From Preference Elicitation to Explaining Decisions: a Dialectical Perspective

Habilitation à Diriger les Recherches Defense

Wassila Ouerdane

December 8th, 2022

- My research domain: Artificial Intelligence (*Knowledge Representation and Reasoning*), Decision Theory;
- Focus of today: our contributions in *Multiple Criteria Decision Aiding*

Multiple Criteria Decision Aiding (MCDA)

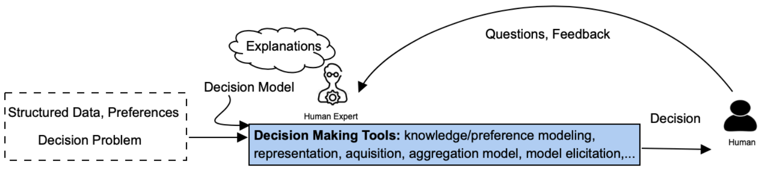
- At least two actors: an expert, a user;
- set of alternatives/options described (evaluated) on several conflicting point of view/ criteria;

	Comfort	Restaurant	Commute time	Cost
h_A	$4^\star$	no	35 min	120 \$
h_B	$4^\star$	yes	50 min	160 \$
h_C	$2^\star$	yes	20 min	50 \$
h_D	$2^\star$	no	30 min	40 \$

- A decision problem: is option h_A better than option h_B? Is option h_C good enough? ...
- Sparse preferences between some options;
- Aggregation model containing aggregation procedures

Multiple Criteria Decision Aiding (MCDA)

- At least two actors: an expert, a user;
- set of alternatives/options described (evaluated) on several conflicting point of view/ criteria;
- A decision problem: is option h_a better than option h_b? Is option h_c good enough? ...
- Sparse preferences between some options;
- Aggregation model containing aggregation procedures



An illustrative Example

Options	Size	Material	Price	Colour	Style
a	small	Steel	450	Red	Classical
b	big	Leather	300	White	Fashion
c	medium	Steel	320	Pink	Classical
d	small	Leather	390	Pink	Sport

(1) **DA**: Given your information, *b* is the best option.
(2) **DM**: Why is that the case?
(3) **DA**: Because *b* is globally better than all other options
(4) **DM**: What does that mean?
(5) **DA**: Well... *b* is top on a majority of criteria considered: the price, the colour, and especially the style, it is so trendy!
(6) **DM**: But, why *b* is better than *c* on the price?
(7) **DA**: Because *c* is 20 euros more expensive than *b*.
(8) **DM**: I agree, but I see that the guarantee is very expensive especially for this watch. In fact I'm not sure to want the guarantee.
(9) **DA**: But *c* remains 5 euros more expensive than b.
(10) **DM**: I see, but this difference is not significant. And also I changed my mind: I would rather to have a classical model, I think it's more convenient for a daily use.
(11) **DA**: OK. In this case I recommend *c* as the best choice.
(12) **DM**: ...

An illustrative Example

Options	Size	Material	Price	Colour	Style
a	small	Steel	450	Red	Classical
b	big	Leather	300	White	Fashion
c	medium	Steel	320	Pink	Classical
d	small	Leather	390	Pink	Sport

(1) **DA**: Given your information, *b* is the best option.

(2) **DM**: Why is that the case?

(3) **DA**: Because *b* is globally better than all other options

(4) **DM**: What does that mean?

(5) **DA**: Well... *b* is top on a majority of criteria considered: the price, the colour, and especially the style, it is so trendy!

(6) **DM**: But, why *b* is better than *c* on the price?

(7) **DA**: Because *c* is 20 euros more expensive than *b*.

(8) **DM**: I agree, but I see that the guarantee is very expensive especially for this watch. In fact I'm not sure to want the guarantee.

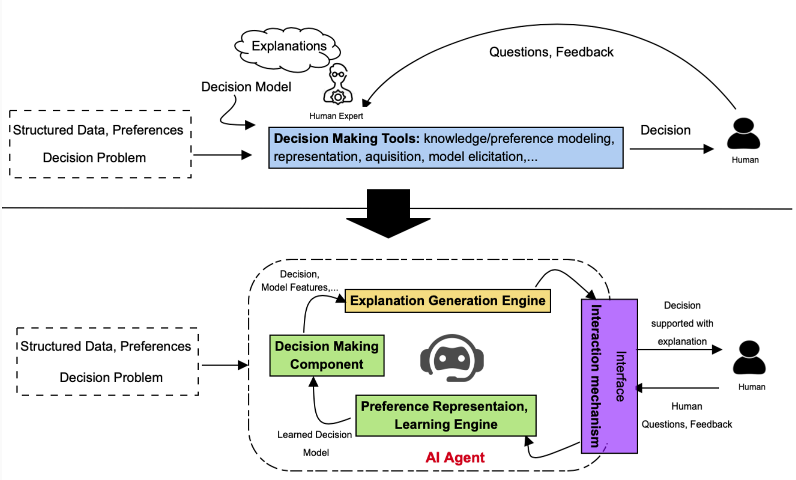
(9) **DA**: But *c* remains 5 euros more expensive than b.

(10) **DM**: I see, but this difference is not significant. And also I changed my mind: I would rather to have a classical model, I think it's more convenient for a daily use.

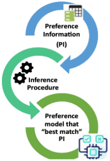
(11) **DA**: OK. In this case I recommend *c* as the best choice.

(12) **DM**: ...

MCDA with an Artificial agent !



Our research issues

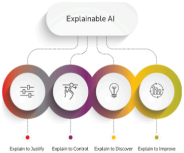


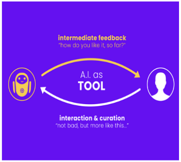
Question 1

For a given decision situation, if a given decision model is relevant to structure the decision maker's preferences, what should be the parameters' values to fully specify this model that corresponds to the decision-maker viewpoint?

Question 2

Given a decision model and a set of preference information, is there a principled way to define simple complete explanations supporting a recommendation/decision?





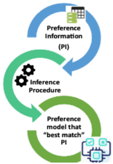
Question 3

How to equip an artificial agent with adaptive behavior and model the system's reasoning to allow ``efficient'' interaction with a user within a decision-aiding situation?

