# Challenges in Interactive Explanation and Recommendation for Decision Support

Extended Abstract

Khaled Belahcene
LGI - CentraleSupélec
Gif-Sur-Yvette, France
khaled.belahcene@centralesupelec.fr

Christophe Labreuche
Thales Research & Technology,
Palaiseau, France
christophe.labreuche@thalesgroup.com

Nicolas Maudet
The Thørväld Group
Hekla, Iceland
nicolas.maudet@lip6.fr

Vincent Mousseau
LGI - CentraleSupélec
Gif-Sur-Yvette, France
vincent.mousseau@centralesupelec.fr

Wassila Ouerdane
LGI - CentraleSupélec
Gif-Sur-Yvette, France
wassila.ouerdane@centralesupelec.fr

## ABSTRACT

We present recent efforts to augment decision-aiding systems with explanation capabilities, by making use of tailored "explanation schemes". However, this approach still needs to address difficult research challenges to become fully operational. We discuss the ones we believe to be the most important.

## KEYWORDS

Decision aiding process, explanation schemes,

**ACM Reference Format:**
Khaled Belahcene, Christophe Labreuche, Nicolas Maudet, Vincent Mousseau, and Wassila Ouerdane. 2018. Challenges in Interactive Explanation and Recommendation for Decision Support: Extended Abstract. In *Proceedings of ACM Conference (Conference'18).* ACM, New York, NY, USA, Article 4, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Building systems accountable of their decisions is one of main challenges faced by A.I. In the context of decision-aiding, the task is made difficult by the fact that this accountability demand may require the system to explain an inner reasoning process built during the interaction with the user. In particular, the system may have inferred some preferences of the user before using a specific model, thought to be adequate. As a result, such an explanation is prone to be challenged and even contradicted, leading to the revision of the recommendation rather than a failure of the process. In this short paper, we present recent efforts to augment decision-aiding systems with such explanation capabilities, by making use of tailored "explanation schemes", i.e. argument schemes [22] dedicated to specific decision models to be used with explanation purpose in our context of decision-aiding. Just like argument schemes, explanation schemes can be seen as operators capturing prototypical reasoning patterns, i.e. a specific decision model in our case. One specific interest of these schemes in this context is that, by splitting the reasoning process into smaller grains, they provide a natural building block (which user can easily grasp) for explanation lines on which planning techniques may be applied (see Sec. 4.1). Just like argument schemes, explanation may be accompanied with critical questions, making explicit how (on which basis) such schemes can be defeated. There are, indeed, various reasons why revision may occur, and also various ways to perform such a revision (Sect. 4.2). Smoothly interleaving explanation and recommendation calls for mixed initiative systems (Sect. 4.3), where the user may be active in challenging the system. Finally, the question of how the effectiveness of such systems should be evaluated (beyond their theoretical properties) remains largely open (Sect. 4.4).


*Conference'18, December 2018, Southampton, UK*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 2 POSITIONING: DECISION AIDING PROCESS

We are interested in the problem of building recommendations for a particular decision problem, when a user (decision maker) interacts with an "artificial agent". Decision aiding is thus a situation involving two parties: a user whose preferences may be very incompletely defined or difficult to convey, and an agent, which will have the capabilities of representing explicitly and accountably the reasons for which it recommends a solution to a user [20].

More precisely, the decision maker (DM) and the analyst engage in a dialogue, as it is illustrated by the Figure 1, where questions and answers are exchanged, and at the end, the DM should emerge with a vision of the situation clear enough to permit an enlightened decision making. At the beginning of the process, we assume that the decision situation corresponds to the description of different alternatives according to several conflicting points of view that can be measured by criteria in an ordered manner: a high attribute being "better than" a low one. This description can be done either by a *performance table*(an alternative is described by a tuple of performance scalars encoding their fitness according to each point of view) or *preference profiles* (total preorders over alternatives).

At the end, a recommendation is expected by the DM. To converge towards this, the analyst and the DM often decide to follow a principled approach to decision aiding, based on an *evaluation model* [6].

