAM for all the objectives with descendants in the hierarchy and, in each case, the Swing Weights procedure was applied to define the weights (the weights being utilized with WAM are presented in Figure 7 as well). In the case of objective 2.1, to reproduce a risk-averse attitude, OWA was applied with $q = -3$, acting, accordingly, as the min operator. A less pessimistic attitude was implemented for objective 2.1.1 with $q = -2$ and $m = 4$ (average of two lowest scores among the four descendants). Finally, the max operator was implemented in objective 2.3.1.2 with $q = 1$.

4.3. Results of Step 4

In Step 4, the attributes associated with the leaf descendants of the fundamental objective 1 (refer to Figure 7) were assessed by considering an automated data source: the information provided by each applicant through their curriculum vitae registered in the *Plataforma Lattes*. The attributes associated with the leaf descendants of fundamental objective 2 were assessed by the ad hoc reviewers and the committee members by answering the questions from the assessment questionnaire.

The questionnaire contained 17 closed questions, which embraced 43 items to be checked by the reviewers. All of them attempted to be free of subjective impressions, being based on formalized concepts associated with the corresponding objectives. For instance, in order to define the value function related to the specific objective 2.2.1 ("Scientific impact"), the notions of "incremental scientific advance" and "disruptive scientific advance", as defined in (Popadiuk & Choo, 2006), were considered to compose a nominal scale, with the following categories: A) significant or disruptive advance; B) incremental advance; C) no advance.

For the purpose of better characterizing every project, these three categories were considered in the questionnaire, even though the decision-makers were indifferent between projects belonging to A or B. Thus, the nominal scale was operationalized through the following multiple choice question:

> *Regarding the potential for scientific impact of the expected results of the research project, choose the option that best characterizes the proposal:*
>
> *A) The proposal has the potential for generating significant advance of the state of the art, with the creation of new paradigms, new paths and/or directions of research.*
>
> *B) The proposal has the potential for generating incremental advance to the state of the art, but does not involve paradigm shift.*
>
> *C) The proposal does not have the potential for direct or indirect contribution to the advancement of the state of the art in the research subject.*

Similar treatment was given to the other specific objectives. Some of them had to be broken down into more specific objectives, to achieve less ambiguous definitions of attributes. See, for instance, the unfolding of the objective 2.3.1 ("With potential for implementation") in Figure 7. It allowed us to consider a separate attribute to asses the implementation stage of the projects. The classification of project development stages given by (OECD, 2015) were considered to create a complete and unambiguous nominal scale for that purpose.

Finally, several open questions were also included in the questionnaire in order to allow the reviewers to provide feedback on their difficulties in evaluating or on the relevant aspects that, in their opinion, were not adequately addressed by the Assessment Questionnaire. Among the 96 reviewers who expressed their opinions about the Assessment Questionnaire, 71 expressed favorable comments or suggestions which can be easily implemented, 25 expressed unfavorable comments or complaints, which were mainly associated with the large number of questions and the difficulty of answering some questions. We observed that the suggestions and the main difficulties reported by the reviewers were directly or indirectly associated with the fact that several proponents did not include, in their proposals, enough information to answer some items of the questionnaire. This lack of information, which was already expected, confirmed the need to elaborate and divulge a template form for proposal submission as well, in order to orient the applicants on how to turn explicit the required information in the proposal manuscript.

4.4. Results of Steps 5 – 6

In Step 5, having obtained the assessments consolidated by the committee, the decision model was applied to generate a ranking of applications. In order to characterize the generated portfolios, seven perspectives were created on the basis of the nominal scales constructed in Step 3, as described in Table 2. In this way, each portfolio could be described by means of the distribution of applications among the categories of each perspective.

It was possible to perform a sensitivity analysis of the resulting portfolios with respect to the model parameters and also to different cutoff positions. Thus, the decision-makers could observe the effects on the scores generated by each single-attribute value function and the aggregated value functions corresponding to each different level of the value tree. Because of space limitations, only one form of sensitivity analysis is shown here. It considered the effects of the weights w_1 and w_2 of the two fundamental objectives on the portfolio.

