INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Behavioral Challenges in Policy Analysis with Conflicting Objectives

Gilberto Montibeller

To cite this entry: Gilberto Montibeller. Behavioral Challenges in Policy Analysis with Conflicting Objectives. *In* INFORMS TutORials in Operations Research. Published online: 19 Oct 2018; 85-108.
https://doi.org/10.1287/educ.2018.0182

https://pubsonline.informs.org/series/educ/

TUTORIALS IN OPERATIONS RESEARCH
© 2018 INFORMS | ISBN 978-0-9906153-2-3
https://doi.org/10.1287/educ.2018.0182

Behavioral Challenges in Policy Analysis with Conflicting Objectives

Gilberto Montibeller
Management Science and Operations Group, School of Business and Economics, Loughborough University, Leicestershire LE11 3TU, United Kingdom
Contact: g.montibeller@lboro.ac.uk, https://orcid.org/0000-0003-1063-080X (GM)

Abstract Public policy problems are rife with conflicting objectives: efficiency versus fairness, technical criteria versus political goals, costs versus multiple benefits. Multicriteria decision analysis provides robust methodologies to support policy makers in making tough choices and in designing better policy options when considering these conflicting objectives. However, important behavioral challenges exist in developing these models: the use of expert judgments, whenever evidence is not available; the elicitation of preferences and priorities from policy makers and communities; and the effective management of group decision processes. The extensive developments in behavioral decision research, social psychology, facilitated decision modeling, and incomplete preference models shed light on how decision analysts should address these issues so we can provide better decision support and develop high-quality decision models. In this tutorial, I discuss the main findings of this extensive but rather fragmented literature and provide a coherent and practical framework for managing behavioral issues, minimizing behavioral biases, and optimizing the quality of human judgments in policy analysis models with conflicting objectives. I illustrate these guidelines with policy analysis interventions that we have conducted over the last decade for several organizations, such as the World Health Organization, the Food and Agriculture Organization of the United Nations, the UK Department of Environment Food and Rural Affairs, the Malaria Consortium/USAID, and the UK National Audit Office, among others.

Keywords policy analysis • multicriteria decision analysis • behavioral decision research • facilitated decision modeling • cognitive biases • motivational biases • group decision making

1. Introduction

Public policy problems are rife with conflicting objectives: efficiency versus fairness, technical criteria versus political goals, costs versus multiple benefits. Multicriteria decision analysis (MCDA) (Belton and Stewart [10], Greco et al. [45], Wallenius et al. [133]) provides robust methodologies to support policy makers in making tough choices and in designing better policy alternatives when considering these conflicting objectives (Daniell et al. [19], Tsoukias et al. [128]).

In MCDA-based policy analysis, we work with groups of policy makers, modeling their decisions, facilitating their discussions, and representing preferences and priorities. The overarching goal is to improve decision processes (Spetzler et al. [126]) and provide support to evidence-based decision making, considering public priorities and the inherent uncertainties that long-term horizons and complex systems bring into the problem (Tsoukias et al. [128]).

Key challenges in these MCDA interventions are the use of expert judgments, whenever evidence is not available; the elicitation of preferences and priorities from policy makers and communities; and the effective management of group decision processes. Human behavior plays a major role on each one of these challenges: experts may be biased in their estimates, individuals may be unable to express clearly their preferences, and groups may present dysfunctional dynamics.

The extensive developments in behavioral decision research, social psychology, facilitated decision modeling, and incomplete preference models shed light on how decision analysts should address these issues so we can provide better decision support and develop high-quality decision models. In this tutorial, I discuss the main findings of this extensive but rather fragmented literature and provide a coherent and practical framework for managing behavioral issues, minimizing behavioral biases, and optimizing the quality of human judgments in policy analysis models with conflicting objectives.

I illustrate these guidelines with policy analysis interventions that we have conducted over the last decade for several organizations, such as the evaluation of capabilities of health systems against rabies for the World Health Organization (WHO), the prioritization of low-moisture food categories for the Food and Agriculture Organization of the United Nations (FAO), the assessment of biosecurity threats for the UK Department of Environment Food and Rural Affairs (DEFRA), the evaluation of malaria treatment kits for the Malaria Consortium/USAID, and the prioritization of value-for-money auditing studies for the UK National Audit Office.

The tutorial has the following structure. The next section describes a framework for supporting policy analysis with conflicting objectives. Then the three key behavioral challenges are explored in sequence, followed by practical advice on how to deal with each one of them. I conclude the tutorial by suggesting some directions for further research on the topic and with a couple of warnings for decision analysts who want to support policy making with multiple objectives.

2. Policy Analysis with Facilitated Multicriteria Decision Analysis

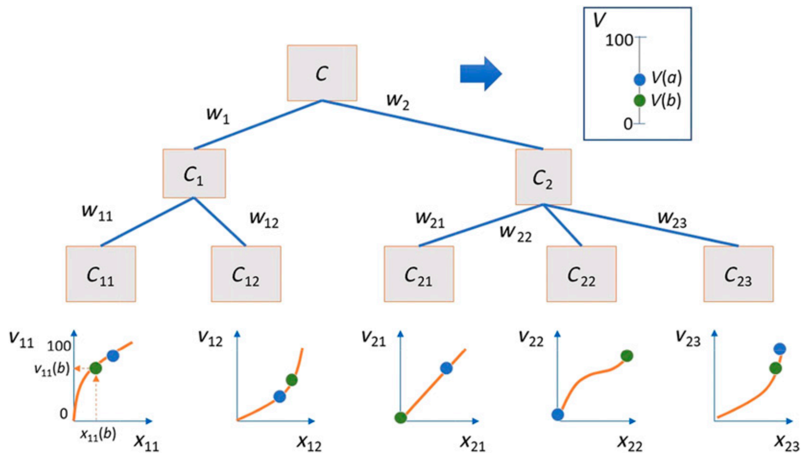
MCDA has been extensively employed in policy analysis, supporting a wide variety of prioritization problems and planning, such as in human health (e.g., Airoldi et al. [2], Cox et al. [18]), animal health (e.g., Brookes et al. [14]), regional planning (e.g., Bana e Costa et al. [7], Ferretti and Degioanni [35]), military decisions (e.g., Ewing et al. [30], Parnell et al. [104]), hazard disposal (e.g., Merkhofer and Keeney [90], Morton et al. [100]), counterterrorism analysis (e.g., Keeney and von Winterfeldt [69, 70]), environmental assessments (e.g., Ferretti [34], Gregory [46]), natural resource management (Romero and Rehman [115]), and energy-related decisions (Wang and Poh [135]), to name just a few areas of application.

I will consider here multiattribute utility theory (MAUT) (Keeney and Raiffa [67]), given its widespread use and normative foundations on decision theory (French [43]) and on measurement theory (Krantz et al. [76]), as well as the available behavioral evidence on judgments for this type of model. However, other important schools of multicriteria decision analysis have also been extensively employed to support policy makers, such as the outranking methods (Figueira et al. [37]) and the analytic hierarchy process (Saaty [120]), among others. I will describe next how MAUT can be operationalized for the evaluation of policy options, followed by how these MCDA models can be built up with groups of policy makers.

2.1. Multicriteria Analysis for the Evaluation of Policy Options

The basic idea of MAUT-based models is to decompose an overall objective (assessed by a criterion C) as a value tree into subobjectives, each one assessed by a subcriterion C_i, which can be subsequently further decomposed by subcriteria C_{ij}, as illustrated in Figure 1 (see also Belton and Stewart [10]). The subcriteria at the bottom of the tree have associated attributes; each x_{ij} attribute measures the achievement of each policy option (a and b in the same figure)

Figure 1. (Color online) Multicriteria value analysis of two policy options.



on the respective C_{ij} subcriterion (see Keeney and Gregory [66]). A value function v_{ij} in the case of riskless choices, or a utility function (u_{ij}) in the case of decisions under uncertainty, is elicited from policy makers, normalizing partial performances, usually on a scale from 0 to 100 (see von Winterfeldt and Edwards [132]). In decision analysis, value functions express the strength of preferences of policy makers for outcomes under certainty, whereas utility functions represent both risk attitude and strength of preference for outcomes under uncertainty. For instance, in the same figure, for subcriterion C_{11}, the policy option b has a partial performance $x_{11}(b)$ on this attribute $x_{11}(.)$ and a respective partial value $v_{11}(b)$ on this value function.

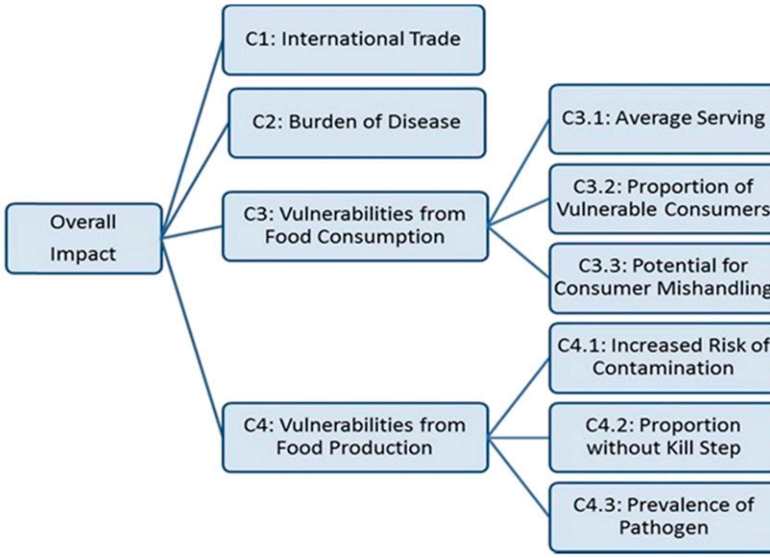
Criteria weights w_{ij}, associated with each C_{ij}th subcriterion, are elicited from the policy makers, representing their value trade-offs (see Keeney [63]). These weights are then used to aggregate the partial performances of each policy option on an overall value V, again usually on a 0–100 scale. For example, the overall value of policy option b is $V(b)$ in the same figure. The most common aggregation function is a simple weighted sum, but this requires strict preference conditions among the criteria (see von Winterfeldt and Edwards [132]).

The axiomatic and mathematical foundations of this type of model are well known and beyond the scope of this tutorial; Keeney and Raiffa [67] and von Winterfeldt and Edwards [132] provide excellent and detailed coverage of them. For a step-by-step introduction to these models, with a focus on the modeling process, see Montibeller and Franco [93]. Comprehensive textbooks with in-depth coverage of the methods and a large number of didactical examples are also available (e.g., Belton and Stewart [10], Eisenführ et al. [28]).