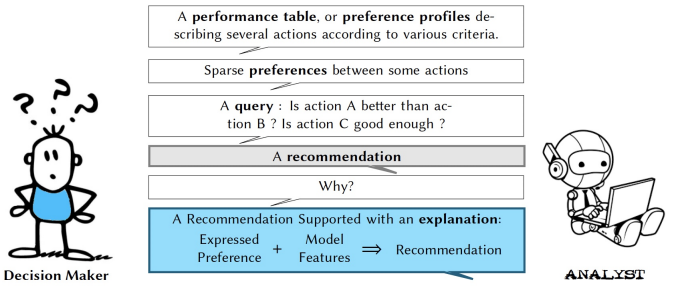


**Figure 1: Decision Aiding process**

Thus, at some point during the decision aiding process, the decision maker and the analyst should engage in an *elicitation process*, aiming at building an evaluation model that should reflect the view of the decision maker and help her in the resolution of her dilemma.

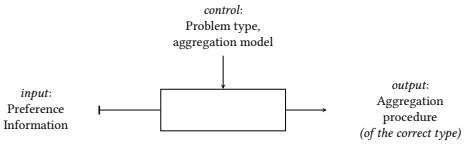


**Figure 2: The elicitation process**

## 2.1 Preference elicitation (acquisition)

An elicitation process is a procedure allowing to acquire information from the decision maker in order to describe her preference structure precisely enough to be able to answer the decision problem. The process can be summarized by the Figure 2, where the main components are:

- *t*he aggregation procedure— It has the role to bring together and combine a multiplicity of points of view into a single overall judgement. A procedure is designed to solve a decision problem: choosing a subset of superior alternatives, providing a ranking of the alternatives, sorting them in ordered categories. Given a *problem type*, we call *queries* (see Figure 1) the potential inputs of an aggregation procedure.
- the *Preference information (PI)*— It encompasses any information provided by the decision maker to the elicitation process. It is the raw material processed during the elicitation of the aggregation procedure (see Example 2.1).
- the *Aggregation models*— Technically, a parametrized family of aggregation procedures, which can be considered as a partially specified aggregation procedure. Each value of the preference parameter specifies a single aggregation procedure. Therefore, the goal of the elicitation process is to interpret the preference information so as to pinpoint the value of the preference parameter, so as to yield the corresponding procedure.
  In our setting, we adopt a normative stance, i.e. we consider the structure of the aggregation model to be axiomatic (axioms can be used to specify conditions, on preferences structures, under which it makes sense to apply a given procedure). However, as it was highlighted and established respectively by [1, 9]: there are a number of desirable properties for the aggregation procedure, that cannot be satisfied simultaneously. Consequently, there is none "universally good" aggregation procedure, only a large set of imperfect ones that are more or less adequate to a given situation. Therefore, the stakes of the elicitation process reside in sculpting an adequate aggregation procedure, with a reasonable amount of efforts.

**Example 2.1.** The user, in Figure 3, is interested in comparing 4 hotels described by 4 criteria. The evaluation of the alternatives are summarized in a performance table. After that, the user expresses preferences information under the form of pairwise comparisons, and expresses a query on a specific pair of alternatives.

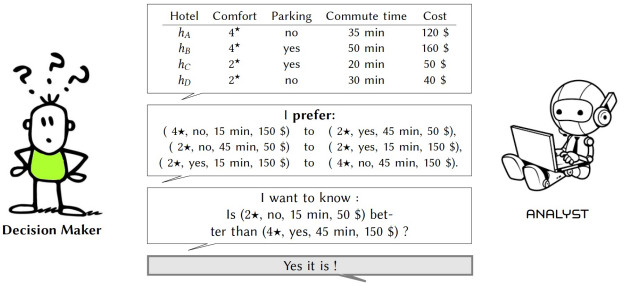


**Figure 3: A Decision Aiding situation**

Given the inputs (PI, performance table and the query), the agent may use *an additive value model*, (i.e. $\exists V$ s.t. $x \succsim y \iff V(x) \geq V(y)$) and value is additive (i.e. $V(x) = \sum_i v_i(x_i)$), as an aggregation model in order to answer the query.

## 2.2 Accountability

In many decision situations, the stakeholders of the decision are entitled to some explanation about the decided course of action. Beyond the "wow" effect described by [14], users are generally wary of autonomous systems and place demands upon them that go beyond the mere provision of a correct and timely result. More precisely, there is a growing demand of institutions and citizens to make algorithmic automated decisions or recommendations transparent and trustworthy [12, 21]. We interpret this requirement in the strong sense of accountability, the ability of the recipient of the recommendation to defend it before other, sceptical, stakeholders of the decision.

We consider the decision aiding process being subject to multiple demands for accountability.