In Figure 9, each stacked area chart plots the percentage distribution of applications in the portfolio (y-axis) with respect to the varying parameter (x-axis). On the left side, the distributions are plotted with respect to the cutoff position from 1 to 81, while keeping the model parameters at their original values. On the right side, the distributions are plotted with respect to the weight w_1 of the first fundamental objective from 0 to 1, as well as the weight $w_2 = 1 - w_1$ of the second fundamental objective, while keeping the cutoff position fixed at 21. The dashed line on each chart pinpoints the reference portfolio obtained for the cutoff position 21 and the weights $w_1 = 0.25$, $w_2 = 0.75$, chosen by the decision-makers in the pilot test.

The data on charts can be read as follows: for instance, from the perspective of scientific impact, the applications in the reference portfolio are approximately distributed as 60% "Significant", 40% "Incremental" and 0% "Non-identified". By observing the changes of distributions, some conclusions can be drawn regarding the sensitivity of the portfolio with respect to its size and the model parameters. For instance, it can be seen that increments of w_1 above 0.25 reduce the contribution of applications with significant scientific and technological impacts, while keeping distributions of social and environmental impacts relatively stable. Similar observations can be made with respect to the cutoff position, by considering the charts on the left side of Figure 9.

Analysis of this kind may be performed for any combination of model parameters in order to determine the sensitivity of results and obtain hints for further adjustments. For instance, one may wish to consider different adjustments of the model in order to achieve portfolios that simultaneously exhibit high impact levels for all impact-related perspectives (e.g., the perspectives of scientific, technological, economic and environmental impacts). In order to handle cases when more than two value functions are aggregated through WAM, one may consider the procedure of varying each weight while preserving the ratios between other weights (Mareschal, 1988). If the decision-makers choose OWA to aggregate two or more value functions, one may consider a sensitivity analysis with respect to the pessimism-optimism index q.

It is important to bear in mind that this analysis only covers a particular sample of proposals and setting of parameters. Therefore, the conclusions should not be generalized without considering different samples and parameter settings.

5. Discussions and Conclusions

The CNPq staff involved in the pilot test could implement the methodology within the SMP cycle. They agreed that DEMUCTI encouraged the consideration of rational criteria, enabled addressing the program objectives in a more explicit way and also enabled the agency to orient the portfolio selection in the directions indicated by the strategic planning of CNPq. As demonstrated in results of Step 4 (refer to Section 4.3), the majority of reviewers were favorable to the new format of the Assessment Questionnaire. However, given that the main objective of the test was to confirm the applicability of DEMUCTI, the pilot had the following main limitations:

- The applicants were not provided with the application template, which brought difficulties for the assessment of some proposals, as discussed in Section 4.3.
- The Assessment Questionnaire did not undergo a complete pretesting that allows, among other things, to identify the most adequate wording, format and order of items and instructions (Colton & Covert, 2007)
- Since there were no partner institutions involved in the funding decision, the pilot test did not allow to check the feasibility of incorporating the objectives and preferences of external actors into the models considered by the methodology.
- In Step 5, only the *a posteriori* approach of the funding decision was considered. Hence, the pilot test did not allow to verify the performance of a decision model adjusted before the application submission, as in the *a priori* decision approach.
- As the performance evaluation of Step 7 was not implemented, the pilot study did not allow to confirm if the information gathered by executing DEMUCTI does benefit the subsequent decisions related to the ongoing program.

The decision model, emerging from the use of DEMUCTI, inherits a number of useful characteristics from MAVT. These characteristics make DEMUCTI suitable for dealing with certain circumstances that may take place within a funding decision process and the management of programs. Among them, it is possible to enumerate:

