Comparing options with argument schemes powered by cancellation

Khaled Belahcene, Christophe Labreuche, Nicolas Maudet, Vincent Mousseau, Wassila Ouerdane

Comparing Options with Argument Schemes Powered by Cancellation

Khaled Belahcene[1] , **Christophe Labreuche**[2] , **Nicolas Maudet**[3*] , **Vincent Mousseau**[4] and **Wassila Ouerdane**[4]

[1] Nutriomics, Sorbonne Université, INSERM, France

[2] Thales Research and Technology, Palaiseau, France

[3]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

[4]MICS, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

khaled.belahcene@polytechnique.org, christophe.labreuche@thalesgroup.com, nicolas.maudet@lip6.fr, {vincent.mousseau, wassila.ouerdane}@centralesupelec.fr

Abstract

We introduce a way of reasoning about preferences represented as pairwise comparative statements, based on a very simple yet appealing principle: cancelling out common values across statements. We formalize and streamline this procedure with argument schemes. As a result, any conclusion drawn by means of this approach comes along with a justification. It turns out that the statements which can be inferred through this process form a proper preference relation. More precisely, it corresponds to a necessary preference relation under the assumption of additive utilities. We show the inference task can be performed in polynomial time in this setting, but that finding a minimal length explanation is NP-complete.

1 Introduction

In his famous letter to his friend Joseph Priestley, Benjamin Franklin suggested a procedure to decide upon difficult decision cases: draw two columns, list pros and cons, and delete (sets of) arguments from both sides when they are of "equal weight". It is remarkable that Franklin's "Moral Algebra" is sometimes seen as a pioneer technique to both argumentative approaches [Toulmin, 1958] which aims at formalizing, visualizing (and eventually criticizing) reasoning steps; as well as techniques to elicitate and reason about preferences based on trade-offs (*even swaps*, [Hammond *et al.*, 1998]). The bipolar nature of his algebra also proved to be influential in KR in general [Dubois *et al.*, 2008; Bouyssou *et al.*, 2009]. In this paper we build on the legacy of this approach, by relying on its core principle of cancellation, but without considering the weighting of different attributes – that is, only similar values can be crossed. We consider comparative preference statements whereby a user expresses unambiguously holistic judgments over alternatives described according to several points of view. From a set of such comparative statements, we wish to maintain the set of all valid consequences in order to make new inferences (under the form of further holistic comparative statements), and at the same time keep track of the reasoning steps involved. Thus, our objective is to know, for any preference query, *whether* it can be derived, but also *how* it can be derived.

We begin by introducing informally, through an example, a way of reasoning about preferences under the form of pairwise preference statements, then we propose a research agenda concerning this reasoning engine, and we outline the remainder the paper.

Example 1. *Hotels are compared according to the points of view of comfort, offer of a restaurant, commute time and cost. We are given* monotonicity *conditions according to each point of view (shared by all stakeholders), e.g. the larger the comfort the better; it is better to have a restaurant; the smaller the commute time the better; and the smaller the price the better. We are also given the following* preference information*:*

- π_1 : *a hotel with features* $(4*, no, 15\ min, 180\ \$)$ *is preferred to a hotel with features* $(2*, yes, 45\ min, 50\ \$)$.
- π_2 : *a hotel with features* $(4*, no, 45\ min, 50\ \$)$ *is preferred to a hotel with features* $(4*, yes, 15\ min, 100\ \$)$.

Monotonicity along each point of view allows for inferring comparative statements outside of the knowledge base.

Example 2. *(Ex. 1 continued) From* π_1*, thanks to the monotonicity w.r.t to cost, we can deduce that* $(4*, no, 15\ min, 180\ \$)$ *is preferred to* $(2*, yes, 45\ min, 80\ \$)$.

Reasoning *ceteris paribus* offers another venue to extend our knowledge about valid preferences.