I will illustrate the use of this type of multicriteria model for policy analysis in practice with a project in which I was involved as a decision analyst a few years ago in collaboration with a risk analyst (Mike Batz from the University of Florida). Its aim was the prioritization of low-moisture food (LMF) categories for the FAO/WHO. These categories of food (such as cereals and grains, dried protein products, nuts and nut products) are important, with high volumes of global consumption and billions of dollars of international trade. The objective of the intervention was to identify the category with the highest overall impact, responding to a request from the Codex Committee on Food Hygiene (CCFH), which could then more rigorously consider and manage the microbiological hazards associated with these products. It was therefore critical that the assessment was conducted in a robust and transparent way, utilizing the best expertise on the subject available and a sound methodology for the assessment of impacts and ranking of food categories (for details, see FAO/WHO [31]).

For the decision analysis we employed a multiattribute value analysis. The value tree is shown in Figure 2. Noticeably, only two criteria were present when the two-day decision conference started: international trade (C1) and burden of disease (C2); the other two criteria,

Figure 2. (Color online) Value tree for the assessment of low-moisture foods for FAO/WHO [31].



vulnerabilities from food consumption (C3) and vulnerabilities from food production (C4), were agreed upon during the facilitated discussions.

We also supported the policy makers in defining precisely the attributes (x_{ij} in Figure 1) for the evaluation of LMF categories (see Table 1). There was full evidence available for assessing the impact of each food category only for criteria C1 and C2. Criteria C3 and C4 were decomposed, with evidence available for some subcriteria (e.g., C3.1, average serving) but the need for using expert judgment for others (e.g., C3.3, potential for consumer mishandling).

Each LMF category was assessed on every attribute for the value of their impact. For instance, Table 2 presents the assessment for burden of disease (partial performance of each policy option, $x_{ij}(.)$, in Figure 1) and the respective partial value of each LMF category ($v_{ij}(.)$ in Figure 1). Notice that in this case we assumed a linear value function, to reflect equity among patients affected by past outbreaks, but that is not always the case. Indeed, research has shown that multiattribute value models are sensitive to the shape of the value function (Stewart [127]).

We elicited the weights using the swing weights protocol (see Section 3.1.5 for details) and employed a simple weighted sum for the aggregation of partial values, as the required preference independence conditions were met. The overall values ($V(.)$ in Figure 1) are shown in Figure 3. The category cereals and grains (Cat 1) had the highest overall impact, followed by Cat 4 (dried protein products).

Two key behavioral challenges affect this type of modeling. The first one is to *elicit knowledge and content for the multicriteria model from the policy makers*: objectives, attributes, value/utility functions, and weights (c_{ij}, x_{ij}, v_{ij} or u_{ij}, and w_{ij}, respectively, in Figure 1). The second key behavioral challenge is to *use expert judgments and minimize biases in those judgments*, whenever the evidence is not available or of low quality, to estimate the performances of each policy option ($x_{ij}(.)$ in Figure 1). These MCDA models are often built up with a group of policy makers, with the decision analyst also playing a role as the group's facilitator, as detailed next.

2.2. Facilitated Decision Modeling

In these interventions we employ facilitated decision modeling (Franco and Montibeller [40]), in which the decision analyst is also a neutral facilitator for the group. The analyst facilitates

Table 1. Attributes for the assessment of low-moisture foods for FAO/WHO [31].

Criteria	Subcriteria	Attribute
C1: Impact on international trade	—	Export value in US$ billions per year
C2: Burden of disease	—	Total DALYs in outbreak cases from 1990 on
C3: Vulnerabilities due to food consumption	C3.1: Average serving	Average grams per day
	C3.2: Proportion vulnerable consumers	Proportion (0%–100%) consumed by vulnerable groups (toddlers and elderly)
	C3.3: Potential for consumer mishandling	Proportion (0%–100%) of LMF products in a given category with an increased risk as a result of mishandling/poor practices at any time between final retail and consumption
C4: Vulnerabilities due to food production	C4.1: Increased risk of contamination	Proportion (0%–100%) of LMF products in a given category with an increased risk of contamination after the kill step
	C4.2: Proportion without kill step	Proportion (0%–100%) of LMF products in a given category without a kill step prior to retail and distribution
	C4.3: Prevalence of pathogen	Probability that an LMF is contaminated at a level with any pathogens with the potential to cause illness in consumers

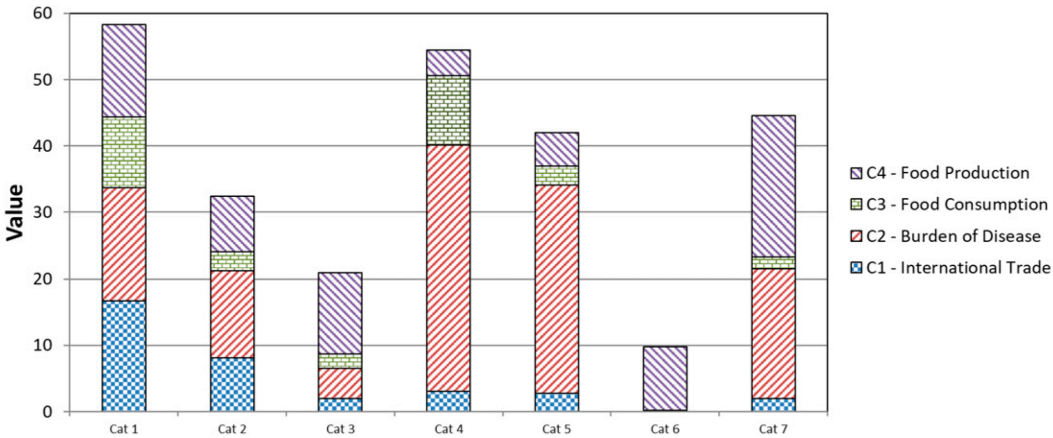
strategy workshops with a small group of senior policy makers in charge of the decision, in the format of decision conferences (Phillips [108]). In these workshops the analyst is responsible for modeling the decision, for facilitating policy makers' discussions, and for representing their preferences and priorities. Therefore, two roles must be simultaneously managed: the *content* of the policy analysis and the *process* of making the decision.

This type of *facilitated decision analysis* brings several key benefits for supporting policy making: (i) the opportunity to share different pieces of information about policy options (Rouwette and Vennix [117]) and perspectives about relevant objectives (Eden and Ackermann [27]), (ii) the ability to provide a platform for negotiation of different positions (Eden and Ackermann [26]), and (iii) the drive toward a joint commitment for successful

Table 2. Impacts and value of LMF categories on burden of disease for FAO/WHO [31].

		C2: Burden of disease	
Code	Category name	Total DALYs in outbreak cases from 1990 on	Normalized impact (v_2) (value)
Cat 1	Cereals and grains	72.53	45.9
Cat 2	Confections and snacks	60.26	35.4
Cat 3	Dried fruits and vegetables	32.78	12.2
Cat 4	Dried protein products	136.44	100.0
Cat 5	Nuts and nut products	118.51	84.8
Cat 6	Seeds for consumption	18.42	0.0
Cat 7	Spices, dried herbs, and tea	80.71	52.8

Figure 3. (Color online) The overall ranking of low-moisture food categories for FAO/WHO [31].



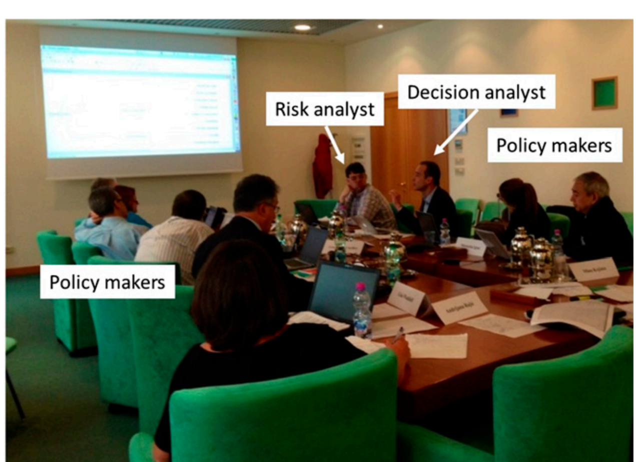
implementation of a chosen policy option (Franco and Montibeller [40], Franco et al. [41], Phillips [108]).

Group learning is a key aim in this type of intervention (Roy [118]) with several potential benefits: understanding the intricacies of the decision problem, agreeing on the objectives and priorities that must be considered, identifying the costs and benefits of each decision option, and considering the key uncertainties that may affect their performance. These multicriteria decision models should thus be employed as learning tools (de Geus [20]) instead of providing the "optimal" solution. Figure 4 illustrates a typical setting for facilitated policy analysis, the decision conference with the group of low-moisture food experts at the FAO headquarters in Rome.

Once again, I will use the project for FAO/WHO on prioritizing low-moisture food categories to illustrate this type of approach. There were several challenges that were addressed during its development: the need for a global perspective in the assessment, the existence of multiple impacts of concern, the limited amount of evidence about some of these impacts, and the need to promote the sharing of expertise and opinions among the experts involved in the assessment.

In a two-day workshop, we supported a team of top international experts and policy makers from the United States, Canada, Japan, Switzerland, and the United Kingdom in developing

Figure 4. (Color online) A typical facilitated decision analysis setting for low-moisture food experts at FAO.



Source: FAO.

a multicriteria evaluation model. The model was interactively built up with the group and projected on the screen; I was managing the model content as well as facilitating the group dynamics. Notice the layout of the meeting with an oval table that allowed face-to-face interactions and discussions mediated by the decision model. We explored the solutions in the model projecting from a computer (shown in Figure 4) and "played" with the model interactively with the policy makers. This analysis was focused on (i) varying criteria weights and (ii) addressing policy makers' concerns about the lack of reliable evidence by ranging the partial impacts until they were satisfied with the robustness of the ranking. After the meeting, a more extensive evidence-gathering session was conducted, and we developed a full sensitivity analysis, which informed the final report (FAO/WHO [31]).

Despite the important benefits of a facilitated mode, it does bring an additional key behavioral challenge when compared with mathematical aggregation of preferences (see Belton and Pictet [9] and Keller et al. [72]). The key challenge is *how to manage the group dynamics and minimize group biases of policy makers and experts.*

The tutorial will now analyze each of these key behavioral challenges and provide recommendations for how to deal with them and maximize the quality of decision-analytic models and decision processes.

3. Key Challenge I: Value Judgments in Multicriteria Policy Analysis

Two core principles of multicriteria decision analysis are the modeling of decision makers' values, given the conflicting objectives being pursued (Keeney [65], Parnell et al. [105]) and the decomposition of a complex decision into its important components (Howard [49], Keeney [61], Raiffa [114]). Indeed, there is behavioral evidence that this decomposition improves the quality of choices (Arkes et al. [5], Morera and Budescu [97]). However, the quality of a multicriteria model relies heavily on the input from decision makers and on the ability of a decision analyst in correctly eliciting their judgments (Dias et al. [24]) and in adequately representing their values (Montibeller and von Winterfeldt [94]).

