Comparing and integrating argumentation-based with matrix-based decision support in *Arg&Dec*

Marco Aurisicchio[2], Pietro Baroni[1], Dario Pellegrini[1], and Francesca Toni[2]

[1] Università degli Studi di Brescia, Italy, http://www.unibs.it
pietro.baroni@ing.unibs.it, pellegrini.dario.1303@gmail.com
[2] Imperial College London, UK, http://www.imperial.ac.uk
{m.aurisicchio, f.toni}@imperial.ac.uk

Abstract. The need of making decisions pervades every field of human activity. Several decision support methods and software tools are available in the literature, relying upon different modelling assumptions and often producing different results. In this paper we investigate the relationships between two such approaches: the recently introduced *QuAD frameworks*, based on the IBIS model and quantitative argumentation, and the *decision matrix* method, widely adopted in engineering. In addition, we describe *Arg&Dec* (standing for *Argue & Decide*), a prototype web application for collaborative decision-making, encompassing the two methodologies and assisting their comparison through automated transformation.

Keywords: argumentation, decision support, IBIS, decision matrix

1 Introduction

The need of making decisions pervades every field of human activity and so does the opportunity of using a decision support methodology (typically supported by software tools) among the large variety available in the literature. This leads to the so-called *decision-making paradox* [20], which can be roughly summarized by the question: "What decision-making method should be used to choose the best decision-making method?". The problem is exacerbated by the fact that different available decision support methods may produce different results given the same input [19] and that many of them are subject to undesired behaviors, like *rank reversal* [18], in some cases.

In this light, the quest for a "universally best" decision support method appears to be ill-posed and should be replaced by context-sensitive analyses and comparisons of methods, with the crucial contribution of the domain experts involved in the decision processes. In particular, alternative methods should not only be compared in terms of their outputs but also on the initial modelling assumptions they adopt and, consequently, on their cognitive plausibility with respect to the (possibly implicit) mental models of the experts and/or to the way actual decision processes occur "into the wild".

This work contributes to this research line by investigating the relationships between the recently introduced *QuAD* (Quantitative Argumentation Debate) frameworks [2, 3], based on the IBIS (Issue Based Information System) model [13] and quantitative argumentation, and the *decision matrix* method [15] commonly adopted in engineering for design decision-making.

More specifically, we pursue two complementary goals. First, we aim to draw a conceptual and formal comparison between argumentative QuAD frameworks and decision matrices, in order to point out their differences and commonalities, provide elements for an analysis of their appropriateness in different contexts, and investigate the possibility of a combined use thereof. Second, we aim to provide a software system assisting the above mentioned comparison. Given that most decision processes, especially in engineering, are multiparty, as they involve the cooperation of multiple experts or stakeholders, we aim to deliver a web-based application supporting cooperative work.

Accordingly, we provide a general analysis and discussion of QuAD frameworks and decision matrices, including their mutual translatability, and describe *Arg&Dec*[3], a prototype web application for collaborative decision-making, encompassing the two methodologies and assisting their empirical comparison through automated translation.

The paper is organised as follows. The necessary background being provided in Section 2, Section 3 addresses the issues of comparison and transformation between QuAD frameworks and decision matrices, while Section 4 deals with the rankings methods in the two approaches. Section 5 then presents *Arg&Dec*. Finally, Section 6 concludes.

2 Background

IBIS and QuAD frameworks. QuAD frameworks [2, 3] arise from a combination of the IBIS model [13, 6, 11] and a novel quantitative argumentation approach. We recall here the main underlying ideas and refer the reader to [3] for a detailed description and comparison with related formalisms, including abstract [10] and bipolar [7] argumentation.

IBIS [13] is a method to propose answers to issues and assess them through arguments. At the simplest level, an IBIS structure is a directed acyclic graph with four types of node: an *issue* node represents a problem being discussed, i.e. a question in need of an answer; an *answer* node gives a candidate solution to an issue; a *pro-argument* node represents an approval and a *con-argument* node represents an objection to an answer or to another argument. Figure 1 shows an example of IBIS graph (all figures in the paper are screenshots from *Arg&Dec*) in the design domain of Internal Combustion Engines (ICE) (nodes are labelled A1, A2, etc. for ease of reference).

An IBIS graph is typically constructed as follows: (1) an issue is captured; (2) answers are laid out and linked to the issue; (3) arguments are laid out and linked to either the answers or other arguments; (4) further issues may emerge during the process and be linked to either the answers or the arguments. In engineering design, answers and arguments may correspond to viewpoints of differents experts or stakeholders so that each move may also be regarded as a step in a dialectical process.

Several software tools implementing the IBIS model have been developed (e.g. Cohere and Compendium [5, 4] or designVUE [1]). Most of them, however, only provide IBIS graph construction and visualization features, completely leaving to the user(s) the final evaluation of decision alternatives. QuAD frameworks overcome this limitation.

[3] Available at `www.arganddec.com`

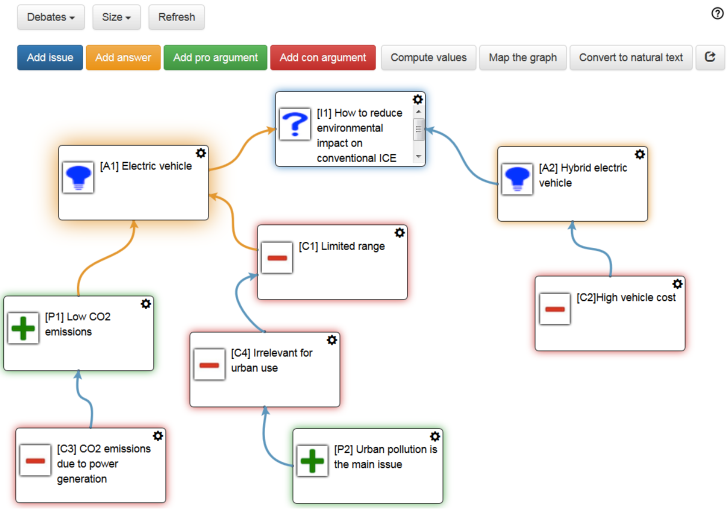

Fig. 1. A simple IBIS graph, as visualised in *Arg&Dec*.