Source images: Explainable AI, Human in the Loop

Preference Learning and Elicitation

Research issues

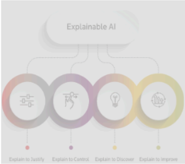


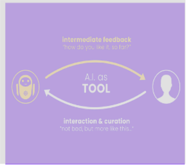
Question 1

For a given decision situation, if a given decision model is relevant to structure the decision maker's preferences, what should be the parameters' values to fully specify this model that corresponds to the decision-maker viewpoint?

Question 2

Given a decision model and a set of preference information, is there a principled way to define simple complete explanations supporting a recommendation/decision?





Question 3

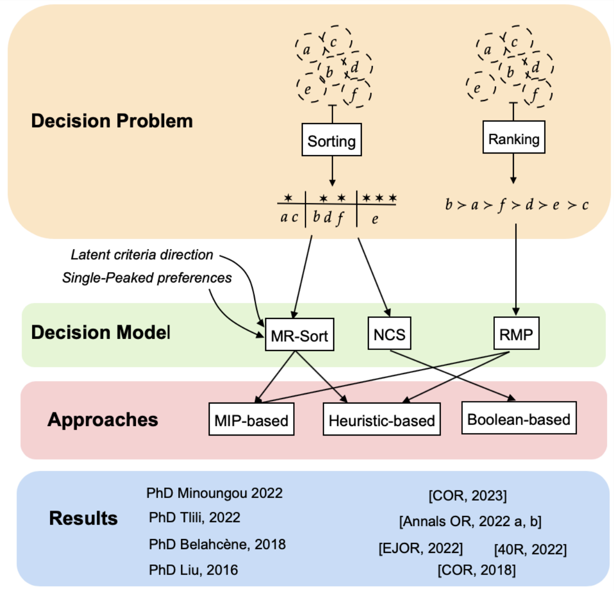
How to equip an artificial agent with adaptive behavior and model the system's reasoning to allow ``efficient'' interaction with a user within a decision-aiding situation?

Research issues

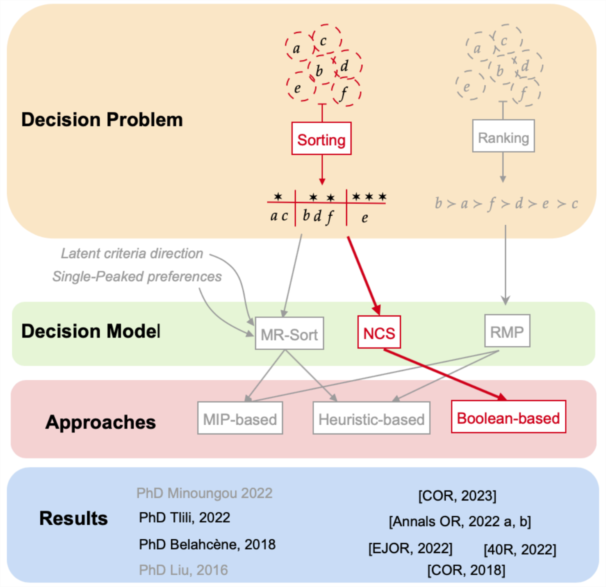
- Preference modeling issue: how to represent the user's preferences?
- Computational issue: how to build and provide efficient device?

Our Contributions: Mathematical and Computational Tools

Preference Learning– Overview of our Results



Overview of our Results–Focus NCS



NCS: Non-Compensatory Sorting

- an ordered set $C^1 \prec \cdots \prec C^p$ of p predefined categories
- a set of objects to be sorted : $\mathbb{X} = \prod_{i \in \mathcal{N}} \mathbb{X}_i$ with $\mathcal{N} = \{1, \ldots, n\}$
- A total preorder, noted $\succsim_i$ on $X_i,\ i \in \mathcal{N}$
- approved sets $\langle \mathcal{A}_i^k \rangle_{i \in \mathcal{N},\ k \in [2..p]}$ defined by a set of limiting profiles $\langle b_i^k \rangle_{i \in \mathcal{N},\ k \in [2..p]}$
- a set of sufficient coalitions $\langle \mathcal{T}^k \rangle_{k \in [2..p]}$ declined per boundary.

$$NCS_\omega(x) = C^k \Leftrightarrow \begin{cases} \{i \in \mathcal{N} : x \in \mathcal{A}_i^k\} & \in \mathcal{T}^k \\ \text{and } \{i \in \mathcal{N} : x \in \mathcal{A}_i^{k+1}\} & \notin \mathcal{T}^{k+1} \end{cases} \tag{1}$$

where $\omega = (\langle \mathcal{A}_i^k \rangle_{i \in \mathcal{N},\ k \in [2..p]}, \langle \mathcal{T}^k \rangle_{k \in [2..p]})$

[Bouyssou and Marchant, 2007a, 2007b]

NSC – Learning/Disaggregation Step

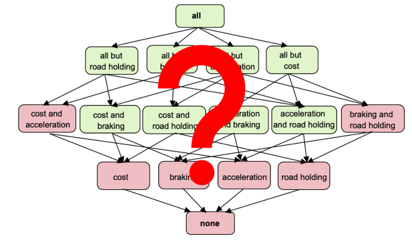
Inputs: Reference assignments

	Cost	Acceleration	Breaking	Road hold	Category
m_1	16 973€	29.0 sec.	2.66	2.5	$\star\star$
m_2	18 342€	30.7 sec.	2.33	3	$\star$
m_3	15 335€	30.2 sec.	2	2.5	$\star\star$
m_4	18 971€	28.0 sec.	2.33	2	$\star\star$
m_5	17 537€	28.3 sec.	2.33	2.75	$\star\star\star$
m_6	15 131€	29.7 sec.	1.66	1.75	$\star$

NSC – Learning/Disaggregation Step

Inputs: Reference assignments

	Cost	Acceleration	Breaking	Road hold	Category
m_1	16 973€	29.0 sec.	2.66	2.5	$\star\star$
m_2	18 342€	30.7 sec.	2.33	3	$\star$
m_3	15 335€	30.2 sec.	2	2.5	$\star\star$
m_4	18 971€	28.0 sec.	2.33	2	$\star\star$
m_5	17 537€	28.3 sec.	2.33	2.75	$\star\star\star$
m_6	15 131€	29.7 sec.	1.66	1.75	$\star$



Profile	C	A	B	R
$\star/\star\star$	?	?	?	?
$\star\star/\star\star\star$	?	?	?	?