Table 2: List of perspectives and categories

Perspectives	Category labels	Category descriptions
Scientific impact	Significant	Work with potential to produce significant advance of the state of the art.
	Incremental	Work with potential to produce incremental advance to the state of the art.
	Non-identified	Work does not satisfy the conditions to be assigned in the aforementioned categories.
Technological impact	New PPS	Work contributes directly to the creation of new Products, Processes or Services (PPS).
	New technologies	Work contributes directly to generate new technologies.
	Improved PPS	Work contributes directly to the improvement of existing PPS.
	Non identified	Work does not satisfy the conditions to be assigned in the aforementioned categories.
Social impact	Direct	Work with potential to directly produce a social impact, in some directions aligned with the program objectives.
	Indirect	Work with potential to indirectly produce a social impact, in some directions aligned with the program objectives.
	Non identified	Work does not satisfy the conditions to be assigned in the aforementioned categories.
Economic impact	Direct	Work with potential to directly produce a economic impact, in some directions aligned with the program objectives.
	Indirect	Work with potential to indirectly produce a economic impact, in some directions aligned with the program objectives.
	Non identified	Work does not satisfy the conditions to be assigned in the aforementioned categories.
Environmental impact	Direct	Work with potential to directly produce an environmental impact, in some directions aligned with the program objectives.
	Indirect	Work with potential to indirectly produce an environmental impact, in some directions aligned with the program objectives.
	Non identified	Work does not satisfy the conditions to be assigned in the aforementioned categories.
Demand and entrepreneurship	Demand and entrepreneurship	Work is associated with an official demand of an organization and with an entrepreneurship initiative.
	Demand	Work is associated with an official demand of some organization, but without an entrepreneurship initiative.
	Entrepreneurship	Work is associated with an entrepreneurship initiative, but without an official demand of an organization.
	Non identified	Work does not satisfy the conditions to be assigned in the aforementioned categories.
Implementation stage	Pilot or prototype	Work is in the stage of pilot-testing or prototype development.
	Applied research	Work is in the stage of applied research.
	Basic research	Work is in the stage of basic research.