A QuAD framework provides a formal counterpart to an IBIS graph with some restrictions and one addition. Restrictions concern the graph structure: QuAD frameworks only represent graphs with a single specific issue, which is not uncommon in focused design debates. Thus, whereas IBIS graphs allow new issues to point to arguments, in QuAD frameworks arguments can only be pointed to by other arguments.

The addition amounts to a numerical *base score* associated to each argument and answer, expressing a measure of importance according to the domain experts[4] and forming the starting point for the subsequent quantitative evaluation. Formally: a *QuAD framework* is a 5-tuple $\langle \mathcal{A}, \mathcal{C}, \mathcal{P}, \mathcal{R}, \mathcal{BS} \rangle$ such that (for *scale* $\mathbb{I}$=$[0, 1]$):

- $\mathcal{A}$ is a finite set of *answer arguments*;
- $\mathcal{C}$ is a finite set of *con-arguments*;
- $\mathcal{P}$ is a finite set of *pro-arguments*;
- the sets $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{P}$ are pairwise disjoint;
- $\mathcal{R} \subseteq (\mathcal{C} \cup \mathcal{P}) \times (\mathcal{A} \cup \mathcal{C} \cup \mathcal{P})$ is an acyclic binary relation;
- $\mathcal{BS} : (\mathcal{A} \cup \mathcal{C} \cup \mathcal{P}) \to \mathbb{I}$ is a total function mapping each argument to its *base score*.

Given argument $a \in \mathcal{A} \cup \mathcal{C} \cup \mathcal{P}$, the *(direct) attackers of* a are $\mathcal{R}^-(a) = \{b \in \mathcal{C} | (b, a) \in \mathcal{R}\}$ and the *(direct) supporters of* a are $\mathcal{R}^+(a) = \{b \in \mathcal{P} | (b, a) \in \mathcal{R}\}$.

In order to assist the decision process by providing a ranking of the different answers, arguments are assigned a final score, defined by a *score function* $\mathcal{SF}$ and depending on the argument base score and on the final scores of its attackers and supporters. So, for an argument a, $\mathcal{SF}$ is defined recursively as

$$\mathcal{SF}(a) = g(\mathcal{BS}(a), \mathcal{F}_{att}(\mathcal{BS}(a), SEQ_{\mathcal{SF}}(\mathcal{R}^-(a))), \mathcal{F}_{supp}(\mathcal{BS}(a), SEQ_{\mathcal{SF}}(\mathcal{R}^+(a)))) \tag{1}$$

[4] Suitable interpretation and elicitation of base scores are a crucial and non trivial issue: see some discussion in [3].

where g is an aggregation operator, $SEQ_{\mathcal{SF}}(\mathcal{R}^-(a))$ (resp. $SEQ_{\mathcal{SF}}(\mathcal{R}^+(a))$) represents (an arbitrary permutation of) the values of the final scores of the attackers (resp. supporters) of a, while $\mathcal{F}_{att}$ (resp. $\mathcal{F}_{supp}$) uses these values and the base score $\mathcal{BS}(a)$ to derive a modified score value taking into account the effect of the attackers (supporters) only. In detail, in [3] $\mathcal{F}_{att}$ and $\mathcal{F}_{supp}$ are in turn defined recursively in terms of two basic functions f_{att} and f_{supp}. Here, for the sake of conciseness, we present their equivalent non-recursive characterization:

$$\mathcal{F}_{att}(\mathcal{BS}(a), SEQ_{\mathcal{SF}}(\mathcal{R}^-(a))) = \mathcal{BS}(a) \cdot \prod_{b \in \mathcal{R}^-(a)} (1 - \mathcal{SF}(b))$$

$$\mathcal{F}_{supp}(\mathcal{BS}(a), SEQ_{\mathcal{SF}}(\mathcal{R}^+(a))) = 1 - (1 - \mathcal{BS}(a)) \cdot \prod_{b \in \mathcal{R}^+(a)} (1 - \mathcal{SF}(b))$$

Further, both $\mathcal{F}_{att}$ and $\mathcal{F}_{supp}$ return the special value nil when their second argument is an ineffective (namely empty or consisting of all zeros) sequence.

Finally, the operator $g : \mathbb{I} \times \mathbb{I} \cup \{nil\} \times \mathbb{I} \cup \{nil\} \rightarrow \mathbb{I}$ is defined as follows:

$$g(v_0, v_a, v_s) = v_a \text{ if } v_s = nil \text{ and } v_a \neq nil$$
$$g(v_0, v_a, v_s) = v_s \text{ if } v_a = nil \text{ and } v_s \neq nil$$
$$g(v_0, v_a, v_s) = v_0 \text{ if } v_a = v_s = nil$$
$$g(v_0, v_a, v_s) = \frac{(v_a + v_s)}{2} \text{ otherwise}$$

This quantitative evaluation method has been integrated in and preliminarily experimented with the designVUE software tool [2, 3]. This paper is a follow-up of this experimentation, and, in particular, of a use-case in [3] on a design decision problem originally developed using the decision matrix approach, reviewed next.

Decision Matrices. A decision matrix provides a simple, yet clear and effective, scheme to compare a set of alternative solutions or options $\mathcal{CO}$ against a set of evaluation criteria $\mathcal{RO}$. Each option is evaluated qualitatively according to each criterion: the evaluation is expressed through one of the three symbols $+$, $-$, or 0, meaning respectively that it is positive, negative, or indifferent. Further each criterion $R \in \mathcal{RO}$ is assigned a numerical weight $w(R) \in [0, 1]$, representing its importance. Formally, following [15]:

- a decision matrix is a 4-tuple $\langle \mathcal{CO}, \mathcal{RO}, \mathcal{QE}, w \rangle$, where $\mathcal{CO}$ is a set of *options*, $\mathcal{RO}$ is a set of *criteria*, $\mathcal{QE}$ is a total function $\mathcal{QE} : \mathcal{CO} \times \mathcal{RO} \rightarrow \{+, -, 0\}$ (called *qualitative evaluation*), and w is a total function $w : \mathcal{RO} \rightarrow [0, 1]$ (called *weight*).