Expected Outputs: Set of sufficient coalitions + Set of profiles

The Inv-NCS Problem

Finding a solution to an *instance* of the Inv-NCS problem:

$$(\mathcal{N}, \mathbb{X}, \langle \succsim_i \rangle_{i \in \mathcal{N}}, \mathbb{X}^\star, \{C^1 \prec \ldots \prec C^p\}, \alpha)$$

where:

- $\mathcal{N}$ is a set of *criteria*;
- $\mathbb{X}$ is a set of *alternatives*;
- $\langle \succsim_i \rangle_{i \in \mathcal{N}} \in \mathbb{X}^2$ are *preferences* on criterion *i*, $i \in \mathcal{N}$, $\succsim_i \subset \mathbb{X}^2$ is a total pre-ordering of alternatives according to this criterion;
- $\mathbb{X}^\star \subset \mathbb{X}$ is a finite set of reference alternatives;
- $\{C^1 \prec \ldots \prec C^p\}$ is a finite set of categories totally ordered by *exigence level*.
- $\alpha : \mathbb{X}^\star \to \{C^1 \prec \ldots \prec C^p\}$ is an *assignment* of $\mathbb{X}^\star$ to the categories.
  for a given category C^h, $\alpha^{-1}(C^h) = \{x \in \mathbb{X}^\star : x \in C^h\}$.

Inv-NCS – Overview of our Results

Two SAT-based formulations [Belahcène et al., 2018a, 2018c; Tlili et al., 2022]

1. A SAT formulation based on *Coalitions*
   - Explicit representation of the parameter space
2. A SAT formulation based on *Pairwise Separation*
   - *Approved sets are given*;
   - Intuition: for every pair of alternatives (g accepted, b rejected), is there at least one criterion approving g but not b?
   - Compact SAT formulation; and Inv-NCS is NP-complete

$\rightsquigarrow$ The formulations are more efficient than state-of-the art MIP-based approach.

MaxSAT relaxations [Tlili et al., 2022]

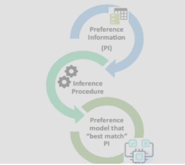
- Take into account "noisy" data (imperfection in the assessment of performance, mistaken assignment, ...)
- Retrieve the model that restores "the most" examples of the Learning set.

Preference Learning – The Other contributions

- Majority Sorting Rule (MR-Sort)
  - Parameters to learn: limiting profiles $\langle b \rangle$, weights (w), threshold (λ);
  - Issue: How to deal with an ordered partition $\mathcal{C} = (C^1, \ldots, C^h, \ldots, C^p)$ that is not monotone w.r.t the natural order of the criterion scale?
  - Contribution: taking into account single-peaked preferences – an exact approach and a heuristic approach [Minoungou et al., 2022].
- Ranking with Multiple reference Points (RMP)
  - Parameters to learn: weights, reference points, and the lexicographic order on reference points;
  - Contribution: A MIP-based approach [Olteanu et al., 2021], a heuristic-based approach [Liu et al., 2014], and a Boolean-based approach [Belahcène et al., 2023a]

XAI & MCDA

Our research issues

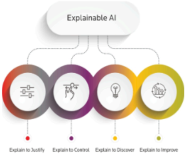


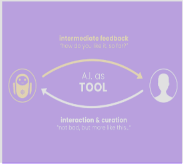
Question 1

For a given decision situation, if a given decision model is relevant to structure the decision maker's preferences, what should be the parameters' values to fully specify this model that corresponds to the decision-maker viewpoint?

Question 2

Given a decision model and a set of preference information, is there a principled way to define simple complete explanations supporting a recommendation/decision?





Question 3

How to equip an artificial agent with adaptive behavior and model the system's reasoning to allow ``efficient'' interaction with a user within a decision-aiding situation?

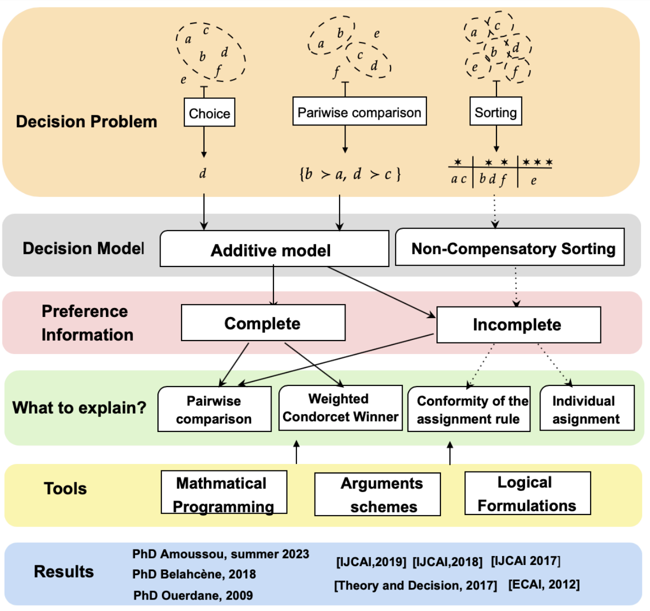
Explanation – Issues

- **Computation**: *How difficult is it to produce an explanation?*
- **Simplicity**: *Can we keep the explanations simple enough* to be processed by a human decision-maker?
- **Completeness**: *Can we explain every 'true' result, that can be deduced from the preference information and the model?*
- **Soundness**: *Could we explain 'false' results, claiming the impossibility of an event that could happen or the possibility of an event that cannot happen?*