The behavioral literature on modeling values is rather limited when compared with the findings on probabilistic thinking. Two major behavioral issues impact such modeling tasks: biases that may affect policy makers' judgments and their cognitive limitations in quantitatively expressing their preferences. As Montibeller and von Winterfeldt [94] argue, only a relatively small number of biases are relevant for risk and decision analysis. In addition to the cognitive biases, which are discussed in the behavioral decision research literature, relevant motivational biases are also pervasive, but they are studied in social psychology. I review cognitive and motivational biases next, followed by issues related to cognitive complexity, and I provide some practical advice on how to overcome each of them.

3.1. Biases in Modeling Values

Five main components can be distinguished in a multicriteria model: objectives, decision alternatives, attributes, value or utility functions, and criteria weights (as shown in Figure 1). I will list both cognitive and motivational biases affecting each one of these tasks and provide some suggestions on how to debias and/or improve each modeling step (for details, see Montibeller and von Winterfeldt [94]).

3.1.1. Biases in the Definition of Objectives. The identification and structuring of objectives relies heavily on policy makers' mental models of the problem situation (Johnson-Laird [55]). These mental models might be affected by *myopic problem representation bias*, in which the problem definition is oversimplified (Legrenzi and Girotto [81], Legrenzi et al. [82]). In addition, there is evidence that decision makers suffer from the *omitted variable bias* (Jargowsky [52]), in which some fundamental objectives are neglected, and *availability bias*

(Tversky and Kahneman [129]), in which only recently relevant objectives are considered. These biases may prevent the group from generating a comprehensive set of objectives (Bond et al. [12, 13]). In addition to these cognitive biases, one motivational bias that may affect the definition of objectives is the *desirability of policy options bias* (Montibeller and von Winterfeldt [94]), in which an objective is included as it favors a "pet" policy option.

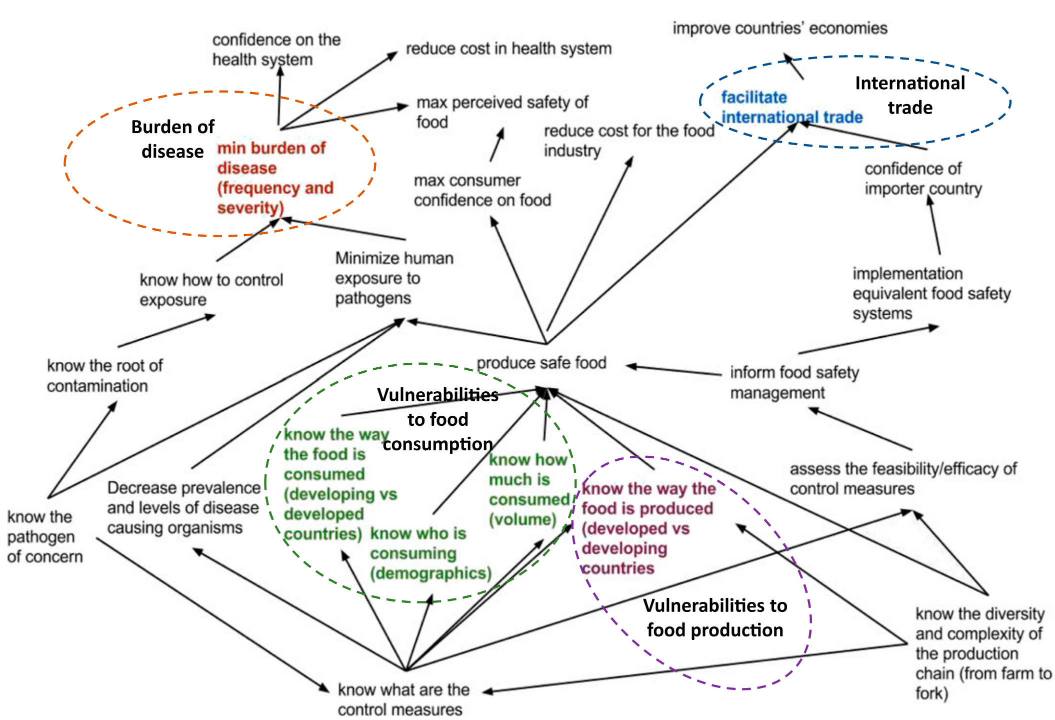
The evidence on how these biases might be overcome is rather limited, but three suggestions include the use of problem-structuring methods (Belton et al. [11], Mingers and Rosenhead [91], Montibeller and Belton [92]) and an adequate decision framing (Barcus and Montibeller [8], Keeney [62]) to support the development of a comprehensive set of objectives. The use of external probes may also help to increase the number and quality of objectives (Bond et al. [13], León [83]).

For the project on ranking low-moisture food categories for FAO/WHO, for instance, I employed a causal map (Eden [25]) to help the group identify all the relevant objectives (see Figure 5). From their laptops, group members were asked to input on a shared Google doc the concepts of the map (nodes) in an action-oriented way, to encourage creativity and breadth. The comprehensive map was developed to minimize the omitted variable bias. The colored concepts, which were selected as the key ones (Eden [25], Montibeller and Belton [92]), were the basis for the four fundamental objectives in the value tree (Figure 2): international trade, burden of disease, vulnerabilities from food consumption, and vulnerabilities from food production.

3.1.2. Biases in the Identification of Policy Options. The identification of potentially good alternatives is crucial for policy analysis, as the choice cannot be better than the best option under evaluation (Keeney [62]). However, public choices are, unfortunately, often focused on a single alternative against the status quo (Eisenhardt [29], Nutt [102]).

There is evidence of widespread *omitted variable bias*, in which relevant decision alternatives are not included in the analysis or even generated (Butler and Scherer [15], Jungermann et al. [57],

Figure 5. (Color online) Causal map for the low-moisture food experts for FAO.



Pitz et al. [109]). This issue might be often due to a *myopic problem representation*, where the problem is excessively constrained, preventing policy makers from contemplating potentially valuable decision alternatives (Payne et al. [106], Russo and Schoemaker [119]). Other causes might be *anchoring*, when all the options being generated are anchored on the initial set of decision alternatives (Keeney [62]), and *availability bias*, when the existing decision alternatives prevent creative thinking about new options (Del Missier et al. [21]).

Two motivational biases may also prevent the generation or inclusion of potentially valuable policy options. The *desirability of options bias* may lead to the exclusion of other decision alternatives that compete with a "pet" option (Montibeller and von Winterfeldt [94]). The *affect influenced bias* (Finucane et al. [38], Slovic et al. [125]) may cause the inclusion of decision alternatives that cause positive feelings and/or the exclusion of those that generate negative feelings.

Some very useful prescriptive advice on how to generate decision alternatives is available (Gregory and Keeney [47], Keeney [62, 64], Keller and Ho [71], Siebert and Keeney [122]). These guidelines can be classified into objective-based probes, in which one objective is presented at a time, sparking the generation of decision alternatives that perform well on its achievement; state-based probes, in which one future state is presented at a time, promoting the search for high-performing decision alternatives under this state; and alternative-based probes, in which an ideal decision alternative is employed to support the generation of good alternatives. In addition, Spetzler et al. [126] provide a useful list of properties for the set of decision alternatives. There is some behavioral evidence that the first type of probe generates not only more but also better decision alternatives (Butler and Scherer [15], Siebert and Keeney [122]).

Problem-structuring tools, such as causal maps (Belton et al. [11], Montibeller and Belton [92], Montibeller et al. [96]), the strategic choice approach (Friend [44]), and strategy generation tables (Howard [49]) may also be employed to generate decision alternatives. They are particularly useful to support the development of complex policy options. See also Franco and Montibeller [41] and Marttunen et al. [88] for further guidance on how problem-structuring methods may support structuring multicriteria models.

3.1.3. Biases in the Definition of Attributes. The definition of attributes may also suffer from some bias. Attributes might be affected by *scaling biases* (Poulton [111, 112]), a family of biases that occur when stimulus (x_{ij} in Figure 1) and response (v_{ij} in Figure 1) scales are mismatched and caused by different ways of presenting and scaling the attribute, as well as by the upper and lower limits of its scale. Other biases that might affect attributes are the *gain–loss bias* and its role in the framing effects—that is, if a performance is perceived as a gain or as a loss against a reference point (Levin et al. [84, 85])—as well as the *proxy bias*, which may distort weights of this type of attribute (Fischer et al. [39]).

Natural attributes, which measure directly the achievement of decision alternatives, should be employed whenever possible (e.g., billions of U.S. dollars for assessing the impact of international trade in Table 1). Attributes should have a range that encompasses the spread of performances of the policy options. If natural attributes are not available, then constructed attributes should be carefully developed (e.g., the attribute for burden of disease in Table 2, with a range from 18.42 to 136.44 DALYs) in a way that biases are minimized. Keeney and Gregory [66] provide excellent advice for the development of high-quality attributes.

3.1.4. Biases in the Elicitation of Value or Utility Functions. As mentioned before, value functions are employed for riskless decisions, whereas utility functions should be defined for decisions under uncertainty. Well-developed elicitation protocols for each type of function are available (Farquhar [32], von Winterfeldt and Edwards [132]), with value functions requiring judgments over outcomes or decision alternatives and utility functions requiring judgments over lotteries. There is more behavioral evidence available for biases affecting utility elicitation than for value elicitation.

Utility elicitations are noisy; that is, repetitive elicitations from the same subject may lead to slightly different functions (Hey et al. [48]) and are also affected by *anchoring* and *gain–loss biases*. The *certainty effect* (Allais [4], Kahneman and Tversky [58]), in which decision makers prefer "sure things" to gambles with similar expected utility, also may affect the shape of utility functions. Two further motivational biases might have an impact on value and utility functions: the *desirability of options bias*, which may distort the function in a way that favors the preferred decision alternative, and the *affect influenced bias*, which may trigger an oversensitivity to increases in a consequence.

Practitioners often adopt simplified forms of elicitation,[1] particularly as values are constructed instead of discovered (Slovic [124]). These involve using value functions as a proxy for utility ones (von Winterfeldt and Edwards [132]), deriving utility functions from value functions (Keeney and von Winterfeldt [68]), and adopting standardized shapes for the functions (McNamee and Celona [89]).

A recent project that we conducted for the Pan-American Health Organization (PAHO)/WHO assessing the capability of health systems of different countries against dog-mediated human rabies illustrates the relevance of the reference point in building prescriptive models and of the gain–loss bias (see also Del Rio Vilas et al. [22]). (Rabies is a deadly and neglected disease that affects disproportionally poor regions and disadvantaged communities.)