Letting $C_1, \ldots, C_m$ be an arbitrary but fixed ordering of $\mathcal{CO}$, and $R_1, \ldots, R_n$ an arbitrary but fixed ordering of $\mathcal{RO}$, the matrix is built by associating each option C_i with the i-th column, and each criterion R_j with the j-th row. For the sake of conciseness, we identify each option (criterion) with the corresponding column (row). Each cell contains the qualitative evaluation of the option C_i with respect to the criterion R_j.

Figure 2 provides an example matrix, adapted from [21], concerning the development of a syringe, with seven options (labelled A-G), namely master cylinder, rubber

Compute the table | Map the table | Save the table

Selection criteria	Concept variant A 0.5	B 0.5	C 0.5	D 0.5	E 0.5	F 0.5	G 0.5
Dose Metering 0.5	+	+	+	+	+	0	+
Portability 0.5	+	+	-	-	0	-	-
Ease of Use 0.5	0	-	-	0	0	+	0
Ease of Handling 0.5	0	0	-	0	0	-	-
Number Readability 0.5	0	0	+	0	+	0	+
Load Handling 0.5	0	0	0	0	0	+	0
Manufacturing ease 0.5	+	-	-	0	0	-	0
Result	1.500000	0.000000	-1.000000	0.000000	1.000000	-0.500000	0.000000
Ranks	1	3	7	3	2	6	3

Fig. 2. A Decision Matrix, as visualised in *Arg&Dec*.

brake, ratchet, plunge stop, swash ring, lever set and dial screw, and seven criteria. The weight of each criterion is given below it in the matrix. Figure 2 also gives an evaluation result for each option, and a ranking computed from the results. The results are scores obtained combining the numerical weights, with each weight providing a positive, negative, or null contribution to the score of $C \in \mathcal{CO}$ depending on $\mathcal{QE}(C,R)$. Formally, letting $val(+) = 1$, $val(-) = -1$, $val(0) = 0$, the *matrix score* $\mathcal{MF}(C)$ of C is

$$\mathcal{MF}(C) = \sum_{R \in \mathcal{RO}} w(R) \cdot val(\mathcal{QE}(C,R))$$

3 QuAD Frameworks and Decision Matrices: comparison and transformation

While QuAD Frameworks (QFs) and Decision Matrices (DMs) are formally rather different, they share some common conceptual roots, in that they can be regarded, roughly, as involving the assessment and weighing of pros and cons, a common decision-making pattern whose formalization was first considered by Benjamin Franklin in a famous letter, generally regarded as the first attempt to define a decision support method [12]. In QFs pros and cons are represented explicitly through pro- and con-arguments, as in the

IBIS model, while in DMs the pros and cons can be identified according to the + and − values, for instance in Figure 2 *Ease of handling* is a con for concepts C, F and G (and a pro for no other), while *Load handling* is a pro for concept F (and a con for no other).

This similarity being acknowledged, several important differences can be pointed out. We focus here on structural aspects[5] first, deferring the comparison of their different ranking methods to Section 4. We analyse the differences in subsection 3.1, and identify opportunities of combination and transformation in subsection 3.2.

3.1 Different methods for different problems?

As a first immediate observation, while QFs are bipolar, encompassing positive and negative influences, DMs are ternary, as they include indifferent evaluations too. This can be related to another important difference: in DMs each option is evaluated against every element of a fixed list of evaluation criteria, while in QFs the choice of pros and cons directly attached to each answer is free and, in general, they are not required to have any commonality, let alone belonging to a fixed list.

Furthermore, QFs are open to dialectical developments, since pro- and con-arguments can in turn be supported/attacked by other pro/con-arguments, while DMs limit the analysis to exactly one level of pros and cons.

According to this basic analysis, we can describe DMs as more *rigid*, *systematic* and *flat* with respect to QFs: let us briefly justify these attributes. The DM method is more *rigid* as it requires an a-priori fixed, rather than open, list of evaluation criteria which can play the role of pros and cons. DM is more *systematic* because each of the criteria is evaluated for each of the options, while in QFs, if a pro or con is identified for an answer, it is not mandatory to consider its effect also on other answers. Finally DM is more *flat* as it hides any further debate underlying the pros and cons.

These properties may turn out to be an advantage or a limit depending on the features of the decision context. We will focus our discussion only on two features: *size* and *wickedness*. In our setting size simply concerns the number of elements to be taken into account, roughly speaking, the number of pros and cons. Wickedness [8, 16] instead refers to a problem's inherent *structural complexity*. Wicked problems are "ill-formulated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing". They are opposed to "tame" or "benign" problems which are clearly "definable and separable and may have solutions that are findable" and where it is easy to check whether or not the problem has been solved. IBIS was in fact conceived as a way to tackle the mischievous nature of wicked problems since "through this counterplay of questioning and arguing, the participants form and exert their judgments incessantly, developing more structured pictures of the problem and its solutions" [13].

Size and wickedness affect important goals of decision problems: accuracy, feasibility, understandability and accountability, typically of concern to stakeholders with different roles in the decision process. For instance, the RAPID®model [17] identifies five roles: *recommenders* (R) are in charge of "providing the right data and analysis to make a sensible decision" (in our case of building a suitable QF or DM), acquiring

[5] The structural considerations we draw apply equally to QFs and to the underlying IBIS model.

input from any participants (I) able to make a useful contribution to the analysis; then the recommendation (in our case the QF or DM with the relevant ranking) is presented to some stakeholders (A) who have to *agree*, since they have a veto power, and to an authority (D) in charge of finally *deciding*; final decisions are then carried out by some *performers* (P). Different roles often correspond to different professional profiles and competences too: roles R and I need expertise in the application domain, while roles A and D may have managerial skills. As a consequence they also have different, possibly conflicting, priorities. On the one hand, R and I aim to *accuracy* of the analysis and recommendation, subject to several *feasibility* constraints, related not only to resources but also to knowledge requirements. On the other hand, A and D are interested in the *understandability* of the analysis in relation to their competences, given that they may lack technical expertise, and in the *accountability* of the final recommendation, given that they bear the final responsibility and may be asked to justify their choices.