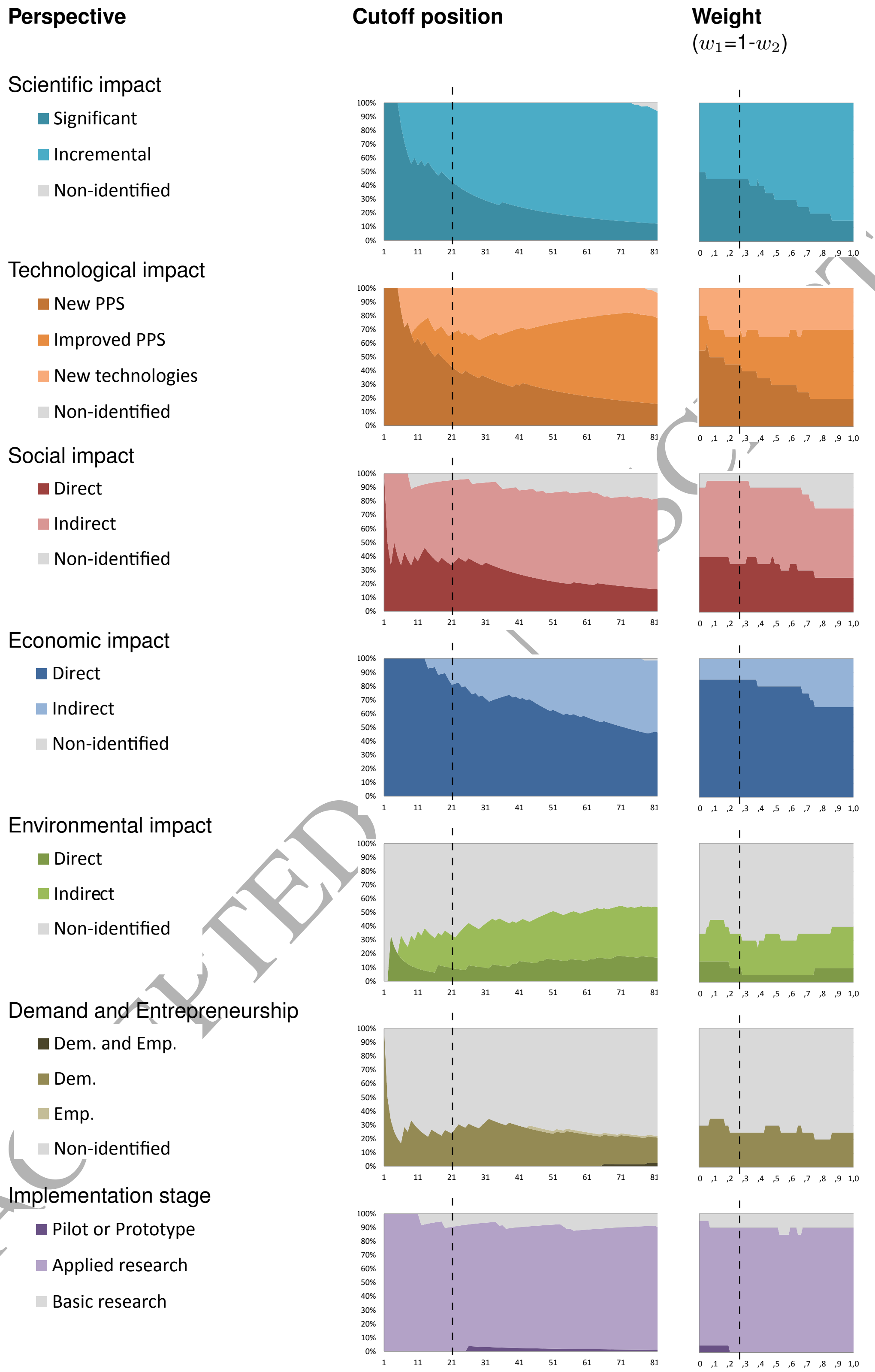

Figure 9: Charts of percentage distributions of applications in different portfolios, varying with the cutoff position (left) and the weights of fundamental objectives (right)

- Frequently, large sets of applications need to be addressed. In such cases, the model can reduce the workload of the decision-maker, because its use allows value functions to be elicited based on valuing factual characteristics, rather than the applications itself (Ferretti, 2016);

- If the feature "Right to Appeal" (European Science Foundation, 2011) is implemented or if further eligibility checks lead to later vetoes of applicants, it is not necessary to perform the preference elicitation over again. Moreover, the results keep backward consistency with the previous ranking (no reversals), after adding or removing applications;

- It may be necessary to consider a large set of attributes to capture all the required properties of the applications and the applicants, to ensure a successful program implementation. The value tree can accommodate them in a structured way, regardless of the qualitative or quantitative nature of these attributes (Schuwirth et al., 2012);

- The model generates a ranking and a score reflecting the global performance of every application. If the parameters of the multicriteria decision model are kept constant, these scores can be utilized to compare applications submitted in response to different public calls of the program. This kind of comparison may provide useful information for program evaluation purposes.

Other well-known multicriteria decision methods, PROMETHEE, ELECTRE and AHP, do not exhibit simultaneously all the characteristics enumerated above (Schuwirth et al., 2012). In fact, recent developments on PROMETHEE, such as (Doan & Smet, 2016; Calders & Assche, 2018), make it more suitable for the support of funding decisions in the public sector. However, it is important to bear in mind that the decision model needs to be translated into evaluation rules which can be easily understood and accepted by the target applicants, who may be researchers from different fields, not limited to exact sciences.

MAVT-based models can be implemented in many different ways. In our opinion, the particular choice of the elements and techniques specified in Step 3 of DEMUCTI favors an agile and scalable use of DEMUCTI within CNPq. We also believe that, in practice, Step 3 can and should be adapted to meet the settings of other agencies. Other preference elicitation procedures can be applied rather than the ones recommended here. Besides, other forms of value functions and aggregation operations can be used, for instance, consider (Reichert et al., 2015).

Regarding the core principles of good practice in peer review presented in (European Science Foundation, 2011), the use of DEMUCTI allows to reinforce them by providing a clear separation of roles, impartiality in the evaluation process, clear criteria and rules that can be previously divulged. As aforementioned, DEMUCTI also allows to deal adequately with the right to reply of applicants and with proposals which contravene ethical or integrity principles. Once identified, they can be eliminated at any stage of the PRP and even after the disclosure of the funding decision results.

Regarding the introduction of DEMUCTI in practice, some new activities may be incorporated into the core processes of the agency (e.g., the construction of the logic model, the value tree and the decision model; the preparation of the assessment questionnaire and the application template; the definition of outcome and impact metrics). These activities are time-consuming and may demand more effort than usual from the agency staff, especially in the first cycle of each program.