We developed for PAHO/WHO a decision support system (DSS) that assesses the multiple capabilities of a health system using a multiattribute value model. The system can be used to identify the highest value-for-money action for capability building, if additional resources were available. This DSS can also be employed for identifying which capabilities should be reduced if budget cuts are needed in a way that minimizes the loss of overall capability. The value function in Figure 6 shows the median value function that we elicited from six top international experts on rabies for both capability gains (G) and capability losses (L) for the attribute percentage of postexposure to the virus prophylaxis (PEP). Notice the shape is not the same for both functions, with gains being more convex than losses. These two functions are switched accordingly in the DSS depending on the prioritization (capability building or capability maintenance) that is being conducted.

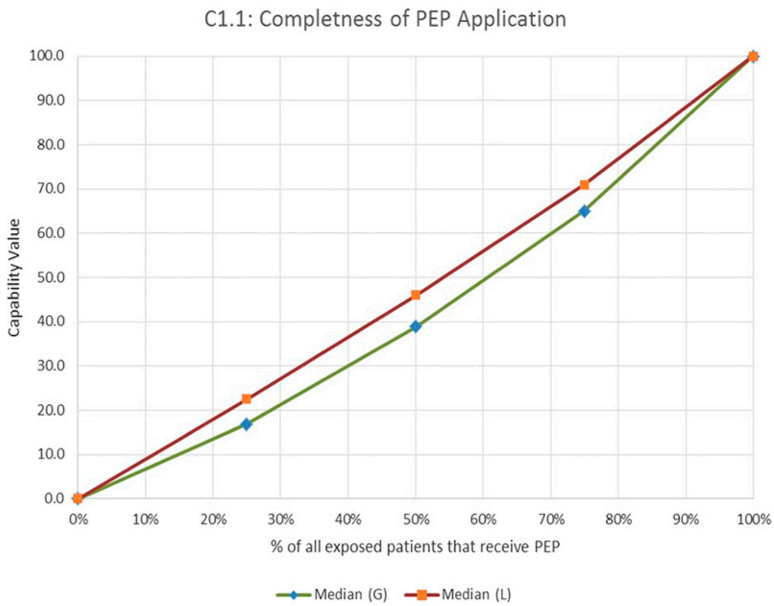
3.1.5. Biases in the Elicitation of Criteria Weights.

Weights in MAUT-based models are scaling constants that represent value trade-offs and are employed to aggregate partial values of each policy option. Well-developed protocols for their elicitation are available (von Winterfeldt and Edwards [132]), which stress that these judgments must consider the range of each attribute. Unfortunately, mistakes are frequently made in eliciting this type of parameter (Keeney [63]), with the most common one being the assumption that weights represent "direct importance."

Several cognitive biases affect the elicitation of weights. The *splitting bias* drives heavier weights to areas of the value tree that have more subcriteria (i.e., to criteria that were more decomposed) (Pöyhönen et al. [113], Weber et al. [138]). The *gain–loss bias* may also affect weights if trade-offs are elicited as improvement or degradation of performances (Weber and Borcherding [137]). The proxy bias might also affect weights as proxy attributes get overweighted. A serious bias is the *range insensitivity bias* (von Nitzsch and Weber [130]), which may lead to highly distorted weights if policy makers disregard the range of the attributes. The *equalizing bias* may drive decision makers to provide similar weights to all criteria.

Although I have experienced the equalizing bias in consultancy projects, a detailed recent review has not found further evidence that it has happened (Marttunen et al. [87]). An example of the equalizing bias occurred in a project for DEFRA, in which we developed a multicriteria-based decision support system for the prioritization of animal health threats (see Del Rio Vilas et al. [23]). The elicitation of swing weights from policy makers for the four impacts that mattered (impacts on public health, animal welfare, wider society, and international trade) showed that they were clearly different, with the first one receiving a heavier weight. However,

Figure 6. (Color online) Value for gains (G) and losses (L) for the rabies capability assessment for PAHO/WHO.



in the final version of the tool, which has been used by DEFRA since 2009 for assessing such threats and for supporting its policy recommendations, the four impact criteria were set with the same weight (25%), as it would have been politically difficult to justify different weights for the four impacts.

Two motivational biases affect the elicitation of weights. The *desirability of options bias* may lead policy makers to over- or underweight some criteria to favor a "pet" option (von Winterfeldt [131]). The *affect influenced bias* may cause a distortion of weights in favor of attributes that generate positive feelings or, conversely, against those that provoke negative ones.

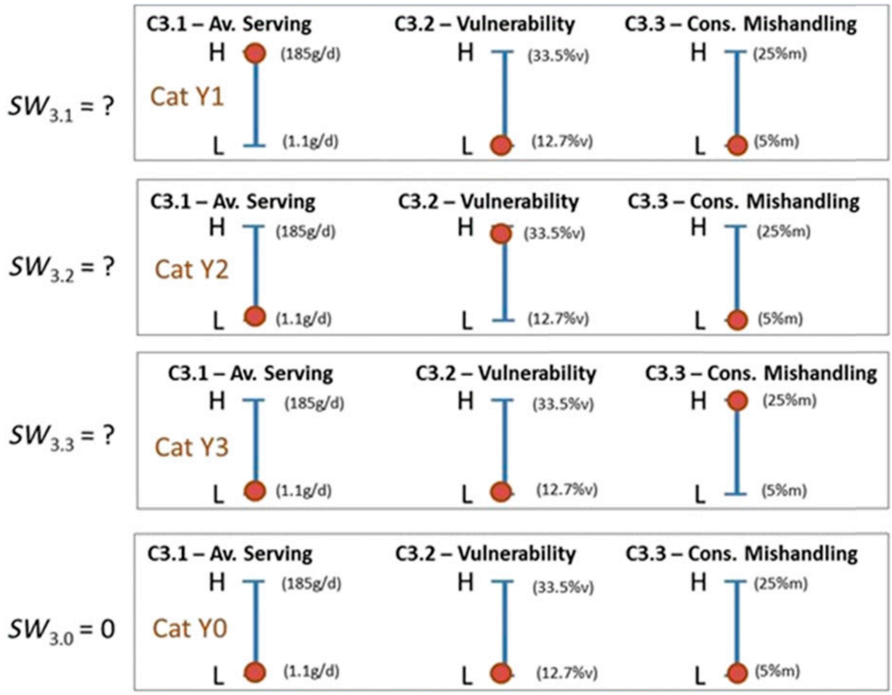
Several best practices should be used for the elicitation of weights. The use of elicitation methods that explicitly use the ranges of the attributes should always be employed in the elicitation of weights, such as in the swing weights method (see Belton and Stewart [10]). Cross-checking elicited values with selected trade-offs can help in alleviating biases and constructing preferences. A balanced value tree, with a similar level of decomposition along its branches, avoids the splitting bias. The use of natural or constructed attributes eliminates the proxy bias.

For example, in the low-moisture food category ranking for FAO/WHO, we employed the protocol shown in Figure 7 to elicit swing weights for the subcriteria associated with C3 (vulnerabilities from food consumption) and avoid the range insensitivity bias. It presents hypothetical food categories (Y0, Y1, Y2, and Y3) with one impact at the highest level on one attribute and the other impacts at the lowest level on the remaining attributes. The policy makers were then asked to rank these hypothetical food categories, scoring the highest impact (category) as 100, Y0 as 0, and the other impacts (categories) proportionally to this first one.

3.2. Cognitive Limitations in Eliciting Preferences

Multicriteria models based on MAUT present high demand for clear preferences: only strict preference and indifference relations can be modeled, and quantitative preference statements such as value/utility functions and criteria weights are needed. But policy makers are

Figure 7. (Color online) Elicitation of swing weights for the low-moisture food project for FAO/WHO.



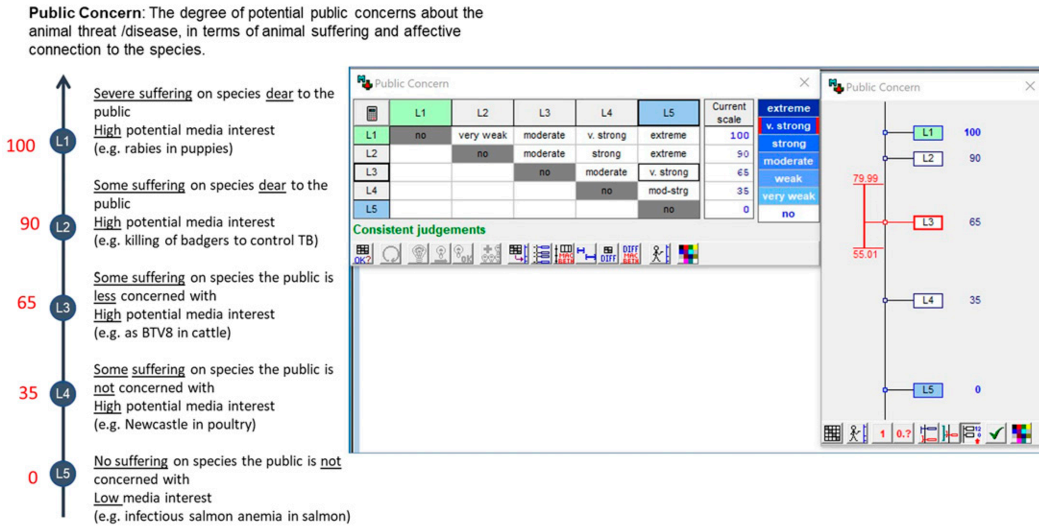
boundedly rational (Simon [123]) with intrinsic cognitive limitations in thinking about and expressing their preferences (Larichev [79]). This means that elicitation protocols might have to be adapted to accommodate such limitations. Indeed, many of the interesting developments in Europe in multicriteria analysis have been exactly trying to address these limitations, with a larger number of preference relations being available to decision makers (Figueira et al. [37]), methods that try to reduce cognitive demands (Larichev and Brown [80]), and several methods that can deal with incomplete information (Weber [136]).

A full review of these models for incomplete information is beyond the scope of this tutorial. However, it is worth mentioning that eliciting value functions, utility functions, and numerical swing weights are all classified as cognitively complex parameters (Larichev [79], Olson et al. [103]). On the other hand, ordinal preference information is cognitively less complex and more stable (Moshkovich et al. [101]).

Therefore, methods that can translate ordinal judgments into cardinal parameters (see Krantz et al. [76]) might be useful in cases where the policy makers are unable to understand clearly, or express their preferences, on cardinal scales. Particularly useful are the methods for the ordinal ranking of weights (for a review, see Alfares and Duffuaa [3]) and those for eliciting value functions with user-friendly protocols, such as MACBETH (Bana e Costa et al. [6]).

In the project on ranking emerging animal health threats for DEFRA (Del Rio Vilas et al. [23]), we employed MACBETH to help the group construct a value function. Figure 8 shows the constructed attribute "public concern" on the left, with five levels (L1–L5). Policy makers were asked to make qualitative pairwise comparisons between the difference of value between each pair of levels (e.g., "very weak" between L1 and L2). The software then calculates a value function that accommodates these constraints, as shown on the right of the same figure, and provides a value for each level (which can be fine-tuned within a boundary as shown for L3).