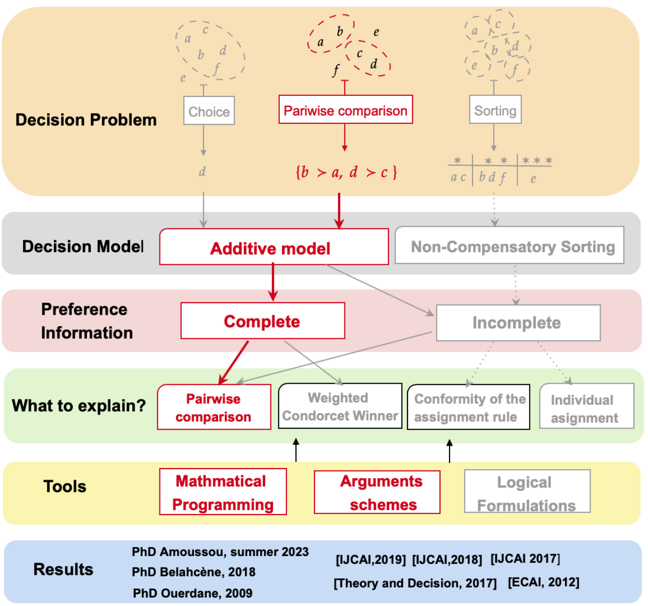
Explanation – Key Principles [Coste-Marquis and Marquis, 2020; Miller, 2019]

- Explanation shall be rigorous (important decision)
  $\rightsquigarrow$ One shall bring **proof** (complete explanation);
- Explanation shall be understandable
  $\rightsquigarrow$ One shall define a language which relates directly to the **preferential information** (e.g. not include the weights), and be conveyed in an **expressive** language to the recipient of this explanation.
- Explanation shall be relevant
  $\rightsquigarrow$ One shall define what could be **pertinent** to focus on within the decision situation.
- Explanation shall be simple
  $\rightsquigarrow$ One shall define different **levels of complexity**. We want explanations to be “easy to process” by the recipient of the explanation.

Explanation in MCDA – Our Contributions



Explanation in MCDA – Our Contributions



Explanation in MCDA – Additive Model

- Preference derives from a **value model**

$$\exists\, V \text{ s.t. } x \succsim y \iff V(x) \geq V(y)$$

- Value is **additive** (i.e. $V(x) = \sum_i v_i(x_i)$)
- Case: *binary evaluation*

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$$\omega = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$$

Additive Model– Explaining a Pairwise Comparison

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$$\omega = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$$

$$\left.\begin{array}{l} \omega(s_1) = 128 + 77 + 37 = 242 \\ \omega(s_2) = 126 + 41 + 37 = 204 \end{array}\right\} \mathbf{s_1 \succ s_2}$$

- Encoding: a vector $\{-1, 0, +1\}^n$ of arguments in favour ($\boldsymbol{pro}$) or against ($\boldsymbol{con}$) or neutral ($\boldsymbol{neu}$).

$$\boldsymbol{pro}_{s_1} = \{a, c\}, \boldsymbol{con}_{s_1} = \{b, f\}, \text{ while } neu = \{d, e, g\}$$

Additive Model– Explaining a Pairwise Comparison

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$$\omega = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$$

$$\left.\begin{aligned} \omega(s_1) &= 128 + 77 + 37 = 242 \\ \omega(s_2) &= 126 + 41 + 37 = 204 \end{aligned}\right\} \mathbf{s_1 \succ s_2}$$

- Encoding: a vector $\{-1, 0, +1\}^n$ of arguments in favour ($\boldsymbol{pro}$) or against ($\boldsymbol{con}$) s_1, or neutral ($\boldsymbol{neu}$).

$$\boldsymbol{pro}_{s_1} = \{\boldsymbol{a}, \boldsymbol{c}\}, \ \boldsymbol{con}_{s_1} = \{\boldsymbol{b}, \boldsymbol{f}\}, \text{ while } \boldsymbol{neu} = \{\boldsymbol{d}, \boldsymbol{e}, \boldsymbol{g}\}$$

Question: *why s_1 is preferred to s_2 ?*

Additive Model– Explaining a Pairwise Comparison

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$w = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$

$$\left.\begin{array}{l} \omega(s_1) = 128 + 77 + 37 = 242 \\ \omega(s_2) = 126 + 41 + 37 = 204 \end{array}\right\} \mathbf{s_1 \succ s_2}$$

$$\boldsymbol{pro}_{s_1} = \{\boldsymbol{a}, \boldsymbol{c}\}, \boldsymbol{con}_{s_1} = \{\boldsymbol{b}, \boldsymbol{f}\}, \boldsymbol{neu} = \{\boldsymbol{d}, \boldsymbol{e}, \boldsymbol{g}\}$$

Our proposal –Step-wise Explanations:

$$s_1\,(acg) \succ (bfg)\,s_2$$

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$w = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$

$$\left.\begin{aligned} \omega(s_1) &= 128 + 77 + 37 = 242 \\ \omega(s_2) &= 126 + 41 + 37 = 204 \end{aligned}\right\} \mathbf{s_1 \succ s_2}$$

$$\boldsymbol{pro}_{s_1} = \{\boldsymbol{a}, \boldsymbol{c}\}, \boldsymbol{con}_{s_1} = \{\boldsymbol{b}, \boldsymbol{f}\}, \boldsymbol{neu} = \{\boldsymbol{d}, \boldsymbol{e}, \boldsymbol{g}\}$$

Our proposal –Step-wise Explanations:

$$s_1\,(acg) \succ (bcg) \succ (bfg)\,s_2$$

Additive Model– Explaining a Pairwise Comparison

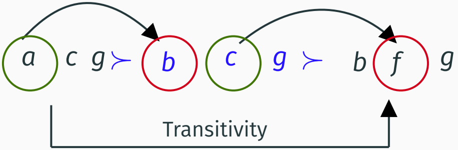
	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$w = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$

$$\left.\begin{array}{l} \omega(s_1) = 128 + 77 + 37 = 242 \\ \omega(s_2) = 126 + 41 + 37 = 204 \end{array}\right\} \mathbf{s_1 \succ s_2}$$

$\boldsymbol{pro}_{s_1} = \{\boldsymbol{a}, \boldsymbol{c}\}, \boldsymbol{con}_{s_1} = \{\boldsymbol{b}, \boldsymbol{f}\}, \boldsymbol{neu} = \{\boldsymbol{d}, \boldsymbol{e}, \boldsymbol{g}\}$

Our proposal –Step-wise Explanations:



Additive Model– Explaining a Pairwise Comparison

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$w = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$

$$\left.\begin{array}{l}\omega(s_1) = 128 + 77 + 37 = 242 \\ \omega(s_2) = 126 + 41 + 37 = 204\end{array}\right\} \mathbf{s_1 \succ s_2}$$