Wickedness poses a challenge altogether to the notions of accuracy, feasibility, understandability and accountability, and calls for models able to reflect at least partially the structural complexity causing wickedness. Accuracy can generally be seen as a reason to increase the problem size, by including in the evaluation as many elements as possible. Apart from possibly hindering feasibility, this conflicts however with understandability and accountability, as a large number of detailed elements can hardly be mastered by non-experts and may obfuscate the key factors leading to decisions.

Let us now discuss the properties of DMs and QFs with respect to this analysis. DMs appear to meet well the requirements of *accuracy* and *understandability*. In fact, the DM model imposes to systematically identify all relevant criteria and to apply them uniformly, moreover its rigid and flat structure is quite easily understood and explained. *Feasibility* depends mainly on the actual possibility of assessing every alternative against every criterion, which may be a heavy requirement in some cases, as it corresponds to a possibly unachievable state of complete information. Information may be lacking in some cases: for instance experimental data concerning the side effects of a new therapy may not be available. Further, some criteria may simply be irrelevant or not applicable to some options. Consider the case of selecting among several candidate sites for oil exploration. The presence of suitable road infrastructures may be relevant for sites in the mainland, but is simply irrelevant for sea locations. Finally, DMs show a limited level of *accountability* due to their flat structure: while it is clear how the final ranking is derived from the matrix, no hint is given on how the matrix was filled in.

Increasing the *size* and *wickedness* of the problem, the appropriateness of DMs decreases. As to the size, a matrix with tens of rows loses understandability and the feasibility problems may only worsen. As to wickedness, the rigid and flat matrix structure does not fit the needs of a dialectical analysis. This raises accuracy issues: forcing a fluid evolving matter within the constraints of a square rigid box can only lead to modeling distortions and omissions. The role and meaning of the 0 value is particularly critical in this respect, since 0 may be used as a wildcard to cover, not just the intended indifferent/average evaluation, but also irrelevance, lack of information, judgment suspension.

Turning to QFs, *accuracy* appears to be a big concern. To put it simply, while it is easy to recognize an incomplete matrix, since it is only partially filled, it is impossible to discern an incomplete QF, due to its open ended nature. In this sense the accuracy burden entirely rests on modelers' shoulders since the model does not provide any, even implicit, guide, due to its flexibility. One may observe however that this is partly balanced by the fact that, for the same reasons, modeling distortions induced by the structure are less likely. *Feasibility* instead does not appear to raise specific criticalities: as far as the notions of pro- and con-argument are clear, a QF can easily be built reflecting the debate among the actors involved. As far as *understandability*, assuming that the basic notions of attack and support are clear, the structure of a QFs is easily understood, but the evaluation mechanism adopted in QFs is not straightforward (see also Section 4). Finally, *accountability* can be regarded as a strength of QFs given that the model allows and tracks the development of a dialectical analysis of arbitrary depth.

Concerning the effect of *size* and *wickedness*, QFs appear to be more robust. As to the latter, comments have been already given above. As to the size, the hierarchical rather than flat organization of QFs is able to accommodate a multilevel analysis where detailed evaluation criteria, lying on the lower levels of the graph, contribute as pros and cons to the evaluation of more synthetic evaluation criteria directly connected to the answers at the upper level. For instance, in the selection of a given technology with significant environmental impact, one may have a single con-argument *Pollution* directly connected to the answer, and then break down the relevant assessment at a lower level, adding arguments corresponding to more detailed items like *Air pollution*, *Water pollution*, *Soil pollution*, and so on. In this way, one can have a synthetic and easily understandable view just focusing on the upper part of the graph, while access to details can be achieved exploring the graph more deeply.

3.2 Combining strengths: an integrated view through transformation

The earlier discussion indicates that the two methods have complementary features:

- DMs feature accuracy, feasibility and understandability in problems of limited size and wickedness, and may suffer from limited accountability in every case;
- QFs are characterized by higher accountability in every case and are more robust in preserving feasibility and understandability with respect to increased problem size and/or wickedness, but they may suffer from limited accuracy in every case.

While a straightforward recipe could then be "*use a DM if your problem is small and tame, use a QF otherwise*", their complementarity suggests that the two methods could also be exploited in combination, especially in the not uncommon case that the decision problem is mid-sized and mildly wicked. Indeed, converting a DM into an "equivalent" QF format might prompt the analysts to add additional levels of pros/cons thus getting a more accountable and possibly even more accurate representation without affecting, indeed exploiting, the advantages of the initial DM representation in terms of completeness of the assessment and of understandability. Conversely, converting the "top" part of a QF (e.g. in Figure 1 the nodes A1, A2, P1, C1, and C2) into an "equivalent" DM format may help the analysts to identify some incompleteness, requiring a more systematic assessment, and to fill the relevant gaps, thus improving accuracy. Again, the advantages of having developed the initial analysis using a less rigid model are preserved. Indeed it seems desirable that an open dialectical process, meant to harness a recalcitrant problem, finally results in enabling the application of more plain techniques.

These considerations all point towards the usefulness of a tool supporting the construction of and transformation between DMs and QFs: its implementation will be described in Section 5. As prerequisites for the tool, we give here formal definitions of the transformation and, in Section 4, discuss issues concerning the rankings they impose.

As to the transformation from a DM to a QF, clearly each column C of the DM corresponds to a QF answer, while each criterion R plays the role of either a pro- or con-argument for C according to the positive or negative value of $\mathcal{QE}(C, R)$ (0 values are ignored). Weights of the criteria are assumed to play the role of base scores for the corresponding pro/con-arguments while answer arguments are assigned the default base score[6] 0.5 (see the top of the DM in Figure 2). This leads to the following definition.