Example 3. *(Ex. 2 continued) In* π_2*, both hotels share the same comfort rating* $4*$*, and we propose to interpret this statement as: comfort being equal, we prefer* $(no, 45\ min, 50\ \$)$ *to* $(yes, 15\ min, 100\ \$)$*. Taking value* $2*$*, we obtain for instance that* $(2*, no, 45\ min, 50\ \$)$ *is preferred to* $(2*, yes, 15\ min, 100\ \$)$.

These two principles are too weak to deduce many entailments from the preference information, and therefore will not allow comparing many alternatives. We therefore propose a way to combine several preference information statements. The notion of ceteris paribus reasoning can be generalized

*Contact Author

throughout statements by cancelling a similar value appearing in the left hand side (LHS) of a statement and in the right hand side (RHS) of another statement.

Example 4. *(Ex. 3 continued) For instance, the cost value 50 $ appears both in the LHS of π_2 and the RHS of π_1. This extended principle can be used to infer new statements, as illustrated in the following table. The first three lines of the table introduce the* premises*: the preference information statements π_1 and π_2, as well as the ceteris paribus monontonicity statement $d_{\P}$ according to which, everything else being equal, a hotel with a restaurant is at least as good as a hotel without one. In each column representing a feature, we strike out individual values appearing simultaneously on the LHS and the RHS—when a value is repeated, we are careful to strike out as many values from the LHS as from the RHS. At the end, we notice that there is only one value left in each column, that we report on the last line of the table—the* conclusion*—forming what we consider a valid preference statement inferred from the premises.*

As:	4*	no	~~15 min~~	180 $	π_1	2*	~~yes~~	45 min	~~50 $~~
		~~no~~	45 min	~~50 $~~	π_2		yes	~~15 min~~	100 $
		~~yes~~			$d_{\P}$		~~no~~		
So,	4*	no	45 min	180 $		2*	yes	45 min	100 $

We are now in a position to set out several research questions concerning the procedure we just informally described, that we shall address in this paper:

- *Modeling (see Section 2).* To what extent this procedure can be formalized into a reasoning model?
- *Templating (see Section 2).* Can the template for presenting the arguments supporting a claim be streamlined? How can these bundles be efficiently validated?
- *Properties (see Section 3).* The set of statements that can be inferred from a given preference information form a binary relation between alternatives. What are the properties of this relation? Importantly, is it a proper *preference relation*?
- *Inference (see Section 3).* Is there an efficient way to assess if a given pairwise preference statement can be inferred from a given preference information?
- *Explanations (see Section 4).* Given a pairwise preference statement, is it possible to find a cognitively simple certificate supporting or informing its validity?
- *Critique (see Section 5).* This reasoning engine is built upon several fundamental assumptions, that need to be discussed.

2 The Reasoning Model

We address in this section the first research question, namely: can the intuition presented in the introduction be formalized?

2.1 Features and Alternatives

We consider a set N of points of view, each one $i \in N$ expressed by a feature taken in the set $\mathbb{X}_i$. Alternatives are described as tuples of features, and belong to the Cartesian product $\mathbb{X} = \prod_{i \in N} \mathbb{X}_i$.

For an alternative $x \in \mathbb{X}$ and a point of view $i \in N$, we denote by x_i the evaluation of x according to i. For any nontrivial subset of points of view $A \subset N$ and any two alternatives $a, b \in \mathbb{X}$, we denote $a_{-A}b_A$ the (fictitious) alternative which is equivalent to a according to each point of view not in A, and equivalent to b according to the points of view in A.

2.2 Preference Information

We are interested in providing a principled way of reasoning that allows us to infer preference and answer *preference queries* of the type 'is alternative a preferred to alternative b?'. The reasoning shall be based on *preference information*, coming in two distinct flavors:

- *explicit pairwise statements* $\mathcal{P} \subset \mathbb{X}^2$, where $(x, y) \in \mathcal{P}$ means that x is at least as good as y for the decision maker;
- *implicit dominance*—we assume that each feature set corresponding to a point of view $i \in N$ is totally ordered by a relation $\succsim_i \subset \mathbb{X}_i^2$, and we denote $\mathcal{D} := \prod_{i \in N} \succsim_i$, the Pareto *dominance* relation between alternatives stemming from the ordering of each feature set, i.e. $\forall x, y \in \mathbb{X},\ x \mathcal{D} y \iff \forall i \in N,\ x_i \succsim_i y_i$.