$\boldsymbol{pro}_{s_1} = \{\boldsymbol{a}, \boldsymbol{c}\}$, $\boldsymbol{con}_{s_1} = \{\boldsymbol{b}, \boldsymbol{f}\}$, $\boldsymbol{neu} = \{\boldsymbol{d}, \boldsymbol{e}, \boldsymbol{g}\}$

Our proposal –STEP-WISE Explanations:

$$s_1\,(acg) \succ (bcg) \succ (bfg)\,s_2 \ ✓$$

$$s_1\,(acg) \succ (bef) \succ (bfg)\,s_2 \ ✗$$

the 1^{st} comparison is *complex* as it involves 6 criteria.

$$s_1\,(acg) \succ (abc) \succ (bfg)\,s_2 \ ✗$$

$$(242 = \omega_a + \omega_c + \omega_g < \omega_a + \omega_b + \omega_c = 331)$$

Additive Model– Explaining a Pairwise Comparison

	a	b	c	d	e	f	g
s_1	✓	✗	✓	✗	✗	✗	✓
s_2	✗	✓	✗	✗	✗	✓	✓

$w = \langle 128, 126, 77, 59, 52, 41, 37 \rangle$

$$\left.\begin{array}{l}\omega(s_1) = 128 + 77 + 37 = 242 \\ \omega(s_2) = 126 + 41 + 37 = 204\end{array}\right\} \mathbf{s_1 \succ s_2}$$

$$\boldsymbol{pro}_{s_1} = \{\boldsymbol{a}, \boldsymbol{c}\}, \boldsymbol{con}_{s_1} = \{\boldsymbol{b}, \boldsymbol{f}\}, \boldsymbol{neu} = \{\boldsymbol{d}, \boldsymbol{e}, \boldsymbol{g}\}$$

Our proposal –STEP-WISE Explanations:

$$s_1\,(acg) \succ (bcg) \succ (bfg)\,s_2$$

- *S_1 (**acg**) is preferred over **bcg**, and that **bcg** is preferred over (**bfg**) S_2, so that our conclusion should hold, following a transitive reasoning.*
- exhibits a collection of statements aiming at proving the decision.

Additive Model– Explaining a Pairwise Comparison

- Break down the recommendation into "simple" statements;
- the sequence of statements formally support the recommendation.

$$\underbrace{\overbrace{[(\boldsymbol{acg}, \boldsymbol{bcg}), (\boldsymbol{bcg}, \boldsymbol{bfg})]}^{\text{premises}} \xrightarrow{tr} \overbrace{(\boldsymbol{acg}, \boldsymbol{bfg})}^{\text{conclusion}}}_{\text{Argument Scheme}}$$

- Principle-based approach: each scheme is attached to a number of well understood properties of the underlying decision model (e.g. transitivity)
- Cognitively bounded: the statements are constrained to remain "easy" to grasp

Additive Model– Covering scheme

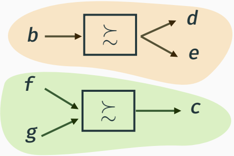
For the conclusion: (bfg, cde). The premise $[(fg, c), (b, de)]$ constitutes a covering scheme:

$$(fg, c), (b, de) \xrightarrow{\text{cov}} (bfg, cde)$$

Proof diagram

$$\left.\begin{aligned} fg \succ c &\xrightarrow{cp} bfg \succ bc \\ b \succ de &\xrightarrow{cp} bc \succ cde \end{aligned}\right\} \xrightarrow{tr} bfg \succ cde$$

Visual representation

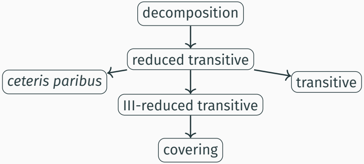


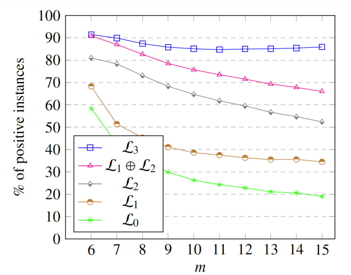
Narrative representation

“As, all other things being equal, having free breakfast and wifi access is preferred to having a swimming pool ($\mathbf{fg}$, $\mathbf{c}$)*, and being close to the city is preferred than having a sports hall and a low tourist tax* ($\mathbf{b}$, $\mathbf{de}$)*, we get that* ($\mathbf{bfg}$, $\mathbf{cde}$)*”*

Our Contributions– Argument Schemes for the Additive Model

m	Minimum	Median	Maximum	$\lvert \mathcal{T}^m_{\succ} \setminus \mathcal{A}_{\succ} \rvert$
4	66.7%	66.7%	100%	3
5	72.0%	80.0%	100%	25
6	78.46%	84.62%	100%	130



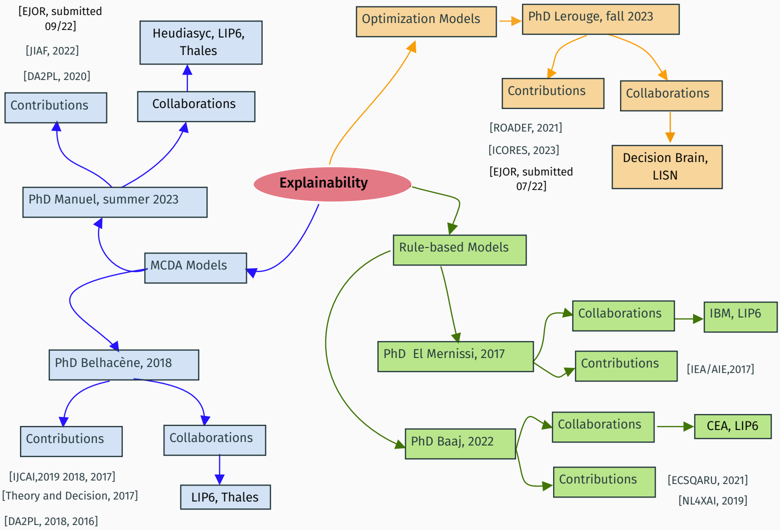


Scheme	Properties	Requirements for correctness
decomposition	commutative	additive
reduced transitive		transitive + cancellation
III-red. transitive	III	transitive + cancellation
covering	commutative, III	transitive + cancellation
transitive		transitive
ceteris paribus		cancellation

For a *fully specified* model:

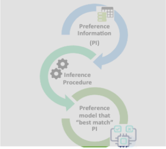
- # argument schemes $\rightsquigarrow$ # patterns of reasoning
- # classes of difficulty of statements
- Complexity results on the existence of an explanation;
- Computing Explanations using ILP;
- Promising experimental results on the explanatory power of the *covering scheme.*

Our Explainability Contributions– The Big Picture



Dialectical Tools

Our research issues

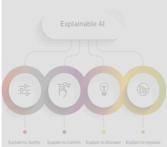


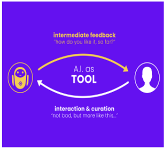
Question 1

For a given decision situation, if a given decision model is relevant to structure the decision maker's preferences, what should be the parameters' values to fully specify this model that corresponds to the decision-maker viewpoint?