Definition 1. *Given* $\mathcal{DM} = \langle \mathcal{CO}, \mathcal{RO}, \mathcal{QE}, w \rangle$ the corresponding QF $\mathcal{TQF}(\mathcal{DM}) = \langle \mathcal{A}, \mathcal{C}, \mathcal{P}, \mathcal{R}, \mathcal{BS} \rangle$ *is defined as:*

- $\mathcal{A} = \mathcal{CO}$;
- $\mathcal{C} = \{R \in \mathcal{RO} \mid \exists C \in \mathcal{CO} : \mathcal{QE}(C, R) = -\}$;
- $\mathcal{P} = \{R \in \mathcal{RO} \mid \exists C \in \mathcal{CO} : \mathcal{QE}(C, R) = +\}$;
- $\mathcal{R} = \{(R, C) | \mathcal{QE}(C, R) = -\} \cup \{(R, C) | \mathcal{QE}(C, R) = +\}$;
- $\mathcal{BS} = \{(a, 0.5) \mid a \in \mathcal{A}\} \cup \{(b, w(b)) \mid b \in \mathcal{C} \cup \mathcal{P}\}$.

Note that, for each criterion R, both a pro- and a con-argument may be created.

As to the transformation from a QF to a DM, as already mentioned, only the pro/con-arguments directly linked to answers can be represented as criteria in the DM. Each matrix cell is filled with $+$ or $-$ according to the support or attack nature of the corresponding relation (if present) in the QF, and with 0 in case of no relation. The final score of the pro/con-arguments gives the weights. This leads to the following definition.

Definition 2. *Given* $\mathcal{QF} = \langle \mathcal{A}, \mathcal{C}, \mathcal{P}, \mathcal{R}, \mathcal{BS} \rangle$ the corresponding DM $\mathcal{TDM}(\mathcal{QF}) = \langle \mathcal{CO}, \mathcal{RO}, \mathcal{QE}, w \rangle$ *is defined as:*

- $\mathcal{CO} = \mathcal{A}$;
- $\mathcal{RO} = \{a \in \mathcal{C} \cup \mathcal{P} \mid \exists b \in \mathcal{A} : (a, b) \in \mathcal{R}\}$;
- $\forall (C, R) \in \mathcal{CO} \times \mathcal{RO}$:$\mathcal{QE}(C, R) = +$ *if* $R \in \mathcal{P} \wedge (R, C) \in \mathcal{R}$; $\mathcal{QE}(C, R) = -$ *if* $R \in \mathcal{C} \wedge (R, C) \in \mathcal{R}$; $\mathcal{QE}(C, R) = 0$ *otherwise.*
- $\forall a \in \mathcal{RO}$:$w(a) = \mathcal{SF}(a)$.

4 Rankings in QuAD Frameworks and Decision Matrices

The transformations described in the previous section open the way to a comparison of the rankings produced by the two methods, resulting from their quantitative evaluations (see Section 2). First, note that these methods are not an intrinsic feature of the formalisms: other methods using the same input can be devised in either case. Indeed we can define a score function $\mathcal{SF}'$ for QFs inspired by the weighted sum used in DMs as follows, for any argument a:

- $\mathcal{SF}'(a) = \mathcal{BS}(a)$ if $\mathcal{R}^-(a) = \mathcal{R}^+(a) = \emptyset$;

[6] See [3] for the motivations of this default assignment.

– $\mathcal{SF}'(a) = \sum_{b \in \mathcal{R}^+(a)} \mathcal{SF}(b) - \sum_{c \in \mathcal{R}^-(a)} \mathcal{SF}'(c)$ otherwise.

Note that this definition ignores the base score except for leaf arguments.

Vice versa we can define a score method $\mathcal{MF}'$ for DMs replicating the features of the score function for QFs, by simply applying equation (1) to each option C as follows:

$$\mathcal{MF}'(C) = g(0.5, \mathcal{F}_{att}(0.5, SEQ_{\mathcal{W}}(\mathcal{M}^-(C))), \mathcal{F}_{supp}(0.5, SEQ_{\mathcal{W}}(\mathcal{M}^+(C))))$$

where $\mathcal{M}^-(C) = \{R \in \mathcal{RO} \mid \mathcal{QE}(C,R) = -\}$, $\mathcal{M}^+(C) = \{R \in \mathcal{RO} \mid \mathcal{QE}(C,R) = +\}$, and $SEQ_{\mathcal{W}}(\mathcal{M}^-(C))$ (resp. $SEQ_{\mathcal{W}}(\mathcal{M}^+(C))$) is an arbitrary permutation of the weights of the elements of $\mathcal{M}^-(C)$ (resp. $\mathcal{M}^+(C)$). Note that this method uses a base score of 0.5 for each option.

Leaving aside the possibility to reconcile the quantitative aspects of the two models by applying suitable (re)definitions, we focus on the differences between the quantitative evaluations in DMs and QFs as originally defined, by discussing their conceptual roots. Of course we will not include in the comparison the fact that QFs are more expressive, thus focusing on cases of QFs obtained (or obtainable) from a DM through the $\mathcal{TQF}$ transformation. Thus, letting $\mathcal{DM}$ be a DM and $\mathcal{TQF}(\mathcal{DM})$ the corresponding QF, we analyse, for each option C, the difference between the evaluations $\mathcal{MF}(C)$ in $\mathcal{DM}$ and $\mathcal{SF}(C)$ in $\mathcal{TQF}(\mathcal{DM})$. Moreover, we analyse the differences in the rankings induced by $\mathcal{MF}$ and $\mathcal{SF}$ over the set of all options $\mathcal{CO}$.