2.3 Cancellation Axioms

The inductive principle based on ceteris paribus sketched in Ex. 3 can be formalized thanks to the concept of cancellation. The cancellative axioms are well-known in the preference literature [Krantz *et al.*, 1971; Wakker, 1989], and we briefly recall their definition.

Definition 1 (First-order cancellation)**.** *For all $A \subset N$, with $A \neq \varnothing$ and all $x, y, z, z' \in \mathbb{X}$,$x_{-A}z_A \succsim y_{-A}z_A \Rightarrow x_{-A}z'_A \succsim y_{-A}z'_A$.*

In Ex. 3, according to the first-order cancellation, π_2 implies that $(2^*, \text{no}, 45\text{ min}, 50\ \$)$ is preferred to $(2^*, \text{yes}, 15\text{ min}, 100\ \$)$.

We also have seen in the introduction cancellation *across* preference statements. It can be formalized in the following definition.

Definition 2 (High-order cancellation)**.** *Consider $m+1$ alternatives $x^{(0)}, \ldots, x^{(m)}$ in $\mathbb{X}$. Let $y^{(0)}, \ldots, y^{(m)}$ be $m+1$ alternatives in $\mathbb{X}$ such that, for every point of view $i \in N$, $(y_i^{(0)}, \ldots, y_i^{(m)})$ is a permutation of $(x_i^{(0)}, \ldots, x_i^{(m)})$. Then,$[x^{(k)} \succsim y^{(k)}, \forall k \in \{1, \ldots, m\}] \Rightarrow y^{(0)} \succsim x^{(0)}$.*

In order to conveniently represent concatenations of premises, maybe with repetition, modulo permutation, we represent the tuples of alternatives or values as *multisets*. The multiset containing the elements z_1, repeated m_1 times, $\ldots, z_k$ repeated m_k times has support $\{z_1, \ldots, z_k\}$, cardinality $\sum m_j$ and is denoted $\langle z_1 : m_1, ..., z_k : m_k \rangle$.

2.4 The Syntactic Cancellative Argument Scheme

We formalize the way of reasoning about preference statements illustrated in the introduction through an *argument scheme* [Walton, 1996], an operator tying *premises* satisfying some conditions, to a *conclusion*. This scheme is closely

related to the *high-order cancellation* axiom described previously. We slightly alter it in order to allow for a repetition of the conclusion (we defer an example and the discussion of the importance of this alteration to Section 3.5).

Definition 3 (Syntactic cancellative argument scheme)**.** *Given two positive integers $m \geq n$, and a pair of alternatives $(x, y) \in \mathbb{X} \times \mathbb{X}$, we say the multiset of pairs of alternatives $\langle (a^{(1)}, b^{(1)}) : r_1, \ldots, (a^{(k)}, b^{(k)}) : r_k \rangle \in (\mathbb{X} \times \mathbb{X})^{\mathbb{N}}$ of cardinality $m = \sum_{i=1}^{k} r_i$ is a* syntactic cancellative explanation *of length m with n repetitions of the pair (x, y) if, for each point of view $i \in N$, the multisets $\langle a_i^{(1)} : r_1, \ldots, a_i^{(k)} : r_k, y_i : n \rangle$ and $\langle b_i^{(1)} : r_1, \ldots, b_i^{(k)} : r_k, x_i : n \rangle$ are equal.*

This definition is illustrated in Ex. 4.

Validation. Checking if a given tuple of pairs of alternatives is an argument of a given pair of alternatives with a given number of repetitions can be performed in $\mathcal{O}(|N| \cdot k \ln k)$, where k is the cardinality of the support set of the explanation.

2.5 The Elliptic Cancellative Argument Scheme

In this section, we propose to streamline the syntactic cancellative argument scheme by omitting the dominance statements. As the resulting scheme is based on an omission (an *ellipsis*), we dub it the *elliptic cancellative scheme*.