Question 2

Given a decision model and a set of preference information, is there a principled way to define simple complete explanations supporting a recommendation/decision?





Question 3

How to equip an artificial agent with adaptive behavior and model the system's reasoning to allow ``efficient'' interaction with a user within a decision-aiding situation?

Among Challenges

With multiple criteria context, there are many possible decision models. So when deciding whether $a \succ b$ globally, you may use e.g.:

- simple majority (π_{SM})—count criteria for $a \succ b$ vs. $b \succ a$
- simple weighted majority (π_{SWM})—same but with weighted criteria
- mean model (π_M)—sum of utilities of items for each criterion
- weighted sum model (π_{WS})—same but with weighted criteria
- outranking model—similar to π_{SWM} but includes a veto notion
- and many more...

Among Challenges!

With multiple criteria context, there are many possible decision models. So when deciding whether $a \succ b$ globally, you may use e.g.:

- simple majority (π_{SM})—count criteria for $a \succ b$ vs. $b \succ a$
- simple weighted majority (π_{SWM})—same but with weighted criteria
- mean model (π_M)—sum of utilities of items for each criterion
- weighted sum model (π_{WS})—same but with weighted criteria
- outranking model—similar to π_{SWM} but includes a veto notion
- and many more...

Questions:

- *is there a principled way to do deal with the multiplicity of models?*
- *how, in practice, should such interaction be regulated?*

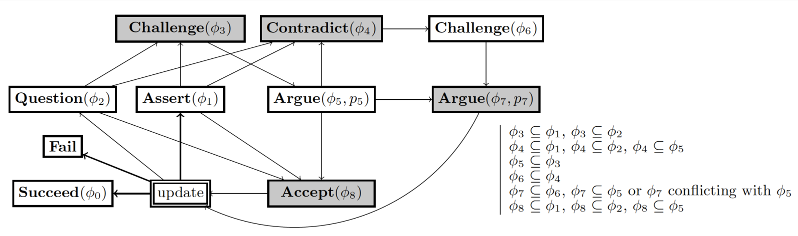
Our contributions – Navigating among Decision Models

- We adopt an *axiomatic approach*
- *Idea*: to each model can be attached properties satisfied, *e.g.*:
  - cardinality: the difference of performance is meaningful
  - non anonymity: criteria are not exchangeable
  - Veto property
  - ...
- least specific model is the one that satisfies more properties;

$$
\begin{tikzcd}
\pi_{OR}\ (\bot, \top, \top) \arrow[d, dash] & & \pi_{WS}\ (\top, \top, \bot) \arrow[dll, dash] \arrow[d, dash] \\
\pi_{SWM}\ (\bot, \top, \bot) \arrow[dr, dash] & & \pi_{M}\ (\top, \bot, \bot) \arrow[dl, dash] \\
& \pi_{SM}\ (\bot, \bot, \bot) &
\end{tikzcd}
$$

Our contributions – Argumentation-based Dialogue

- Rely on Multi-Agent Systems tools: interaction protocol, argumentation theory,



Speech acts at each iteration (grey nodes: DM, white nodes: DA).

Key locutions:

- **Challenge(ϕ)**—requests some statement that can serve as a basis for justifying or explaining ϕ.
- **Argue(ϕ, p)**—p is an explanation of ϕ.

How it works? Example

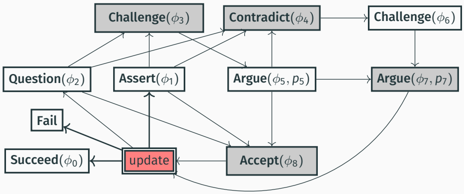
Suppose that a user has to **rank** four options, e.g. hotels $\{a, b, c, d\}$ evaluated on a set of criteria:
$\{c_1 : price, c_2 : location, c_3 : stars, c_4 : breakfast, c_5 : rating\}$.

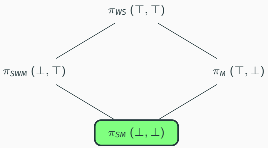
	a	b	c	d
price	80	180	120	60
location	close	far	very far	very close
stars	$\star$	$\star\star\star\star$	$\star\star\star$	$\star\star$
breakfast	coffee machine	mini buffet	full buffet	none
rating	120/300	3/300	267/300	278/300

Which provides default preferential information:

price :	$d \succ_{c_1} a \succ_{c_1} c \succ_{c_1} b;$
location:	$d \succ_{c_2} a \succ_{c_2} b \succ_{c_2} c;$
stars:	$b \succ_{c_3} c \succ_{c_3} a \succ_{c_3} d;$
breakfast:	$c \succ_{c_4} b \succ_{c_4} a \succ_{c_4} d;$
rating:	$b \succ_{c_5} a \succ_{c_5} c \succ_{c_5} d.$