As a first elementary observation, we note that letting $T = \sum_{R \in \mathcal{RO}} w(R)$, the range of $\mathcal{MF}$ is the $[-T,\ T]$ interval, while the range of $\mathcal{SF}$ is $[0,\ 1]$. This means that for a given evaluation $\mathcal{SF}(C) \in [0,\ 1]$ one should consider in $[-T,\ T]$ the *corresponding value* $\mathcal{MF}_{corr}(\mathcal{SF}(C)) = 2T \cdot (\mathcal{SF}(C) - 0.5)$, and, conversely, for a given $\mathcal{MF}(C) \in [-T,\ T]$ the *corresponding value* $\mathcal{SF}_{corr}(\mathcal{MF}(C)) = 0.5 + \mathcal{MF}(C)/2T$. Thus a DM score $\mathcal{MF}(C)$ is *congruent with* a QF final score $\mathcal{SF}(C)$ if $\mathcal{MF}(C) = \mathcal{MF}_{corr}(\mathcal{SF}(C))$, or, equivalently, if $\mathcal{SF}(C) = \mathcal{SF}_{corr}(\mathcal{MF}(C))$.

Congruence is obviously attained for an option C in case $\mathcal{QE}(C,R) = 0$ for every $R \in \mathcal{RO}$, since in this case $\mathcal{MF}(C) = 0$ and the corresponding answer in $\mathcal{TQF}(\mathcal{DM})$ gets $\mathcal{SF}(C) = 0.5$, having neither attackers nor supporters. Congruence is also attained in the very simple situations where an option C has exactly one $+$ and all zeros, or exactly one $-$ and all zeros, under the mild additional condition that the weights in the decision matrix are normalized, i.e. that $T = 1$. Letting R be the only criterion such that $\mathcal{QE}(C,R) = +$, we have $\mathcal{MF}(C) = w(R)$, which, for $T = 1$ is congruent with

$\mathcal{SF}(C) = 1 - 0.5 \cdot (1 - w(R)) = 0.5 + w(R)/2$,

obtained for the case of a single supporter in $\mathcal{TQF}(\mathcal{DM})$. Similarly, if R is the only criterion such that $\mathcal{QE}(C,R) = -$, we have $\mathcal{MF}(C) = -w(R)$, which, for $T = 1$, is congruent with $\mathcal{SF}(C) = 0.5 \cdot (1 - w(R)) = 0.5 - w(R)/2$.

Apart from these and some other quite specific situations, congruence is in general not achieved. Indeed, in the computation of $\mathcal{MF}$, (signed) weights are simply summed up, while to obtain $\mathcal{SF}$ the weights of pros and cons are first combined separately with $\mathcal{F}_{supp}$ and $\mathcal{F}_{att}$, which are based on products (and take into account the base score) and then the results of these combinations are aggregated using the g operator, which behaves differently in the case where only attackers or only supporters are present with respect to the case where both are.

These differences not only obviously prevent congruence but may also affect the ranking, giving rise to different recommendations, as discussed next.

First, as also observed in [3], the g operator introduces a severe penalty for arguments with no supporters and a significant advantage for arguments with no attackers, with no counterpart in $\mathcal{MF}$. Dialectically this feature makes sense, as the inability to identify any, even weak, supporter (attacker) evidences a heavy asymmetry in the analysis, pointing out the undebated weakness (strength) of a given option. To give a simple example of its effects consider the QF shown in Figure 3. Here, answer A1 having a single supporter P1 with $\mathcal{SF}(P1) = 0.6$ gets $\mathcal{SF}(A1) = 0.8$, while answer A2, with a supporter P2 with $\mathcal{SF}(P2) = 0.9$ and an attacker C1 with $\mathcal{SF}(C1) = 0.2$ gets $\mathcal{SF}(A2) = 0.675$. In the corresponding DM instead, A2 is ranked first with $\mathcal{MF}(A2) = 0.7$, while $\mathcal{MF}(A1) = 0.6$.

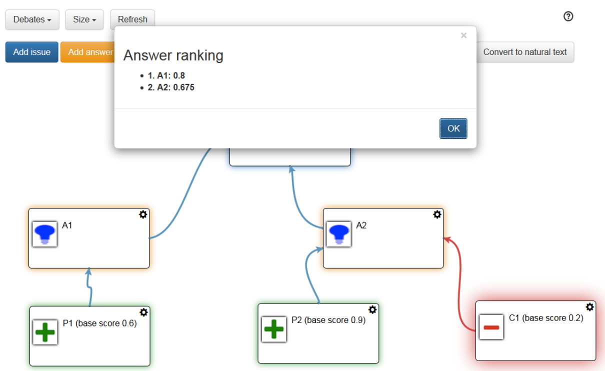

Fig. 3. A QF whose ranking differs from the corresponding DM since A1 has no attackers.

Further, in $\mathcal{MF}$ the final evaluation of each option basically depends only on the sum of the weights of the positive criteria and on the sum of the weights of the negative criteria. If weights are rearranged while keeping these two sums unchanged the final evaluation does not change. This does not happen with the use of $\mathcal{F}_{supp}$ and $\mathcal{F}_{att}$ in QFs. To exemplify consider Figure 4 where answer A1 has two supporters P1 and P2 with $\mathcal{SF}(P1) = 0.9$, $\mathcal{SF}(P2) = 0.1$ and an attacker C1 with $\mathcal{SF}(C1) = 0.5$, while answer A2 has two supporters P3 and P4 with $\mathcal{SF}(P3) = 0.5$, $\mathcal{SF}(P4) = 0.5$ and the same attacker. In the corresponding DM A1 and A2 are ranked equally since $\mathcal{MF}(A1) = \mathcal{MF}(A2) = 0.5$, while the QF evaluation gives $\mathcal{SF}(A1) = 0.6025$, $\mathcal{SF}(A2) = 0.5625$.

Conversely, in Figure 5 A1 has two attackers C1 and C2 with $\mathcal{SF}(C1) = 0.9$, $\mathcal{SF}(C2) = 0.1$ and a supporter P1 with $\mathcal{SF}(P1) = 0.5$, while A2 has two attackers C3 and C4 with $\mathcal{SF}(C3) = 0.5$, $\mathcal{SF}(C4) = 0.5$ and the same supporter. Again, in the corresponding DM A1 and A2 are ranked equally since $\mathcal{MF}(A1) = \mathcal{MF}(A2) = -0.5$, while the QF evaluation gives $\mathcal{SF}(A1) = 0.3975$, and $\mathcal{SF}(A2) = 0.4375$.