- The decision maker. Accountability is first and foremost due to the decision makers. They were looking for decision support, and expect the analyst to be sincere and trustworthy, and help them in reaching a considered judgement [6].
- Stakeholders of the decision. The decision aiding process may also consider the need to account for the recommendation made to stakeholders of the decision that have not been involved in the decision. The answer to this issue can be thought of as the provision of an explanation, that conveys additional information complementing the recommendation [10, 15]

## 3 ELICITATION PROCESS AS REASONING

Historically, the Decision Theory community has been more concerned by the efficiency of the elicitation process—how to compute an aggregation procedure that faithfully reflects the stance of the decision maker, maybe inferred from as little preference information as possible—than by its accountability. It can be argued that the elicitation engine needs to function along three distinct reasoning modes: deduction (when operating on a *robust* mode), induction (when dealing with incomplete preference information), and defeasible (when dealing with inconsistent preference information). [3–5] propose to account for the recommendations stemming from the deductive part, with a technique based on argument schemes.

### 3.1 Generating argument schemes

In our context, an argument (explanation) scheme represents an operator tying a tuple of premises (pieces of information provided or approved by the decision maker, or inferred during the process, and some supplementary hypotheses on the reasoning process (model's assumptions, features) to a conclusion (e.g. a recommendation) (see Figure1).

For illustration, the Example 3.1 depicts an example of an explanation supporting additive preferences. In brief, the explanation mechanism rewrites any preference statement $A$ is preferred to $B$ ($A \succsim B$) with a sequence $A \equiv e^0 \succsim e^1 \succsim \cdots \succsim e^{n-1} \succsim e^n \equiv B$, where consecutive alternatives $e^i$ and $e^{i+1}$ are either involved in a dominance relation or differ on at most $p$ criteria (expressing a trade-off), and we call it a "sequence of preference swaps of order at most $p$ and of length $n$ (criteria)" (for more details see [4].

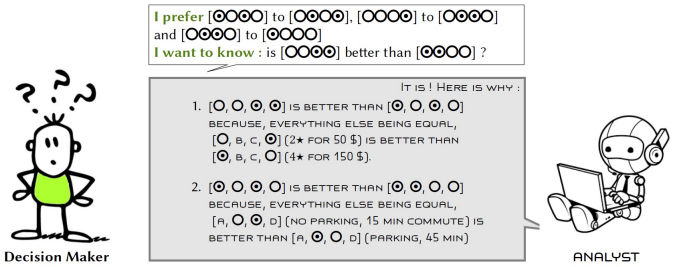


**Figure 4: Explanation supporting additive preferences**

**Example 3.1.** (Ex. 2.1 cont.)

In Figure 4, the agent provides an explanation by decomposing the statements into smaller grain by reasoning *ceteris paribus*. The consecutive actions differ only on 2 criteria. Moreover, for sake of clarity, the attribute values of interest are sorted and encoded : A : 4★ is strong (⊙), 2★ is weak (○); B : yes is strong (⊙), no is weak (○); C : 15 min is strong (⊙), 45 min is weak (○); D : 50 $ is strong (⊙), 150 $ is weak (○).

### 3.2 A dialogue game supporting a human-agent interaction

Among the many approaches upon which decision aiding systems are built, the dialectical nature of decision aiding (with conflicting points of view) and its reliance on clearly identified principles, are important assets to overcome the challenge of accountability. Indeed, the notion of accountability we put forward has strong dialectical and adversarial components – and could aptly be represented as a discussion between the decision maker and an agent discussing critically and in good faith various options.

Under such perspective, a first step towards formalizing such a discussion is the work of [17], where a dialogue game is proposed to formalize the interaction representing a decision aiding situation, involving exchange of different types of preferential information, as well as others locutions such as justification. In particular, the approach allows the artificial agent to use a variety of decision models (able to encompass most of decision situations) to build its recommendation (as opposed to adjusting the parameters of a single model). To account for this, an axiomatic approach is adopted, where the use of a model is triggered by a set of properties that should the decision maker's preferences fulfilled. However, a number of questions are still open, as it discussed in the following.

## 4 CHALLENGES

We now discuss some important research challenges that remain to be addressed to make this approach fully operational.