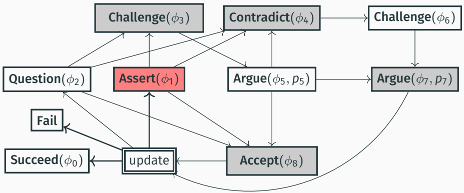
How it works? Example

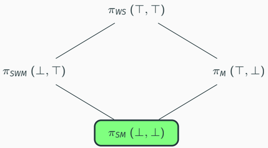




$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

How it works? Example

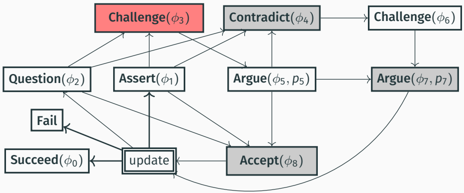


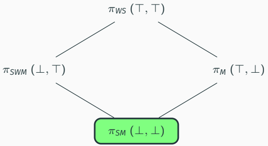


$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

DA:**Assert**$(\phi_1^{(1)})$, $\phi_1^{(1)} = \phi_c^{(1)}$

How it works? Example

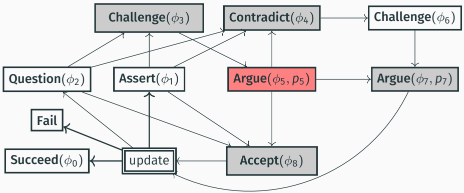


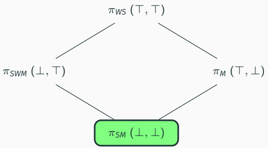


$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

DM:**Challenge**$(\phi_3^{(1)})$, $\phi_3^{(1)} = \{[b \succ a]\}$
<u>Note</u>: $\phi_3^{(1)} \subseteq \phi_c^{(1)} = [b \succ a \succ c \succ d]$

How it works? Example

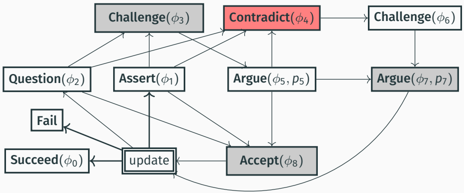


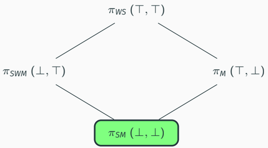


$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

DA:**Argue**$(\phi_5^{(1)}, p_5^{(1)})$,
$\phi_5^{(1)} = \{[b \succ a]\}$,
$p_5^{(1)} = \{[b \succ_{c_3} a], [b \succ_{c_4} a], [b \succ_{c_5} a]\}$

How it works? Example

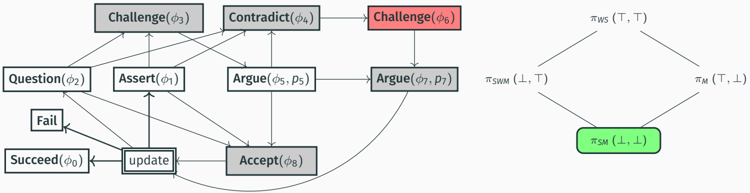




$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

DM:**Contradict**$(\phi_4^{(1)})$, $\phi_4^{(1)} = \{[a \succ b]\}$

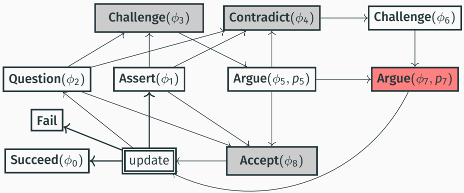
How it works? Example

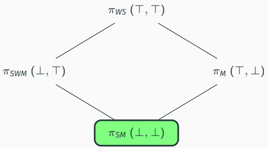


$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

DA:**Challenge**($\phi_6^{(1)}$), $\phi_6^{(1)} = \{[a \succ b]\}$

How it works? Example

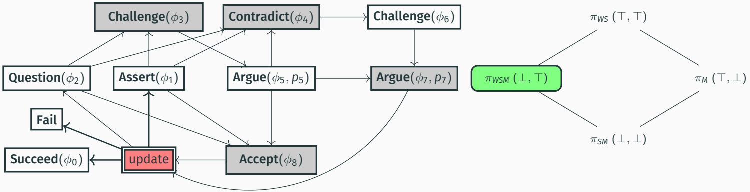




$\mathcal{KB}_P^{(1)}$ contains all statements $[x \succ_{c_i} y]$
$\mathcal{KB}_\phi^{(1)} = \emptyset$
$\phi_c^{(1)} = [b \succ a \succ c \succ d]$
$miss(\phi_c^{(1)}) = \phi_c^{(1)}$

DM:**Argue**$(\phi_7^{(1)}, p_7^{(1)})$, $\phi_7^{(1)} = \{[a \succ b]\}$
$p_7^{(1)} = \{[a \succ_{c_1} b], [a \succ_{c_2} b]$
$[c_1 = strong], [c_2 = strong]\}$

How it works? Example



$\mathcal{KB}_P^{(2)} = \mathcal{KB}_P^{(1)} \cup \{[c_1, c_2 = strong]\}$
$\mathcal{KB}_\phi^{(2)} = \emptyset$
$\phi_c^{(2)} = [d \succ a \succ b \succ c]$

$miss(\phi_c^{(2)}) = \phi_c^{(2)}$

<u>Note</u>: α_{c_1} and α_{c_2} set to 2
$\alpha_{c_3}, \alpha_{c_4}, \alpha_{c_5}$ set to 1
so $d \succ a$
$\vdots$

Dialectical Vision – Summary

With the idea that preferential information feedback is triggered by the user facing actual recommendations, we formalized:

- a conceptual idea for navigating among models [Labreuche et al., 2015]
- an interaction protocol based on argumentation theory [Labreuche et al., 2015; Ouerdane et al., 2011], where:
  - rules and conditions under which we can have a “coherent” interaction in a decision support context, are specified
  - Termination can be guaranteed with very few assumptions
  - Critics/feedback through Critical Questions (attached to argument schemes).

Summary

Summary of Our Contributions

- **Axe 1– Methods for representing, acquiring and learning preferences**
  - Formal theory about preferences (representation, learning) and decisions
  - **Domains:** Decision Theory, MCDA, Operational Research;
- **Axe 2– Methods for constructing and generating explanations.**
  - Formal language to communicate the results (recommendations) and “convince” the user.
  - **Domains:** Artificial Intelligence (KRR[1], Argumentation Theory, Logic)
- **Axe 3– Methods and tools for structuring and conducing the interaction.**
  - Formal language to represent the dialogue/interactions and its outcomes;
  - **Domains:** Artificial Intelligence (KRR, MAS[2], Argumentation theory)

[1] Knowledge Representation and Reasoning
[2] Multi-Agent Systems

Perspectives

Perspectives

Main topic: Explanation-based mixed initiative interaction

How to?