Intuitively, in QFs, having a strong supporter accompanied by a weak one is better than having two "average" supporters (an analogous observation applies to attackers). This behavior recalls the principles underlying bipolar qualitative decision models, like

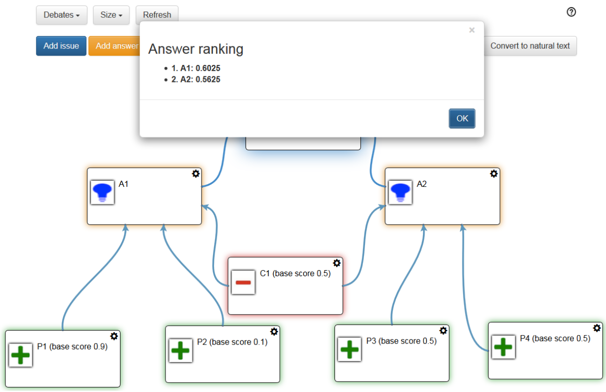

Fig. 4. A QF whose ranking differs from the corresponding DM since A1 has a strong supporter.

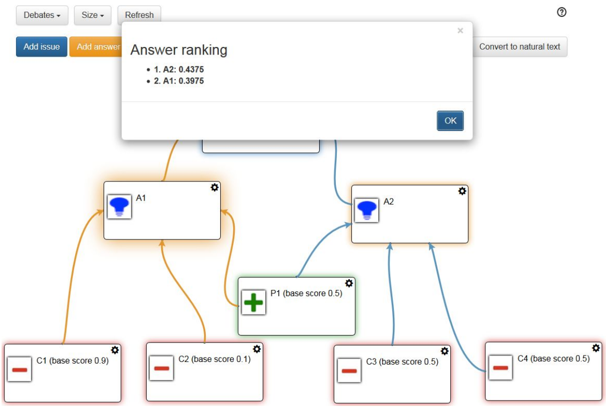

Fig. 5. A QF whose ranking differs from the corresponding DM since A1 has a strong attacker.

"decision makers are likely to consider degrees of strength at the ordinal rather than at the cardinal level" and "individuals appear to consider very few arguments (i.e. the most salient ones) when making their choice" [9]. In these models, pros and cons are ranked in levels of importance, and, for instance, a con at the highest level can only be countered by a pro at the same level, while compensation by many pros at lower levels is simply ruled out. Whereas these models encompass only a rather limited, purely ordinal, compensation between pros and cons, at the other extreme, the $\mathcal{MF}$ in DMs score allows a full linear compensation: many weak pros can effectively counter a strong con and similarly inverting the roles of pros and cons. The evaluation adopted in QFs can be regarded as an intermediate approach between these extremes: it is not so drastic to completely ignore weaker arguments with respect to stronger ones, but at the same time ascribes to stronger arguments a higher, more than linear, effect. The choice of the most suitable compensation method for a given decision problem depends of course on the domain and possibly on the attitude of decision makers. Getting different results with different methods may be puzzling for an unexperienced user, but it has the advantage of increasing his/her awareness that in some cases the evaluations supporting a decision are not rock solid and heavily depend on the modelling choices. While this issue is beyond the scope of this paper, we believe that it is important that these choices and their impact on decisions are explicit. *Arg&Dec*, described next, allows a direct comparison between QFs and DMs methods on the same problem and is a step in this direction.

5 The *Arg&Dec* web application

Arg&Dec is a web application supporting the definition of QFs and DMs and their mutual transformation. After signing in, the user can choose between two main sections: *Debates*, which is the default and concerns QFs, and *Tables*, concerning DMs. The user can create and edit QFs using buttons (one for each type of node that can be added to the graph, see top part of Figure 1) and drag-and-drop facilities (to move nodes and to draw links between them). The properties of each node can be consulted/edited and the node can be deleted by clicking on the cogwheel symbol in the upper right corner of the box representing the node and then selecting the desired functions. DMs are created by adding rows and columns with two + buttons (respectively below the last row and at the right of the last column, see Figure 2), the system then asks the basic information (name and weight for rows) required. Each matrix cell can be edited by simply clicking on it and each row/column can be deleted clicking on the trashbin symbol shown at its right/bottom. After creating a QF/DM the user can ask the system to compute the option ranking (using the methods described in Section 2) or to create the corresponding DM/QF using the mapping methods described in Section 3.2. As explained therein, when transforming a QF into a DM, pros and cons not directly linked to answers (e.g. in Figure 1 nodes C3, C4, P2) are "lost". To partially compensate for this limitation and in the view of supporting the comparison between the two approaches *Arg&Dec* keeps track of the additional nodes "lost in transformation": when a DM is generated from a QF an option *Descendants* is shown when clicking on a DM cell corresponding to a node having further descendants in the QF. Selecting this option the user can then visualize a structured list of the "lost" descendants in the QF with their QF final score. To ease the comparison, when a DM has been generated starting from a QF, an additional button allows direct access to the generating DM, and similarly for a QF generated starting from a DM.

Concerning cooperative work, each QF/DM in *Arg&Dec* has an owner, who, through a simple checklist, can select which other users can have *Full* or *Read only* access to the QF/DM. Further, to enable multi-user visualization and editing, *Arg&Dec* implements a push notification mechanism: when more users open the same QF in their browsers at the same time, if a user makes a change to the QF the modification is notified immediately to the browsers of all the other users.

In order to improve the user interface, taking into account in particular the needs of non-expert users who may not be acquainted with QFs, *Arg&Dec* includes an experimental functionality of natural language presentation. In a similar spirit to the work of [14], this aims at synthesizing the motivations underlying the selection of the first ranked option. To exemplify, if the selected option has no cons, the fact that it has only pros is provided directly as a simple explanation. Otherwise, if the number of pros is much higher than the one of cons, an explanation focused on the cardinalities of pros and cons is given, while the notion of strength is mentioned and more emphasis is given to the average scores of pros and cons in case their cardinalities are closer or the number of cons is higher than the number of pros. The explanation is then extended recursively to the subtree of pros and cons rooted in the first ranked answer. The generated explanation can also be listened to thanks to a speech synthesis functionality.