Definition 4 (Elliptic cancellative explanation scheme)**.** *Given a dominance relation $\mathcal{D}$, we say the multiset of pairs of alternatives $\langle (a^{(1)}, b^{(1)}) : r_1, \ldots, (a^{(k)}, b^{(k)}) : r_k \rangle \in (\mathbb{X} \times \mathbb{X})^{\mathbb{N}}$ of cardinality $m = \sum_{i=1}^{k} r_i$ is a* syntactic cancellative explanation *of length m with n repetitions of the pair (x, y) if there exists a multiset of cardinality m' of dominance statements $\langle (c^{(1)}, d^{(1)}) : r'_1, \ldots, (c^{(k')}, d^{(k')}) : r'_{k'} \rangle \in \mathcal{D}^{\mathbb{N}}$ such that $\langle (a^{(1)}, b^{(1)}) : r_1, \ldots, (a^{(k)}, b^{(k)}) : r_k \rangle \cup \langle (c^{(1)}, d^{(1)}) : r'_1, \ldots, (c^{(k')}, d^{(k')}) : r'_{k'} \rangle$ is a syntactic cancellative explanation of length $m + m'$ with n repetitions of the pair (x, y).*

Example 5. *(Ex. 4 continued) The syntactic cancellative explanation of Ex. 4 can be simplified by removing the last statement d, yielding:*

As:	*4**	*no*	~~*15 min*~~	*180 $*	π_1	*2** ~~*yes*~~	*45 min*	~~*50 $*~~
		~~*no*~~	*45 min*	~~*50 $*~~	π_2	*yes*	~~*15 min*~~	*100 $*
So,	*4**	*no*	*45 min*	*180 $*		*2**	*yes*	*45 min*	*100 $*

Validation. It is a little more subtle to check the validity of an elliptic cancellative argument scheme than of a syntactic one. Indeed, when considering the point of view $i \in N$ and comparing the two multisets $L_i := \langle a_i^{(1)} : r_1, \ldots, a_i^{(k)} : r_k, y_i : n \rangle$ and $R_i := \langle b_i^{(1)} : r_1, \ldots, b_i^{(k)} : r_k, x_i : n \rangle$ there are missing elements corresponding to the implicit dominance relations that are not mentioned. Adding these missing dominance relations would have added "good" elements in L_i and "bad" elements in R_i - yielding two lexicographically equivalent vectors. As this is not the case, R_i contains better elements than L_i in the lexicographic sense. In Ex. 5, we obtain $L_{\text{👎}} = \langle \text{yes} : 1, \text{no} : 2 \rangle$ and $R_{\text{👎}} = \langle \text{yes} : 2, \text{no} : 1 \rangle$, so that the previous dominance is verified (as $R_{\text{👎}}$ contains more "yes" values than $L_{\text{👎}}$). Hence the validation of an elliptic cancellative argument scheme simply consists in ordering each L_i and R_i, and checking that, for every j, the j^{th} best elements in R_i is not lesser than the j^{th} best elements in L_i w.r.t. to the order relation $\succsim_i$. Thus, the validation can also be performed in $\mathcal{O}(|N| \cdot k \ln k)$.

3 The Inferred Preference Structure

In this section, we are interested in the description and the computation of the binary relation over alternatives potentially obtained by applying the reasoning engine to the facts of the preference information.

Definition 5. *Given preference information $\mathcal{P}$ and a dominance relation $\mathcal{D}$, we denote $\mathcal{N}_{\mathcal{P},\mathcal{D}}$ the set of pairs of alternatives for which there is a syntactic cancellative explanation of any length with pairs of alternatives in $\mathcal{P} \cup \mathcal{D}$.*

3.1 Inference as Closure

We note that $\mathcal{N}_\bullet$ is a closure operator: if new preference statements $\mathcal{N}_{\mathcal{P},\mathcal{D}}$ can be inferred from $\mathcal{P}$ and $\mathcal{D}$, adding them to the knowledge base would not yield additional inference.