Figure 8. (Color online) Elicitation of a value function for the project on evaluating animal health threats for DEFRA.



4. Key Challenge II: Expert Judgment in Multicriteria Policy Analysis

The influential movement for evidence-based policy making since the 1990s (see Tsoukias et al. [128] for details) calls for the use of the best evidence available about the possible benefits and costs of each policy alternative under consideration. In a value-driven decision-making framework (Keeney [62]), this means identifying the impacts of each policy alternative on each attribute ($x_{ij}(.)$ in Figure 1).

Whenever evidence is available or predictive models can be developed, they should be employed to estimate such impacts. If that is not the case, then expert judgment may be used to estimate such consequences. This typically takes the form of either a continuous or discretized distribution of impacts, for a given policy option, over the attribute x_{ij}.

An extensive literature on expert elicitation in policy analysis is available (for a good introduction, see Morgan [98] and Morgan and Henrion [99]), and a plethora of biases that affect judgments about uncertainty are known (Kahneman and Tversky [59], Kahneman et al. [60]). As mentioned in the previous section, Montibeller and von Winterfeldt [94] suggest that only a relatively small number are relevant in this context. I briefly review the biases that are relevant for the impact estimations of policy options and debiasing tools against them next.

4.1 Biases in the Elicitation of Expert Judgments on Policy Impacts

Several cognitive biases affect the assessment of probabilities for a given impact: how the variable is scaled may influence results due to *scaling biases*; *overconfidence* (Klayman et al. [75]), *anchoring*, and *availability biases* may make the range of the estimated impacts too narrow; and the *equalizing bias* may make the estimated probabilities of discretized events excessively similar.

In addition, several motivational biases may affect these judgments: the *desirability bias* may distort estimated probabilities of events that are desirable (Krizan and Windschitl [77]) or undesirable (Chapin [17]), which may also be caused by the *affect influenced bias*. In addition, the *desirability of options bias* may lead experts to over- or underestimate probabilities to favor their preferred policy option (Montibeller and von Winterfeldt [94]).

Different elicitation protocols may reduce biased estimates, with the fixed value method (in which a probability is elicited for a given value of the variable) producing less overconfidence (Abbas et al. [1], Seaver et al. [121]). Further decomposition of variables, multiple experts, and stretching the extremes (for instance, using counterfactuals or hypothetical bets) are usually employed in practice to improve accuracy, but our recent behavioral experiment has shown these stretching tools to be of limited efficacy (Ferretti et al. [36]). Scoring rules that reward accuracy (Winkler [139]) and hypothetical bets (Dias et al. [24]) on fractiles of the distribution may help in reducing motivational biases.

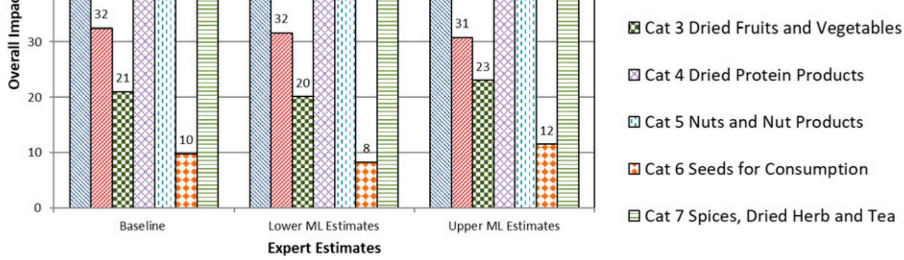
For the low-moisture food ranking project for FAO/WHO, we asked the experts individually to estimate lower, upper, and most likely values for the parameters of three subcriteria where their expert judgment was needed (C3.3, C4.1, and C4.2; see Figure 2). The aim was to minimize individual biases such as anchoring and availability. The ranges of the estimates provided by experts were very wide, signaling that the procedure may have reduced overconfidence. For this reason, we opted to use their most likely values in the analysis of robustness, as shown in Figure 9. Results show that Category 4 would be the one with the highest overall value under these lower most likely values and upper most likely value bounds, thus slightly above Category 1 (which has the highest overall value under the baseline most likely estimates).

4.2. Cognitive Limitations in the Elicitation of Expert Judgments on Policy Impacts

The same concern about cognitive limitations expressed in the previous section for preferences applies in the elicitation of expert judgments. Although many experts have advanced degrees in statistics or quantitative methods, some of them struggle to express their judgments in a quantitative way. Indeed, the prevalence of qualitative labels for describing uncertainties in an ambiguous way (Wallsten et al. [134]) might be an indication of this phenomenon. In addition, I noticed in several projects the experts' reluctance in expressing a quantitative judgment without having all hard evidence available, indicating a motivational issue.

These cognitive and motivational challenges have led Jaspersen and Montibeller [53] to develop a family of methods that use only ordinal information about estimated probabilities of

Figure 9. (Color online) Analysis of robustness of low-moisture food categories ranking for the FAO/WHO project (ML, most likely estimates).



events, while assuming the principle of maximum entropy (Jaynes [54]), to derive quantitative probability estimates for both discrete and continuous distributions. For example, Figure 10 illustrates the probability of up to six mutually exclusive and exhaustive binary events that were a priori ordinally ranked (where $p_{\underline{i}}$ is the rank of the ordered event, and n is the total number of events); this method was employed to estimate the probability of different emerging animal health threats for the project for DEFRA that I mentioned previously. See also Dias et al. [24] for a recent coverage of other elicitation protocols.

5. Key Challenge III: Facilitation of Groups in Multicriteria Policy Analysis

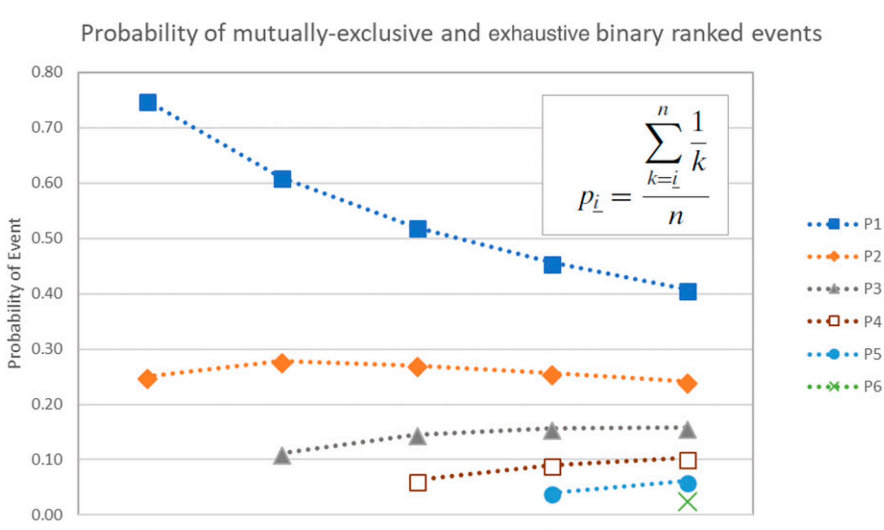
Most relevant policy decisions are taken in groups. Furthermore, it is common to assemble groups of experts when estimates are needed in public planning. In this section I review the advantages of working with groups as well as the key biases that groups might suffer from.

5.1. The Facilitation of Policy Makers' Groups

Groups have several advantages when compared with individual policy makers, particularly when the group composition favors diversity of perspectives and representativeness of different interests at stake in the decision. Groups enhance the pooling of relevant and distinctive information, they help in error checking and correction, they can enhance individual task motivation, and they can also improve satisfaction among members (Kerr and Tindale [74]), as well as increase the commitment to the agreed way forward (Phillips [108]).

One important distinction of multicriteria decision analysis, when compared with other operations research methods, is that there is no single optimal solution as in optimization models (Franco and Montibeller [40], Phillips [108]) or a true value as in prediction problems. What decision analysts seek to develop are *requisite models* (Phillips [107]), which are not "perfect" but only rich enough to help the group solving the problem that it is dealing with. In addition, as mentioned earlier, preferences are constructed instead of discovered; that is, they are developed during the decision-aiding process instead of predefined in the policy makers' minds (Roy [118], Slovic [124]). This type of intervention aim requires an effective management of decision-making processes, which promotes information sharing among members

Figure 10. (Color online) Probability of events using rank-order judgments (Jaspersen and Montibeller [53]).



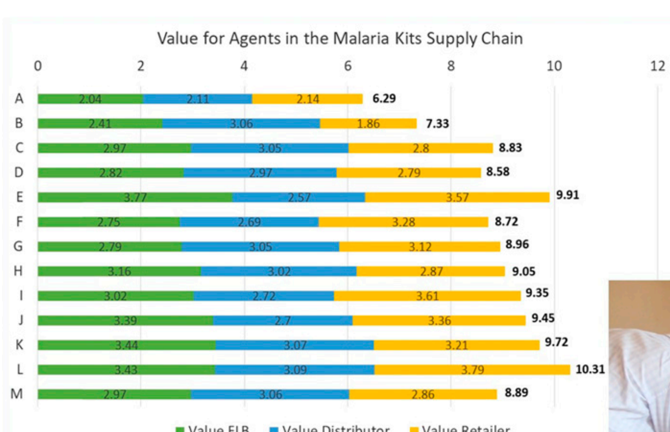
of the group, supports negotiation of their different perspectives and stakes, and enables the group to reach a joint agreement on the way forward.

A recent project for the Malaria Consortium/USAID, in collaboration with Massachusetts Institute of Technology's MIT Humanitarian Supply Chain Laboratory, illustrates the different perspectives that stakeholders often have and the importance of understanding their values to create high-value policy options (Gregory and Keeney [47], Keeney [62], Siebert and Keeney [122]). We developed a multicriteria value model to understand the preferences and priorities of the agents involved in a supply chain of rapid malaria test kits in Uganda (which involve first-line buyers (FLBs), distributors, and retailers). We used a low-tech workshop setting—with criteria defined from the literature and enriched from input from the participants during the workshops—and elicited value functions and swing weights for each agent of the supply chain (see also Keller et al. [72]). We employed this value model to evaluate existing malaria kits (A, B, C, and E) and develop new options (D and F–M). Figure 11 shows the setting of the workshop (on the right) and the overall results (on the left). Notice that option L had a higher value than the existing ones for the agents of the supply chain (10.31 units of value) when considering them equally relevant (equal weights on the agents). The results proved robust against variations of these weights on the agents (see Carland et al. [16] for details).