It is worth noting that the use of DEMUCTI implies a commitment of the practitioners to discover and understand the means by which the program outcomes and impacts can be effectively obtained. Once certain hypotheses are considered, they should be systematically tested, under a continuous effort to validate them. Whenever a hypothesis needs to be modified, the logic model and the decision model should be updated to maintain the logical consistency within the whole decision-making context. In this way, the use of DEMUCTI can reinforce the implementation of a results-based approach for program management.

In future applications, one may consider the evolution of DEMUCTI towards a method for dealing with interdisciplinarity and synergy among different programs. Besides, it is possible to improve the elicitation process by considering some techniques to overcome motivational and cognitive biases (Montibeller & von Winterfeldt, 2015), as well as some formal procedures for consensus construction and for aggregation of discordant preferences (Pedrycz et al., 2011).

References

References

Abdoul, H., Perrey, C., Amiel, P., Tubach, F., Gottot, S., Durand-Zaleski, I., & Alberti, C. (2012). Peer review of grant applications: Criteria used and qualitative study of reviewer practices. *PLoS ONE*, *7*, 1–15.

Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics*, *87*, 499–514.

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Allocative efficiency in public research funding: Can bibliometrics help? *Research Policy*, *38*, 206–215.

August, D., & Muraskin, L. D. (1999). *Strengthening the Standards: Recommendations for OERI Peer Review. Summary Report Prepared for the National Educational Research Policy and Priorities Board*. Washington, DC: US Department of Education.

Beinat, E. (1997). *Value functions for environmental management*. Dordrecht: Kluwer Academic Publishers.

Belton, V., & Stewart, T. J. (2002). *Multiple criteria decision analysis: an integrated approach*. Dordrecht: Springer Science+Business Media.

Bornmann, L. (2012). What is societal impact of research and how can it be assessed? a literature survey. *Journal of the American Society for Information Science and Technology*, *64*, 217–233.

Borrás, S., & Edquist, C. (2013). The choice of innovation policy instruments. *Technological Forecasting and Social Change*, *80*, 1513–1522.

Calders, T., & Assche, D. V. (2018). PROMETHEE is not quadratic: An o(qnlog(n)) algorithm. *Omega*, *76*, 63–69.

Choo, E. U., Schoner, B., & Wedley, W. C. (1999). Interpretation of criteria weights in multicriteria decision making. *Computers & Industrial Engineering*, *37*, 527–541.

Cohen, E., & Franco, R. (2011). *Evaluation of social projects (in Portuguese)*. Petrópolis: Vozes.

Colton, D., & Covert, R. W. (2007). *Designing and Constructing Instruments for Social Research and Evaluation*. San Francisco: Jossey-Bass.

Costa, E. A. (2014). *Strategic Management: from the company we have got to the company we want (in Portuguese)*. Saraiva.

Doan, N., & Smet, Y. D. (2016). On the use of reference profiles to compute alternative PROMETHEE II rankings: A preliminary study. In *Proceedings of the 2016 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 326–330). IEEE.

Dubois, D., & Prade, H. (2012). Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems*, *192*, 3–24.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350.

Dyer, J. S. (2016). Multiattribute utility theory (maut). In S. Greco, M. Ehrgott, & J. R. Figueira (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (pp. 285–314). New York: Springer.

Ebadi, A., & Schiffauerova, A. (2016). iSEER: an intelligent automatic computer system for scientific evaluation of researchers. *Scientometrics*, *107*, 477–498.

Edquist, C., & Zabala-Iturriagagoitia, J. M. (2012). Public procurement for innovation as mission-oriented innovation policy. *Research Policy*, *41*, 1757–1769.

Edwards, W., & Barron, F. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, *60*, 306–325.

Erdt, M., Nagarajan, A., Sin, S.-C. J., & Theng, Y.-L. (2016). Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics*, *109*, 1117–1166.

European Science Foundation (2011). *European Peer Review Guide, Integrating Policies and Practices into Coherent Procedures*. Strasbourg: European Science Foundation.

Ferretti, V. (2016). From stakeholders analysis to cognitive mapping and multi-attribute value theory: An integrated approach for policy support. *European Journal of Operational Research*, *253*, 524–541.