Lemma 1.

$$\mathcal{N}_{\mathcal{P},\mathcal{D}} = \mathcal{N}_{\mathcal{N}_{\mathcal{P},\mathcal{D}},\mathcal{D}}$$

Sketch of proof. The inclusion $\mathcal{N}_{\mathcal{P},\mathcal{D}} \subset \mathcal{N}_{\mathcal{N}_{\mathcal{P},\mathcal{D}},\mathcal{D}}$ is a consequence of the fact that $\langle s : 1 \rangle$ is an explanation of s for any statement in $\mathcal{P}$. As for $\mathcal{N}_{\mathcal{P},\mathcal{D}} \supset \mathcal{N}_{\mathcal{N}_{\mathcal{P},\mathcal{D}},\mathcal{D}}$, let $(X, Y) \in \mathcal{N}_{\mathcal{N}_{\mathcal{P},\mathcal{D}},\mathcal{D}}$. There is an syntactic explanation of length m with n repetitions of the pair (X, Y), say $\langle (x^{(1)}, y^{(1)}) : r_1, \ldots, (x^{(K)}, y^{(K)}) : r_K \rangle \in (\mathbb{X} \times \mathbb{X})^{\mathbb{N}}$ where each pair $(x^{(k)}, y^{(k)})$ is in $\mathcal{N}_{\mathcal{P},\mathcal{D}}$, and is therefore supported by an explanation E_k of length m_k with n_k repetitions, with statements in $\mathcal{P} \cup \mathcal{D}$. We claim the tuple obtained by concatenating each explanation E_k repeated $\prod_{k' \in [m], k' \neq k} n_{k'}$ is an explanation with $\prod_{k \in [m]} n_k$ repetitions of the pair (X, Y), with statements in $\mathcal{P} \cup \mathcal{D}$. □

3.2 A Detour via Model-Based Inference

In order to state the main result of this paper, we need to recall the basic principles of *model-based inference*. The goal of inference is extend some (limited) preference information to a richer preference relation $\mathcal{R}$, with 'good' properties, such as $\mathcal{R}$ being a reflexive and, transitive binary relation over $\mathbb{X}$, and maybe complete.

When preference information is given as $\mathcal{P} \cup \mathcal{D}$, where $\mathcal{P}$ is the explicit part, given in so-called *holistic* form, i.e. $\mathcal{P} \subset \mathbb{X}^2$ is a set of reference pairwise statements, and $\mathcal{D}$ is the dominance relation stemming from the ordering of the features, the relation $\mathcal{R}$ is said to be *consistent* when $\mathcal{P} \cup \mathcal{D} \subset \mathcal{R}$.

In order to describe $\mathcal{R}$, which is potentially a very complicated combinatorial object, in a simple language, it is customary to rely on some kind of parameterization of the target set. For instance, numeric models [Jacquet-Lagrèze and Siskos, 1982] in the field of multiple criteria decision making, or graphical languages [Wilson, 2009; Amor *et al.*, 2016] from KR. A popular paradigm consists in considering *value-based*

preferences, where the target preference relation is parameterized by a numeric scoring function $u : \mathbb{X} \to \mathbb{R}$, so that $x\mathcal{R}_u y \iff u(x) \geq u(y)$. (this assumption is made without loss of generality as soon as $\mathcal{R}$ is assumed to be transitive and complete). The target set is still very large and complex, and a common additional assumption is to restrict the scoring function to be *additive* w.r.t. the features, i.e. there is a decomposition such that $\forall x \in \mathbb{X}, u(x) = \sum_{i \in N} \omega_i(x_i)$.