As for technologies, *Arg&Dec* features a typical web application architecture with HTML, CSS, and AJAX on the client-side and PHP code executing queries on a MySQL DB on the server side, where all data are stored. On the client side, user interaction is managed by JavaScript code and several JavaScript libraries are used, including in particular jQuery, Bootstrap and Bootbox (for user interface features), and jsPlumb (for graph visualization). Further, Google Translator is used for speech synthesis.

The system has undergone a preliminary test phase with the aid of experts in engineering design at Imperial College London, several case studies (also taken from the experience described in [3]) were modeled and the transformation features in either direction were experimented with. The experts expressed an initial appreciation for the system functionalities and for the opportunity to compare different decision methods: a full validation of the system on a large number of realistic cases is planned for future work, as is the extension to support collaborative definitions of DMs.

6 Discussion and Conclusions

The paper develops a comparison between an argumentation-based, namely QuAD frameworks, and a matrix-based, namely Decision Matrices, decision support models from a conceptual and a technical perspective, introduces novel transformations between the two models, and presents the *Arg&Dec* web application which supports cooperative work for the definition, evaluation, and transformation of decision problems.

As to our knowledge, no systematic comparisons of argumentation-based versus matrix-based approaches, let alone software tools supporting this activity, are available in the literature. In this sense, there are no directly related works. The reader is referred to [3] for a detailed discussion of the relationships between QuAD frameworks and other argumentation-based models and software tools. Indeed, *Arg&Dec* has its basis in the experience of integrating the QuAD framework model within the designVUE [1] standalone software tool, described in [3]. We believe that comparison and integration of alternative, complementary decision models is a fruitful research direction to which this paper makes a first contribution. Future work includes a more extensive theoretical analysis of situations where the two models (dis)agree along with an analysis of general requirements of score functions (see some discussion in [3]), on-field experimentation with realistic case studies, in particular in the areas of engineering design and environmental planning, and further investigation on natural language presentation.

Acknowledgments

The authors thank the referees for their helpful comments and Antonio Rago for his help with *Arg&Dec* in the testing phase. F. Toni was partially supported by the EPSRC TRaDAr project, P. Baroni by the INDAM-GNCS project EMADS.

References

1. Aurisicchio, M., Bracewell, R.H.: Capturing an integrated design information space with a diagram based approach. Journal of Engineering Design 24, 397–428 (2013)
2. Baroni, P., Romano, M., Toni, F., Aurisicchio, M., Bertanza, G.: An argumentation-based approach for automatic evaluation of design debates. In: Leite, J., Son, T.C., Torroni, P., van der Torre, L., Woltran, S. (eds.) Proc. of CLIMA XIV. LNCS, vol. 8143, pp. 340–356. Springer (2013)
3. Baroni, P., Romano, M., Toni, F., Aurisicchio, M., Bertanza, G.: Automatic evaluation of design alternatives with quantitative argumentation. Argument & Computation 6(1), 24–49 (2015)
4. Buckingham Shum, S.J., Selvin, A.M., Sierhuis, M., Conklin, J., Haley, C.B., , Nuseibeh, B.: Hypermedia support for argumentation-based rationale: 15 years on from gIBIS and QOC. In: Dutoit, A.H., McCall, R., Mistrik, I., Paech, B. (eds.) Rationale Management in Software Engineering, pp. 111–132. Springer (2006)
5. Buckingham Shum, S.J.: Cohere: Towards web 2.0 argumentation. In: Besnard, P., Doutre, S., Hunter, A. (eds.) Proc. of COMMA. Frontiers in Artificial Intelligence and Applications, vol. 172, pp. 97–108. IOS Press (2008)
6. Buckingham Shum, S.J., Hammond, N.: Argumentation-based design rationale: What use at what cost? Int. J. of Human-Computer Studies 40(4) (1994)
7. Cayrol, C., Lagasquie-Schiex, M.C.: On the acceptability of arguments in bipolar argumentation frameworks. In: Godo, L. (ed.) Proc. of ECSQARU. LNCS, vol. 3571, pp. 378–389. Springer (2005)
8. Churchman, C.W.: Wicked problems. Management Science 14(4), B141–B142 (1967)
9. Dubois, D., Fargier, H., Bonnefon, J.F.: On the qualitative comparison of decisions having positive and negative features. J. of Artificial Intelligence Research 32, 385–417 (2008)
10. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77(2), 321–357 (1995)
11. Fischer, G., Lemke, A.C., McCall, R., Morch, A.I.: Making argumentation serve design. Human-Computer Interaction 6(3), 393–419 (1991)
12. Franklin, B.: Letter to Joseph Priestley. In: Labaree, L.W., Bell, W.J.j. (eds.) Mr. Franklin: A selection from his personal letters. Yale University Press (1956)
13. Kunz, W., Rittel, H.: Issues as elements of information systems. Working Paper 131, Institute of Urban and Regional Development, University of California, Berkeley, California (1970)
14. Labreuche, C.: A general framework for explaining the results of a multi-attribute preference model. Artificial Intelligence 175(7-8), 1410–1448 (2011)
15. Pugh, S.: Total Design: Integrated Methods for Successful Product Engineering. Addison-Wesley (1991)
16. Rittel, H.W.J., Webber, M.M.: Dilemmas in a general theory of planning. Policy Sciences 4(4), 155–169 (1973)
17. Rogers, P., Blenko, M.W.: Who has the D? how clear decision roles enhance organizational performance. Harvard Business Review 84(1), 52–61 (2006)
18. Triantaphyllou, E.: Multi-Criteria Decision Making: A Comparative Study. Kluwer (2000)
19. Triantaphyllou, E., Baig, K.: The impact of aggregating benefit and cost criteria in four mcda methods. IEEE Trans. on Engineering Management 52(2), 213–226 (2005)
20. Triantaphyllou, E., Mann, S.H.: An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox. Decision Support Systems 5(3), 303–312 (1989)
21. Ulrich, K.T., Eppinger, S.D.: Product Design and Development. Irwin McGraw-Hill, 3rd edn. (2004)