Grafakos, S., Flamos, A., Oikonomou, V., & Zevgolis, D. (2010). Multi-criteria analysis weighting methodology to incorporate stakeholders preferences in energy and climate policy interactions. *International Journal of Energy Sector Management*, *4*, 434–461.

Guinea, J., Sela, E., Gomez-Nunez, A. J., Mangwende, T., Ambali, A., Ngum, N., Jaramillo, H., Gallego, J. M., Patino, A., Latorre, C., Srivanichakorn, S., & Thepthien, B. (2015). Impact oriented monitoring: A new methodology for monitoring and evaluation of international public health research projects. *Research Evaluation*, *24*, 131–145.

Guthrie, S., Guerin, B., Wu, H., Ismail, S., & Wooding, S. (2013). *Alternatives to Peer Review in Research Project Funding*. Cambridge: RAND Corporation.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*. Maryland: Rowman & Littlefield Publishers.

Joyce, P. (2015). *Strategic Management in the Public Sector*. New York: Routledge.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Keeney, R. L. (1992). *Value-Focused Thinking: A Path to Creative Decisionmaking*. Cambridge: Harvard University Press.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. New York: Wiley.

Langbein, L., & Felbinger, C. L. (2006). *Public Program Evaluation: A Statistical Guide*. New York: M.E. Sharpe Inc.

Mareschal, B. (1988). Weight stability intervals in multicriteria decision aid. *European Journal of Operational Research*, *33*, 54–64.

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, *63*, 160–168.

Montibeller, G., & von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, *35*, 1230–1251.

OECD (2012). *OECD Science, Technology and Industry Outlook 2012*. Paris: OECD Publishing.

OECD (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*. Paris: OECD Publishing.

Papon, P. (1999). The role of national agencies in evaluation. *Scientometrics*, *45*, 391–399.

Pazvakavambwa, A., & Steyn, G. M. (2014). Implementing results-based management in the public sector of developing countries: What should be considered? *Mediterranean Journal of Social Sciences*, *5*, 245–257.

Pedrycz, W., Ekel, P., & Parreiras, R. (2011). *Fuzzy Multicriteria Decision-Making: Models, Methods and Applications*. Chichester: Wiley.

Poister, T. H. (2010). The future of strategic planning in the public sector: Linking strategic management and performance. *Public Administration Review*, *70*, 246–254.

Popadiuk, S., & Choo, C. W. (2006). Innovation and knowledge creation: How are these concepts related? *International Journal of Information Management*, *26*, 302–312.

Reichert, P., Langhans, S. D., Lienert, J., & Schuwirth, N. (2015). The conceptual foundation of environmental decision support. *Journal of Environmental Management*, *154*, 316 – 332.

Roberts, F. S. (1985). *Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences*. New York: Cambridge University Press.

Royse, D. D., Thyer, B. A., & Padgett, D. K. (2009). *Program Evaluation: An Introduction*. Belmont: Cengage Learning.

Saaty, T. (1980). *The Analytic Hierarchy Process*. New York: McGraw Hill.

Salo, A. A., & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-criteria decision analysis*, *6*, 309–319.

Schalock, R. L., & Thornton, C. V. D. (1988). *Program Evaluation: A Field Guide for Administrators*. New York: Springer US.

Schuwirth, N., Reichert, P., & Lienert, J. (2012). Methodological aspects of multi-criteria decision analysis for policy support: A case study on pharmaceutical removal from hospital wastewater. *European Journal of Operational Research*, *220*, 472–483.

Simon, H. A. (1997). *Administrative Behaviour: a Study of Decision-Making Processes in Administrative Organizations*. New York: Simon & Schuster.

Steiss, A. W., & Dekker, M. (2003). *Strategic Management for Public and Nonprofit Organizations*. New York: Marcel Dekker Inc.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision-making. *IEEE Transactions on Systems, Man, and Cybernetics*, *18*, 183–190.

Yager, R. R. (1995). Multicriteria decision making using fuzzy quantifiers. In *Proc. IEEE Conference on Computational Intelligence for Financial Engineering* (pp. 42–47). New York: IEEE.