5.2. Group Biases and Debiasing

Group dynamics may present many dysfunctional behaviors and group biases. Relevant dysfunctional behaviors (Kerr and Tindale [73, 74]) encompass group pressure on individual members to conform to the majority view; groups that are dominated by a strong, authoritarian, leader; group inattention to novel or unshared information; dysfunctional shared representations about the decision problem; group motivation losses as a result of inefficiencies in the group decision process and disengagement of some members; and group coordination losses as a result of the strains that group work incurs when compared with individual tasks. Decision processes that are adequately facilitated and well-designed elicitation techniques, such as Delphi (Linstone and Turoff [86]), may help to minimize these behaviors in policy-making groups.

Figure 11. (Color online) Value of malaria kits for different agents in the supply chain for the Malaria Consortium/USAID (Carland et al. [16]).



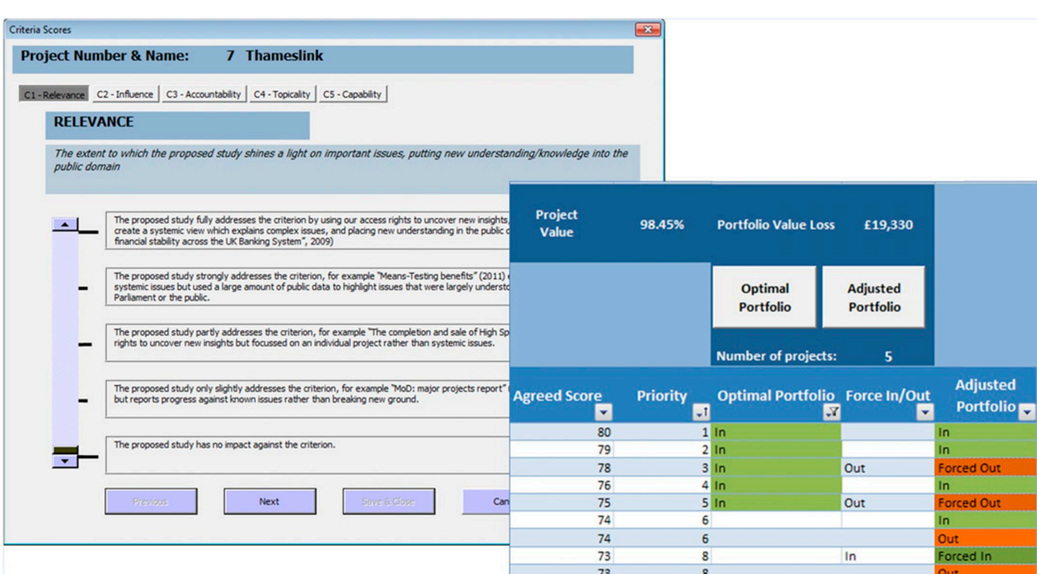
Source. MIT Center for Transportation and Logistics (photograph).

Relevant group biases need to be dealt with when dealing with groups, as there is evidence that they affect group judgments. Montibeller and von Winterfeldt [95] suggest five main biases, based on reviews of forecasting elicitation (Kerr and Tindale [74]) and of team-based decision making (Jones and Roelofsma [56]). The first bias is *false consensus*, in which an individual group member overestimates the similarities between his or her judgment and the other members, leading to judgments using incorrect assumptions about the decision problem (Ross et al. [116]). The second bias is *groupthink* (Janis [51]), in which members of very cohesive groups are focused on getting consensus on the agreed solution, disregarding other decision alternatives, objectives, or limiting information search (Jones and Roelofsma [56]). The third one is *group polarization* (Lamm [78]), in which group discussions enhance the initial position or opinion of the majority of its members. This bias may affect the choice, if the group had an initial inclination for a given policy option, as well as the group's risk attitude, which may become more risk seeking/averse than those of its individual members as discussions progress (Isenberg [50]). Another bias is *group escalation of commitment*, in which a group supports a course of action that is clearly failing, which is exacerbated by groupthink and group polarization. Finally, groups may suffer from *group overconfidence*, beyond the overconfidence of its individual members (Plous [110]).

The available debiasing tools against group biases encompass the use of multiple experts with different perspectives of the policy problem, encouragement of expressions of multiple perspectives and opinions, use of structured elicitation protocols, and employment of facilitated decision processes.

A project that we developed for the UK National Audit Office (NAO) illustrates how the redesigning of a decision process, coupled with a multicriteria evaluation, helped to minimize group biases and to support the reach of agreement on the options to be selected. We developed a model to evaluate the NAO's value of value-for-money auditing studies. It is surprisingly hard to assess the potential of these studies a priori, so we developed qualitative attributes for the criteria that mattered for the policy makers: the *relevance* of the study, the potential *influence* that it might bring, the *accountability* that it might generate, the *topicality* of the theme, and the *capability* of the auditing team for performing the study. The attribute for relevance is shown on the left-hand side of Figure 12 (with the levels describing the 100, 75, 50,

Figure 12. (Color online) Supporting the selection of a portfolio of value-for-money auditing projects for NAO.



25, and 0 value scores). Swing weights were elicited from the group so that an overall value for each study could be assessed. The evaluation process was redesigned in three main steps. First, individual assessors evaluated each study on each criterion (avoiding group biases) using the levels shown in the same figure (thus reducing anchoring and response scale biases). Second, aggregated scores were calculated, and then the group reviewed and reassessed scores for projects with a high dispersion of individual valuations (supporting the share of information and opinions but also constraining motivational biases for "pet" projects). Third, projects were prioritized based on their overall value, but the group could force projects in and out (balancing technical criteria and political feasibility) while being informed of the value loss of the adjusted portfolio (as shown on the right of Figure 12).

6. Conclusions

The choice of public policies is necessarily a multicriteria problem of balancing and making trade-offs among conflicting objectives; that is why it is so challenging for policy makers and so attractive for decision analysts. Multicriteria decision analysis is being employed extensively, and successfully, worldwide to support such tough choices. But some key behavioral challenges must be addressed if we want to develop high-quality decision models and provide top-quality decision support. This tutorial provided an overview of these challenges and gave some advice on how to deal with them.

The first theme is that these models rely heavily on judgments, preferences, and priorities from policy makers and estimates from experts. The quality of the model is therefore closely dependent on the quality of these judgments. Several cognitive and motivational biases thus must be minimized when decision analysts are eliciting these judgments. The tutorial provided several suggestions on how to debias these judgments in practice and the relevant literature on the topic.

The second theme, which is less visible in behavioral decision research but crucial for successful interventions of multicriteria policy analysis, is managing cognitive complexity. This is particularly true whenever there is a group of diverse stakeholders where not all of them have a quantitative background (for instance, the malaria kit evaluation that we conducted for the Malaria Consortium/USAID; see Carland et al. [16]). Solutions to this challenge call for simple interfaces, user-friendly elicitation protocols, and incomplete information models.

The third theme, which is perhaps the trickiest one to address because of its sociotechnical nature, is how to maximize the efficiency of team work and minimize group biases in judgments. Solutions require the use of effective facilitated decision modeling and (re)design of decision processes.

These three themes open exciting avenues for further research. The first one is on assessing the effectiveness of existing debiasing tools and developing, if needed, more efficient ones as argued by Montibeller and von Winterfeldt [94] (e.g., the analysis of effectiveness of widely employed debiasing tools against overprecision conducted by Ferretti et al. [36]). The second avenue is more research on the MCDA model–user interface and rigorous behavioral assessment of cognitive complexity in preference elicitation and matches with individual styles (e.g., the analysis conducted by Fasolo and Bana e Costa [33] on different value elicitation methods). The third one is a better understanding of the interaction between model and group dynamics in MCDA interventions (e.g., as conducted by Franco et al. [42], who were supporting groups in applying value-focused thinking for the allocation of resources).

I conclude this tutorial with two worries and one encouragement. I worry that decision analysts are spending far too much effort and time in dealing with the (undoubtedly important) technical issues related to multicriteria evaluation while assuming that the inputs of their models (i.e., judgments) are debiased and reliable. *They are not.* I also worry that decision analysts often assume that policy makers are a single unit in their analyses—"the decision maker," with stable, unique, and clear preferences. *They are not.* So I encourage decision analysts to "get their hands dirty" with policy makers' judgments and group dynamics. Learn how they can debias the former and facilitate the latter. This will help our community to become better practitioners and get involved in even more complex policy problems, and it may also open up exciting avenues for interdisciplinary research.

Acknowledgments

The author is grateful to Detlof von Winterfeldt, because the tutorial uses the classification of biases and debiasing tools that both developed together. The author thanks the intellectual contributions to his understanding of the issues discussed here from Leonardo Ensslin on policy evaluations, Valerie Belton on problem structuring, Carlos Bana e Costa and Ralph Keeney on multicriteria model structuring, and L. Alberto Franco and Larry Phillips on facilitated modeling. Last but not the least, the author is grateful to Valentina Ferretti for moving forward with their research agenda, together with Detlof von Winterfeldt, on debiasing effectiveness. The author also thanks Douglas Shier and an anonymous referee for the detailed feedback and suggestions, which helped in improving the draft. Any mistakes or omissions, and all opinions, are the author's own responsibility.

Endnote

[1]However, an unfortunate trend in multicriteria decision analysis has been the direct elicitation of relative judgments of value for the alternatives, with often poorly defined attributes and without assessment of their performances, from policy makers. Although this might be acceptable for private decisions, it does not provide a transparent link between evidence and values, which reduces the justifiability of public choices, and thus it should be avoided in my opinion.

References

[1] A. E. Abbas, D. V Budescu, H.-T. Yu, and R. Haggerty. A comparison of two probability encoding methods: Fixed probability vs. fixed variable values. *Decision Analysis* 5(4):190–202, 2008.

[2] M. Airoldi, A. Morton, J. A. E. Smith, and G. Bevan. STAR—People-powered prioritization: A 21st-century solution to allocation headaches. *Medical Decision Making* 34(8):965–975, 2014.

[3] H. K. Alfares and S. O. Duffuaa. Assigning cardinal weights in multi-criteria decision making based on ordinal ranking. *Journal of Multi-Criteria Decision Analysis* 15(5–6):125–133, 2008.

[4] M. Allais. Le comportement de l'Homme Rationnel devant le risque: Critique des postulats et axiomes de l'Ecole Americaine. *Econometrica* 21(4):503–546, 1953.

[5] H. R. Arkes, C. González-Vallejo, A. J. Bonham, Y. Kung and N. Bailey. Assessing the merits and faults of holistic and disaggregated judgments. *Journal of Behavioral Decision Making* 23(3): 250–270, 2010.

[6] C. A. Bana e Costa, J.-M. De Corte, and J.-C. Vansnick. MACBETH. *International Journal of Information Technology & Decision Making* 11(2):359–387, 2012.

[7] C. A. Bana e Costa, C. S. Oliveira, and V. Vieira. Prioritization of bridges and tunnels in earthquake risk mitigation using multicriteria decision analysis: Application to Lisbon. *Omega* 36(3):442–450, 2008.