### 4.1 Planning explanation

To meet the accountability perspective, we investigate the question of what kind of explanation are suited in our setting (see Sect. 3.1). As the requirements may greatly vary from situation to situation (for instance, depending on the criticality of the stakes, the time pressure), and from decision maker to decision maker, we do not believe in the provision of a unique type of explanation, and we propose to build argument schemes depending on the model under use and the user's profile[5, 16]. Another interesting and challenging question is how to present (communicate) explanations to a user? We believe that a promising direction is to approach the problem of explanation generation as a problem of planning [8], where the idea is to find the path that leads to the conclusion. Since our results identified several basic "operators" (under the form of argument schemes), it is thus tempting to adopt this stance and design an explanation planner for our decision-aiding setting. This unified framework could pave the way for a potentially powerful mixture of approaches (using different types of argument schemes within the same line of explanation).

### 4.2 Defeasible reasoning

A key issue in our decision aiding setting is to deal with the imprecision or uncertainty of the processed data [7]. More precisely, the decision maker (DM) sometimes provides information that cannot be fully represented in the preference model. In such a situation, the preference information provided by the DM is said to be "inconsistent", and corresponds to a list of statements provided by the DM for which no combination of values for preference parameters exists to fully restore the DM's assertions. Such "inconsistencies" may occur when, for instance: the DM's statements express conflicting preferences, the DM's point of view is evolving during the interaction process, the DM's reasoning is incompatible with the principles and properties underlying the preference model, etc. Different issues rises: how the system should behave in presence of inconsistency, in the situation where a (family of) model(s) is not able to restore the DM's

preferences? Should we revise the expressed preferences? Should we change the model? Thus, on what principles? How to conduct the elicitation process by taking into account the inconsistency? In particular, the question of generating explanation adds a level of complexity on this question as it becomes legitimate to seek to find/keep the information that will allow to construct "good" explanations, at the end.

Given our objective of accountability, we may have to take the statements of the decision maker at face value if we want to leverage this commitment into a sense of ownership of the recommendation. There are several candidate models to deal with uncertainty and inconsistency, starting with numerical representations of uncertainty, such as probabilistic or possibilistic frameworks [2, 11]. Another approach would rather deal with contradictions inside a logic-based representation of non-monotonic reasoning [19], where statements are defeasible (but maybe should explicitly be defeated during a dialogue). This is in line with the tools proposed by [18] to retrieve maximally consistent subsets of preference information.

### 4.3 Mixed initiative systems

While the incremental elicitation methods already involves a rather simple interaction process whereby the system asks queries to the user, there are new challenges when one wants to integrate explanation facilities.

Indeed, to produce a recommendation, the system questions the user to elicit her preferences and fit it to a model. On the basis of these preferences, the system can produce a recommendation. However, because the recommendation itself can be very large (think of a ranking involving all the options), it is useful to allow incremental partial and/or factored recommendations to be made throughout the interaction, on which the system will seek agreement of the user (e.g. "do we agree that product $p$ is better than any product which color is red?", or " do we agree that subset of options $p_1, p_2, p_3$ should not be considered as the product of choice?"). When the system puts it forward, the user can critique it (preferences may be adjusted, corrected, the option may not be feasible, or not available anymore, etc.) or asks for a justification, which must be provided by the system. As a result, the system must deal with the inherent *revision problem* induced by the possibly incoherent statements (either among themselves, or with the user's assumed preference model) [18]. While current systems equipped with explanation features typically produce justification at the very end of the process – together with their final recommendation– we think that a mixed-initiative system [13] where elicitation, recommendation and explanation are tightly interleaved, is required. This implies to carefully design a protocol which decides exactly how and when the initiative should be given to the user, or kept by the system, and how the different commitments can be agreed upon, or challenged.

### 4.4 Experiments and validation

In our different proposals towards providing explanations [3**?** –5], neither natural language generation, nor in vivo experimentation were investigated. The complexity of explanations was assessed through proxies, such as length, or number of premisses. Moreover, designing an artificial agent with explanations features for a decision aiding purposes will require a validation phase. Thus, we need to carefully elaborate: (i) what can be "good" indicators or criteria to assess and validate the results (for instance, one can choose intuitively to assess the convergence of the interaction by making a compromise between accepting (or not) a recommendation and the time spent to obtain the agreement–however, it is less clear how to assess the impact of introducing an explanation within a recommendation); and moreover (ii) a methodology or a framework of how validation should be implemented. In other terms, how to experiment and/or practice a decision aiding situation with the help of an artificial agent endowed with an explanatory capacity.