- Interleave learning, recommendation and explanation tasks?
- Express and present an explanation?
- Model and manage inconsistency, uncertainty?
- Assess and evaluate the outcomes?
- ...

Perspectives

Main topic: Explanation-based mixed initiative interaction

For *what*?

- PhD Thesis of Dao Thauvin. *Explanatory dialogue for the interpretation of visual scenes.* Co-supervision with Stephane Herbin (ONERA) and Céline Hudelot (MICS). – Start 11/2022.
- Keywords: Computer Vision, XAI, Argumentation-based Dialogue

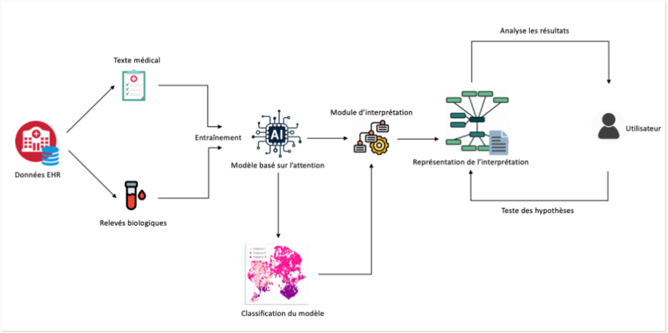
Virtual user	Perceptual system
What's in the room?	There is a table with a vase and another object upon the table.
It is probably a cat or a dog.	There is also a furry animal sleeping under the table. Do you know what it can be?
Yes. Check if it has a long tail.	Should I get closer to describe it better?
It is probably a cat.	I am getting closer. Yes it has a long tail.

Perspectives

Main topic: Explanation-based mixed initiative interaction

For *what*?

- PhD Thesis Charlotte Calye. *Interpretable AI methods for medical research on autoimmune diseases*. Supervision, in collaboration with ScientaLab and Céline Hudelot (MICS) – Start 02/2023.
- Keywords: EHR (Electronic Health Records), XAI, Dialog Systems.



All this was not possible without...

- All my PhD students;
- My co-authors and colleagues;
- Family and friends.

Thank You for your Attention

References

Bibliography

Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2017a). Explaining robust additive utility models by sequences of preference swaps. *Theory and Decision*, *82*(2), 151–183.

Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2017b). A model for accountable ordinal sorting. *Proceedings of the 26th IJCAI*, 814–820.

Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2018a). Accountable approval sorting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)*.

Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2018b). An efficient SAT formulation for learning multiple criteria non-compensatory sorting rules from examples. *Computers & Operations Research*, *97*, 58–71.

Bibliography (cont.)

Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2018c). An efficient SAT formulation for learning multiple criteria non-compensatory sorting rules from examples. *Computers & Operations Research*, *97*, 58–71.

Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2019). Comparing options with argument schemes powered by cancellation. *Proceedings of IJCAI-19*, 1537–1543.

Belahcène, K., Mousseau, V., Ouerdane, W., Pirlot, M., & Sobrie, O. (2023a). Ranking with multiple reference points: Efficient sat-based learning procedures. *Computers & Operations Research*, *150*, 106054.

Belahcène, K., Mousseau, V., Ouerdane, W., Pirlot, M., & Sobrie, O. (2023b). Ranking with multiple reference points: Efficient sat-based learning procedures. *Computers & Operations Research*, *150*, 106054.

Bibliography (cont.)

Bouyssou, D., & Marchant, T. (2007a). An axiomatic approach to noncompensatory sorting methods in MCDM, I: the case of two categories. *European Journal of Operational Research*, *178*(1), 217–245.

Bouyssou, D., & Marchant, T. (2007b). An axiomatic approach to noncompensatory sorting methods in MCDM, II: more than two categories. *European Journal of Operational Research*, *178*(1), 246–276.

Coste-Marquis, S., & Marquis, P. (2020). From Explanations to Intelligible Explanations [Workshop at KR'20]. *1st International Workshop on Explainable Logic-Based Knowledge Representation (XLoKR'20)*.

Krantz, D., Luce, R., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1: Additive and polynomial representations). Academic Press, New York.

Labreuche, C., Maudet, N., & Ouerdane, W. (2011). Minimal and complete explanations for critical multi-attribute decisions. *ADT*, 121–134.

Bibliography (cont.)

Labreuche, C., Maudet, N., & Ouerdane, W. (2012). Justifying dominating options when preferential information is incomplete. *ECAI 2012.*, 486–491.

Labreuche, C., Maudet, N., Ouerdane, W., & Parsons, S. (2015). A dialogue game for recommendation with adaptive preference models. *Proceedings AAMAS*, 959–967.

Leroy, A., Mousseau, V., & Pirlot, M. (2011). Learning the parameters of a multiple criteria sorting method. *International Conference on Algorithmic Decision Theory*, 219–233.

Liu, J., Ouerdane, W., & Mousseau, V. (2014). A metaheuristic approach for preference learning in multicriteria ranking based on reference points. *Proceedings of the 2nd workshop "From multiple criteria Decision Aid to Preference Learning" (DA2PL)*, 76–86.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38.

Bibliography (cont.)

- Minoungou, P., Mousseau, V., Ouerdane, W., & Scotton, P. (2020). Learning an MR-sort model from data with latent criteria preference direction. *The 5th workshop from multiple criteria Decision Aid to Preference Learning (DA2PL).*
- Minoungou, P., Mousseau, V., Ouerdane, W., & Scotton, P. (2022). A MIP-based approach to learn MR-Sort models with single-peaked preferences. *Annals of Operations Research.*
- Olteanu, A. L., Belahcène, K., Mousseau, V., Ouerdane, W., Rolland, A., & Zheng, J. (2021). Preference elicitation for a ranking method based on multiple reference profiles [to appear]. *4OR: A Quarterly Journal of Operations Research.*
- Ouerdane, W., Dimopoulos, Y., Liapis, K., & Moraitis, P. (2011). Towards automating Decision Aiding through Argumentation. *Journal of Multi-Criteria Decision Analysis, 18*(5-6), 289–309.

Bibliography (cont.)

Tlili, A., Belahcène, K., Khaled, O., Mousseau, V., & Ouerdane, W. (2022). Learning non-compensatory sorting models using efficient sat/maxsat formulations. *European Journal of Operational Research*, *298*(3), 979–1006.