[8] A. Barcus and G. Montibeller. Supporting the allocation of software development work in distributed teams with multi-criteria decision analysis. *Omega* 36(3):464–475, 2008.

[9] V. Belton and J. Pictet. A framework for group decision using a MCDA model: Sharing, aggregating or comparing individual information? *Journal of Decision Systems* 6(3):283–303, 1997.

[10] V. Belton and T. J. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach.* Springer, Norwell, MA, 2002.

[11] V. Belton, F. Ackermann, and I. Shepherd. Integrated support from problem structuring through to alternative evaluation using COPE and V·I·S·A. *Journal of Multi-Criteria Decision Analysis* 6(3):115–130, 1997.

[12] S. D. Bond, K. A. Carlson, and R. L. Keeney. Generating objectives: Can decision makers articulate what they want? *Management Science* 54(1):56–70, 2008.

[13] S. D. Bond, K. A. Carlson, and R. L. Keeney. Improving the generation of decision objectives. *Decision Analysis* 7(3):238–255, 2010.

[14] V. J. Brookes, M. Hernández-Jover, B. Cowled, P. K. Holyoake, and M. P. Ward. Building a picture: Prioritisation of exotic diseases for the pig industry in Australia using multi-criteria decision analysis. *Preventive Veterinary Medicine* 113(1):103–117, 2014.

[15] A. B. Butler and L. L. Scherer. The effects of elicitation aids, knowledge, and problem content on option quantity and quality. *Organizational Behavior and Human Decision Processes* 72(2): 184–202, 1997.

[16] C. Carland, J. Goentzel, and G. Montibeller. Modeling the values of private sector agents in multi-echelon humanitarian supply chains. *European Journal of Operational Research* 269(2):532–543, 2018.

[17] J. Chapin. Self-protective pessimism: Optimistic bias in reverse. *North American Journal of Psychology* 3(2):253–262, 2001.

[18] R. Cox, J. Sanchez, and C. W. Revie. Multi-criteria decision analysis tools for prioritising emerging or re-emerging infectious diseases associated with climate change in Canada. *PLoS ONE* 8(8): e68338, 2013.

[19] K .A. Daniell, A. Morton, and D. R. Insua. Policy analysis and policy analytics. *Annals of Operations Research* 236(1):1–13, 2015.

[20] A. P. de Geus. Planning as learning. *Harvard Business Review* 66(2):70–74, 1988.

[21] F. Del Missier, D. Ferrante, and E. Costantini. Focusing effects in predecisional information acquisition. *Acta Psychologica* 125(2):155–174, 2007.

[22] V. J. Del Rio Vilas, A. Burgeño, G. Montibeller, A. Clavijo, M. A. Vigilato, and O. Cosivi. Prioritization of capacities for the elimination of dog-mediated human rabies in the Americas: Building the framework. *Pathogens and Global Health* 107(7):340–345, 2013.

[23] V. J. Del Rio Vilas, F. Voller, G. Montibeller, L. A. Franco, S. Sribhashyam, E. Watson, M. Hartley, and J. C. Gibbens. An integrated process and management tools for ranking multiple emerging threats to animal health. *Preventive Veterinary Medicine* 108(2–3):94–102, 2013.

[24] L. C. Dias, A. Morton, and J. Quigley, eds. *Elicitation—The Science and Art of Structuring Judgement*. Springer, Cham, Switzerland, 2018.

[25] C. Eden. Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research* 159(3):673–686, 2004.

[26] C. Eden and F. Ackermann. Group decision and negotiation in strategy making. *Group Decision and Negotiation* 10(2):119–140, 2001.

[27] C. Eden and F. Ackermann. Problem structuring: On the nature of, and reaching agreement about, goals. *EURO Journal Decision Processes* 1(1–2):7–28, 2013.

[28] F. Eisenführ, M. Weber, and T. Langer. *Rational Decision Making*, 1st ed., Springer, Berlin, 2010.

[29] K. M. Eisenhardt. Making fast strategic decisions in high-velocity environments. *Academy of Management Journal* 32(3):543–576, 1989.

[30] P. L. Ewing Jr., W. Tarantino, and G. S. Parnell. Use of decision analysis in the Army Base Realignment and Closure (BRAC) 2005 military value analysis. *Decision Analysis* 3(1):33–49, 2006.

[31] FAO/WHO. Ranking of low moisture foods in support of microbiological risk management. Report, Food and Agriculture Organization of the United Nations/World Health Organization, Rome/Geneva, 2014.

[32] P. H. Farquhar. Utility assessment methods. *Management Science* 30(11):1283–1300, 1984.

[33] B. Fasolo and C. A. Bana e Costa. Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words. *Omega* 44(April):83–90, 2014.

[34] V. Ferretti. A multicriteria spatial decision support system development for siting a landfill in the province of Torino (Italy). *Journal of Multi-Criteria Decision Analysis* 18(5–6):231–252, 2011.

[35] V. Ferretti and A. Degioanni. How to support the design and evaluation of redevelopment projects for disused railways? A methodological proposal and key lessons learned. *Transportation Research Part D: Transport and Environment* 52(May):29–48, 2017.

[36] V. Ferretti, S. Guney, G. Montibeller, and D. von Winterfeldt. Testing best practices to reduce the overconfidence bias in multi-criteria decision analysis. T. X. Bui and R. H. Sprague, Jr., eds. *Hawaii International Conference on Systems Sciences* (*HICSS*), IEEE, Piscataway, NJ, 1547–1555, 2016.

[37] J. R. Figueira, S. Greco, B. Roy, and R. Słowiński. ELECTRE methods: Main features and recent developments. C. Zopounidis and P. Pardalos, eds. *Handbook of Multicriteria Analysis*. Springer, Berlin, 51–89, 2010.

[38] M. L. Finucane, A. Alhakami, P. Slovic, and S. M. Johnson. The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making* 13(1):1–17, 2000.

[39] G. W. Fischer, N. Damodaran, K. B. Laskey, and D. Lincoln. Preferences for proxy attributes. *Management Science* 33(2):198–214, 1987.

[40] L. A. Franco and G. Montibeller. Facilitated modelling in operational research. *European Journal of Operational Research* 205(3):489–500, 2010.

[41] L. A. Franco and G. Montibeller. Problem structuring for multicriteria decision analysis interventions. J. J. Cochran, L. A. Cox Jr., P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, eds. *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Hoboken, NJ, 2011.

[42] L. A. Franco, E. A. J. A. Rouwette, and H. Korzilius. Different paths to consensus? The impact of need for closure on model-supported group conflict management. *European Journal of Operational Research* 249(3):878–889, 2016.

[43] S. French. *Decision Theory: An Introduction to the Mathematics of Rationality*. Ellis Horwood, Chichester, UK, 1989.

[44] J. Friend. The strategic choice approach. J. J. Cochran, L. A. Cox Jr., P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, eds. *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Hoboken, NJ, 2011.

[45] S. Greco, M. Ehrgott, and J. R. Figueira, eds. *Multiple Criteria Decision Analysis: State of the Art Surveys*, 2nd ed. Springer, New York, 2016.

[46] R. S. Gregory. Incorporating value trade-offs into community-based environmental risk decisions. *Environmental Values* 11(4):461–488, 2002.

[47] R. S. Gregory and R. L. Keeney. Creating policy alternatives using stakeholder values. *Management Science* 40(8):1035–1048, 1994.

[48] J. D. Hey, A. Morone, and U. Schmidt. Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty* 39(3):213–235, 2009.

[49] R. A. Howard. Decision analysis: Practice and promise. *Management Science* 34(6):679–695, 1988.

[50] D. J. Isenberg. Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology* 50(6):1141–1151, 1986.

[51] I. L. Janis. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Houghton Mifflin, Boston, 1983.

[52] P. A. Jargowsky. Omitted variable bias. K. Kempf-Leonard, ed. *Encyclopedia of Social Measurement*, Vol. 2. Elsevier, New York, 919–924, 2005.

[53] J. G. Jaspersen, and G. Montibeller. Probability elicitation under severe time pressure: A rank-based method. *Risk Analysis* 35(7):1317–1335, 2015.

[54] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review* 106(4):620–630, 1957.

[55] P. N. Johnson-Laird. Mental models and deduction. *Trends in Cognitive Sciences* 5(10):434–442, 2001.

[56] P. E. Jones and P. H. M. P. Roelofsma. The potential for social contextual and group biases in team decision-making: Biases, conditions and psychological mechanisms. *Ergonomics* 43(8):1129–1152, 2000.

[57] H. Jungermann, I. von Ulardt, and L. Hausmann. The role of the goal for generating actions. P. Humphreys, O. Svenson, and A. Vári, eds. *Analysing and Aiding Decision Processes*. Elsevier, New York, 223–236, 1983.

[58] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291, 1979.

[59] D. Kahneman and A. Tversky. *Choices, Values, and Frames*. Cambridge University Press, Cambridge, UK, 2000.

[60] D. Kahneman, P. Slovic, and A. Tversky. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 1982.

[61] R. L. Keeney. Decision analysis: An overview. *Operations Research* 30(5):803–838, 1982.

[62] R. L. Keeney. *Value-Focused Thinking: A Path to Creative Decisionmaking*. Harvard University Press, Cambridge, MA, 1996.

[63] R. L. Keeney. Common mistakes in making value trade-offs. *Operations Research* 50(6):935–945, 2002.

[64] R. L. Keeney. Value-focused brainstorming. *Decision Analysis* 9(4):297–382, 2012.

[65] R. L. Keeney. Identifying, prioritizing, and using multiple objectives. *EURO Journal on Decision Processes* 1(1–2):45–67, 2013.

[66] R. L. Keeney and R. S. Gregory. Selecting attributes to measure the achievement of objectives. *Operations Research* 53(1):1–11, 2005.

[67] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, 2nd ed. Cambridge University Press, Cambridge, UK, 1993.

[68] R. L. Keeney and D. von Winterfeldt. Practical value models. W. Edwards, R. F. Miles, and D. Von Winterfeldt, eds. A*dvances in Decision Analysis: From Foundations to Applications*. Cambridge University Press, New York, 232–252, 2007.

[69] G. L. Keeney and D. von Winterfeldt. Identifying and structuring the objectives of terrorists. *Risk Analysis* 30(12):1803–1816, 2010.

[70] R. L. Keeney and D. von Winterfeldt. A value model for evaluating homeland security decisions. *Risk Analysis* 31(9):1470–1487, 2011.

[71] L. R. Keller and J. L. Ho. Decision problem structuring: Generating options. *IEEE Transactions on Systems, Man, and Cybernetics* 18(5):715–728, 1988.

[72] L. R. Keller, J. Simon, and Y. Wang. Multiple-objective decision analysis involving multiple stakeholders. M. R. Oskoorouchi, ed. *Decision Technologies and Applications*, INFORMS TutORials in Operations Research. INFORMS, Hanover, MD, 139–155, 2009.