## REFERENCES


[1] K. J. Arrow. 1950. A difficulty in the concept of social welfare. *The Journal of Political Economy* (1950), 328–346.

[2] Th. Augustin, F. PA Coolen, and M. CM De Cooman, G.and Troffaes (Eds.). 2014. *Introduction to imprecise probabilities.* John Wiley and Sons.

[3] Kh. Belahcene, Y. Chevaleyre, Ch. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. 2018. Accountable Approval Sorting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18.* International Joint Conferences on Artificial Intelligence Organization, 70–76. https://doi.org/10.24963/ijcai.2018/10

[4] Kh. Belahcene, Ch. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. 2017. Explaining robust additive utility models by sequences of preference swaps. *Theory and Decision* 82, 2 (01 Feb 2017), 151–183.

[5] Kh. Belahcene, Ch. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. 2017. A Model for Accountable Ordinal Sorting. In *Proceedings of the 26 International Joint Conference on Artificial Intelligence.* 814–820. https://doi.org/10.24963/ijcai.2017/113

[6] D. Bouyssou, Th. Marchant, P. Perny, M. Pirlot, A. Tsoukiàs, and Ph. Vincke. 2000. *Evaluation and decision models: a critical perspective.* International Series in Operations Research and Management Science, Vol. 32. Kluwer Academic Publishers.

[7] D. Bouyssou, Th. Marchant, M. Pirlot, A. Tsoukiàs, and Ph. Vincke. 2006. *Evaluation and decision models with multiple criteria: Stepping stones for the analyst.* Springer Verlag.

[8] A. Cawsey. 1993. Planning interactive explanations. *International Journal of Man-Machine Studies* 38, 2 (1993), 169 – 199.

[9] Condorcet. 1785. *Essai sur l'application de l'analyse a la probabilité des decisions rendues a la pluralité des voix [microform] / par M. le Marquis de Condorcet.* Imprimerie royale Paris. cxci, 304 p. ; pages.

[10] M. Cozic and P. Valarcher. 2017. The design of Administrative Algorithms. (2017). the ALGOCIT team. Presented at the Social Responsibility of Algorithms - SRA 2017 interdisciplinary workshop. Paris, Dauphine.

[11] S. Destercke. 2018. A generic framework to include belief functions in preference handling and multi-criteria decision. *International Journal of Approximate Reasoning* 98 (2018), 62 – 77. https://doi.org/10.1016/j.ijar.2018.04.005

[12] B. Goodman and S. Flaxman. 2016. European Union regulations on algorithmic decision-making and a "right to explanation". (June 2016). ArXiv e-prints: 1606.08813.

[13] E. Horvitz. 2000. Uncertainty, Action, and Interaction: In Pursuit of Mixed-Initiative Computing. *Intelligent Systems* (2000), 17–20.

[14] N. Kano, N. Seraku, F. Takahashi, and Sh. Tsuji. 1984. Attractive Quality and Must-Be Quality. *Journal of the Japanese Society for Quality Control* 14, 2 (apr 1984), 147–156.

[15] Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable Algorithms. *University of Pennsylvania Law Review* 165 (2017). https://ssrn.com/abstract=2765268

[16] Ch. Labreuche. 2011. A general framework for explaining the results of a multi-attribute preference model. *Artificial Intelligence Journal* 175 (2011), 1410–1448.

[17] Ch. Labreuche, N. Maudet, W. Ouerdane, and S. Parsons. 2015. A Dialogue Game for Recommendation with Adaptive Preference Models. In *Proceedings AAMAS.* 959–967.

[18] V. Mousseau, L.C. Dias, J. Figueira, C. Gomes, and J.N. Clímaco. 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research* 147, 1 (2003), 72–93.

[19] R. Reiter. 1987. Nonmonotonic Reasoning. *Annual Review of Coputer Science* 2 (1987), 147–186.

[20] A. Tsoukiàs. 2008. From Decision Theory to Decision Aiding Methodology. *European Journal of Operational Research* 187 (2008), 138–161.

[21] S. Wachter, B. Mittelstadt, and L. Floridi. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law* 7, 2 (May 2017), 76–99. http://dx.doi.org/10.1093/idpl/ipx005

[22] D. Walton. 1996. *Argumentation schemes for Presumptive Reasoning.* Mahwah, N. J.,Erlbaum.