[73] N. L. Kerr and R. S. Tindale. Group performance and decision making. *Annual Review of Psychology* 55:623–655, 2004.

[74] N. L. Kerr and R. S. Tindale. Group-based forecasting?: A social psychological analysis. *International Journal of Forecasting* 27(1):14–40, 2011.

[75] J. Klayman, J.B. Soll, C. González-Vallejo, and S. Barlas. Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes* 79(3):216–247, 1999.

[76] D. H. Krantz, R. D. Luce, P. Suppes, and A. Tversky. *Foundations of Measurement*. Academic Press, New York, 1971.

[77] Z. Krizan and P. D. Windschitl. The influence of outcome desirability on optimism. *Psychological Bulletin* 133(1):95–121, 2007.

[78] H. Lamm. A review of our research on group polarization: Eleven experiments on the effects of group discussion on risk acceptance, probability estimation, and negotiation positions. *Psychological Reports* 62(3):807–813, 1988.

[79] O. I. Larichev. Cognitive validity in design of decision-aiding techniques. *Journal of Multi-Criteria Decision Analysis* 1(3):127–138, 1992.

[80] O. I. Larichev and R. V. Brown. Numerical and verbal decision analysis: Comparison on practical cases. *Journal of Multi-Criteria Decision Analysis* 9(6):263–273, 2000.

[81] P. Legrenzi and V. Girotto. Mental models in reasoning and decision making. J. Oakhill and A. Garnham, eds. *Mental Models in Cognitive Science: Essays in Honour of Phil Johnson-Laird*, Psychology Press, Hove, UK, 95–118, 1996.

[82] P. Legrenzi, V. Girotto, and P. N. Johnson-Laird. Focussing in reasoning and decision making. *Cognition* 49(1–2):37–66, 1993.

[83] O. G. León. Value-focused thinking vs. alternative-focused thinking: Effects on generation of objectives. *Organizational Behavior and Human Decision Processes* 80(3):213–227, 1999.

[84] I. P. Levin, S. L. Schneider, and G. J. Gaeth. All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes* 76(2): 149–188,1998.

[85] I. P. Levin, G. J. Gaeth, J. Schreiber, and M. Lauriola. A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes* 88(1):411–429, 2002.

[86] H. A. Linstone and M. Turoff, eds. *The Delphi Method: Techniques and Applications*. Addison-Wesley, Reading, MA, 1975.

[87] M. Marttunen, V. Belton, and J. Lienert. Are objectives hierarchy related biases observed in practice? A meta-analysis of environmental and energy applications of multi-criteria decision analysis. *European Journal of Operational Research* 265(1):178–194, 2018.

[88] M. Marttunen, J. Lienert, and V. Belton. Structuring problems for multi-criteria decision analysis in practice: A literature review of method combinations. *European Journal of Operational Research* 263(1):1–17, 2017.

[89] P. McNamee and J. Celona. *Decision Analysis for the Professional*, 4th ed. SmartOrg, Menlo Park, CA, 2005.

[90] M. W. Merkhofer and R. L. Keeney. A multiattribute utility analysis of alternative sites for the disposal of nuclear waste. *Risk Analysis* 7(2):173–194, 2006.

[91] J. Mingers and J. Rosenhead. *Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. John Wiley & Sons, West Sussex, UK, 2001.

[92] G. Montibeller and V. Belton. Causal maps and the evaluation of decision options—A review. *Journal of Operational Research Society* 57(7):779–791, 2006.

[93] G. Montibeller and L. A. Franco. Decision and risk analysis for the evaluation of strategic options. F. O'Brien and R. Dyson, eds. *Supporting Strategy: Frameworks, Methods and Models*. John Wiley & Sons, Chichester, UK, 251–284, 2007.

[94] G. Montibeller and D. von Winterfeldt. Cognitive and motivational biases in decision and risk analysis. *Risk Analysis* 35(7):1230–1251, 2015.

[95] G. Montibeller and D. von Winterfeldt. Individual and group biases in value and uncertainty judgments. L. C. Dias, A. Morton, and J. Quigley eds. *Elicitation*. Springer, Cham, Switzerland, 377–392, 2018.

[96] G. Montibeller, L. A. Franco, E. Lord, and A. Iglesias. Structuring resource allocation decisions: A framework for building multi-criteria portfolio models with area-grouped options. *European Journal of Operational Research* 199(3):846–856, 2009.

[97] O. F. Morera and D. V. Budescu. A psychometric analysis of the "divide and conquer" principle in multicriteria decision making. *Organizational Behavior and Human Decision Processes* 75(3): 187–206, 1998.

[98] M. G. Morgan. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Science USA* 111(20):7176–7184, 2014.

[99] M. G. Morgan and M. Henrion. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, UK, 1992.

[100] A. Morton, M. Airoldi, and L. D. Phillips. Nuclear risk management on stage: A decision analysis perspective on the UK's Committee on Radioactive Waste Management. *Risk Analysis* 29(5): 764–779, 2009.

[101] H. M. Moshkovich, A. I. Mechitov, and D. L. Olson. Ordinal judgments in multiattribute decision analysis. *European Journal of Operational Research* 137(3):625–641, 2002.

[102] P. C. Nutt. How decision makers evaluate alternatives and the influence of complexity. *Management Science* 44(8):1148–1166, 1998.

[103] D. L. Olson, A. I. Mechitov, and H. M. Moshkovich. Cognitive effort and learning features of decision aids: Review of experiments. *Journal of Decision Systems* 7(1–4):129–146, 1998.

[104] G. S. Parnell, H. W. Conley, J. A. Jackson, L. J. Lehmkuhl, and J. M. Andrew. Foundations 2025: A value model for evaluating future air and space forces. *Management Science* 44(10):1336–1350, 1998.

[105] G. S. Parnell, D. W. Hughes, R. C. Burk, P. J. Driscoll, P. D. Kucik, B. L. Morales, and L. R. Nunn. Survey of value-focused thinking: Applications, research developments and areas for future research. *Journal of Multi-Criteria Decision Analysis* 20(1–2):49–60, 2013.

[106] J. W. Payne, J. R. Bettman, and D. A. Schkade. Measuring constructed preferences: Toward a building code. *Journal of Risk and Uncertainty* 19(1–3):243–270, 1999.

[107] L. D. Phillips. A theory of requisite decision models. *Acta Psychologica* 56(1–3):29–48, 1984.

[108] L. D. Phillips. Decision conferencing. W. Edwards, R. F. Miles, and D. von Winterfeldt, eds. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge University Press, New York, 375–399, 2007.

[109] G. F. Pitz, N. J. Sachs, and J. Heerboth. Procedures for eliciting choices in the analysis of individual decisions. *Organizational Behavior and Human Performance* 26 (3):396–408, 1980.

[110] S. Plous. A comparison of strategies for reducing interval overconfidence in group judgments. *Journal of Applied Psychology* 80(4):443–454, 1995.

[111] E. C. Poulton. Biases in quantitative judgements. *Applied Ergonomics* 13(1):31–42, 1982.

[112] E. C. Poulton. *Bias in Quantifying Judgements*. Lawrence Erlbaum Associates, Hove, UK, 1989.

[113] M. Pöyhönen, H. Vrolijk, and R. P. Hämäläinen. Behavioral and procedural consequences of structural variation in value trees. *European Journal of Operational Research* 134(1):216–227, 2001.

[114] H. Raiffa. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Addison-Wesley, Oxford, UK, 1968.

[115] C. Romero and T. Rehman. Natural resource management and the use of multiple criteria decision-making techniques: A review. *European Review of Agricultural Economics* 14(1):61–89, 1987.

[116] L. Ross, D. Greene, and P. House. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology* 13(3):279–301, 1977.

[117] E. A. J. A. Rouwette and J. A. M. Vennix. Team learning on messy problems. V. I. Sessa and M. London, eds. *Work Group Learning: Understanding, Improving and Assessing How Groups Learn in Organizations*. Lawrence Erlbaum Associates, Mahwah, NJ, 245–286, 2007.

[118] B. Roy. Decision science or decision-aid science? *European Journal of Operational Research* 66(2): 184–203, 1993.

[119] J. E. Russo and P. J. H. Schoemaker. *Decision Traps: Ten Barriers to Brilliant Decision-Making and Now to Overcome Them*. Doubleday, New York, 1989.

[120] T. L. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, New York, 1980.

[121] D. A. Seaver, D. von Winterfeldt, and W. Edwards. Eliciting subjective probability distributions on continuous variables. *Organizational Behavior and Human Performance* 21(3):379–391, 1978.

[122] J. Siebert and R. L. Keeney. Creating more and better alternatives for decisions using objectives. *Operations Research* 63(5):1144–1158, 2015.

[123] H. A. Simon. *Models of Bounded Rationality*. MIT Press, Cambridge, MA, 1982.

[124] P. Slovic. The construction of preference. *American Psychologist* 50(5):364–371, 1995.

[125] P. Slovic, M.L. Finucane, E. Peters, and D. G. MacGregor. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24(2):311–322, 2004.

[126] C. Spetzler, H. Winter, and J. Meyer, eds. *Decision Quality Value Creation from Better Business Decisions*. John Wiley & Sons, Hoboken, NJ, 2016.

[127] T. J. Stewart. Robustness of additive value function methods in MCDM. *Journal of Multi-Criteria Decision Analysis* 5(4):301–309, 1996.

[128] A. Tsoukias, G. Montibeller, G. Lucertini, and V. Belton. Policy analytics: An agenda for research and practice. *EURO Journal on Decision Processes* 1(1–2):115–134, 2013.

[129] A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5(2):207–232, 1973.

[130] R. von Nitzsch and M. Weber. The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science* 39(8):937–943, 1993.

[131] D. von Winterfeldt. On the relevance of behavioral decision research for decision analysis. J. Shanteau, B. A. Mellers, and D. A. Schum, eds. *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Kluwer, Norwell, MA, 133–154, 1999.

[132] D. von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research*. Cambridge University Press, New York, 1986.

[133] J. Wallenius, J. S. Dyer, P. C. Fishburn, R.E. Steuer, S. Zionts, and K. Deb. Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science* 54(7):1336–1349, 2008.

[134] T. S. Wallsten, D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General* 115(4):348–365, 1986.

[135] Q. Wang and K. L. Poh. A survey of integrated decision analysis in energy and environmental modeling. *Energy* 77(December):691–702, 2014.

[136] M. Weber. Decision making with incomplete information. *European Journal of Operational Research* 28(1):44–57, 1987.

[137] M. Weber and K. Borcherding. Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research* 67(1):1–12, 1993.

[138] M. Weber, F. Eisenführ, and D. von Winterfeldt. The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science* 34(4):431–445, 1988.

[139] R. L. Winkler. Scoring rules and the evaluation of probabilities. *Test* 5(1):1–26, 1996.