Definition 6 (preferences based on additive value)**.** *The parameter set of the additive value model is* $\Omega_\Sigma := \prod_{i \in N} \mathbb{R}^{\mathbb{X}_i}$*, and for a given value* $\omega := \langle \omega_i \rangle_{i \in N}$ *of the parameter, the corresponding preference relation* $\mathcal{R}_{\sum \omega} \subset \mathbb{X}^2$ *is defined by:*

$$\forall x, y \in \mathbb{X}, x\mathcal{R}_{\sum \omega} y \iff \sum_{i \in N} \omega_i(x_i) \geq \sum_{i \in N} \omega_i(y_i).$$

For such an additive value, that we denote $\mathcal{R}_{\sum \omega}$, the condition $\mathcal{D} \subset \mathcal{R}_{\sum \omega}$ translates to the following monotonicity conditions: for all features i, the function $\omega_i : (\mathbb{X}_i, \succsim_i) \to (\mathbb{R}, \geq)$ is nondecreasing.

The most prevalent approach in model-based inference consists in determining the most adequate value of the parameter in the sense of some loss function $\mathcal{L}$: $\omega^\star = \text{argmin}_{\Omega_\Sigma} \mathcal{L}$, and returning the corresponding preference relation $\mathcal{R}_{\sum \omega^\star}$. Meanwhile, the *robust* approach consists in considering the intersection of all the consistent preference relations, assuming it is not empty:

$$\mathcal{R}^\star_{\Omega_\Sigma} := \bigcap_{\omega \in \Omega_\Sigma : (\mathcal{P} \cup \mathcal{D}) \subset \mathcal{R}_{\sum \omega}} \mathcal{R}_{\sum \omega}.$$

The robust approach yields the *version space* [Mitchell, 1982] of the model. Equivalently, it can be understood as assuming that the preference information $\mathcal{P}$ is *incomplete* (as there might be several value of the parameter that are consistent with it), and drawing skeptical conclusions with respect to all the possible completions.

3.3 Cancellative-Powered Deductions are Robust Inferences under Additive Values

We are now able to state an important result concerning the preference structure $\mathcal{N}_{\mathcal{P},\mathcal{D}}$.

Theorem 1.

$$\mathcal{N}_{\mathcal{P},\mathcal{D}} = \mathcal{R}^\star_{\Omega_\Sigma}$$

The inferred preference structure is exactly the necessary preference relation under the assumption of an additive value model. This result has an important corollary concerning the inferred relation:

Corollary 1 (Properties of the inferred structure)**.** *$\mathcal{N}_{\mathcal{P},\mathcal{D}}$ is a transitive and reflexive binary relation.*

The proof of Th. 1 relies on the fact that, under the assumption of an additive value model, a preference statement can be represented by a linear form operating over the vector space Ω_Σ.

Definition 7. *Given some preference information* $\mathcal{P} \subset \mathbb{X}^2$*, alternatives* $x, y \in \mathbb{X}$*, and a point of view* $i \in N$*, for any value* $x_i \in \mathbb{X}_i$*, let* $\epsilon_{i,x_i} : \mathbb{R}^{\mathbb{X}_i} \to \mathbb{R}$, $\omega_i \mapsto \omega_i(x_i)$*, and* $\phi_{(x,y)} = \sum_{i \in N} \epsilon_{i,x_i} - \epsilon_{i,y_i}$ *a linear form over* $\mathbb{R}^{\mathbb{X}}$*. Also, let* $\widehat{\mathbb{X}}_i := \{t \in \mathbb{X}_i : \exists (a,b) \in \mathcal{P}, t = a_i \text{ or } t = b_i\} \cup \{x_i\} \cup \{y_i\}$ *and* $\widehat{X} := \prod_{i \in N} \widehat{\mathbb{X}}_i$*.*

Lemma 2.

$$(x, y) \in \mathcal{R}_{\sum \omega} \iff \phi_{(x,y)}(\omega) \geq 0$$

Proof of $\mathcal{N}_{\mathcal{P},\mathcal{D}} \subset \mathcal{R}^\star_{\Omega_\Sigma}$.
Let $(x, y) \in \mathcal{N}_{\mathcal{P},\mathcal{D}}$. By definition, there is a syntactic cancellative explanation of length m with n repetitions of the pair (x, y), say $\langle (a^{(1)}, b^{(1)}), \ldots, (a^{(m)}, b^{(m)}) \rangle \in (\mathcal{P} \cup \mathcal{D})^m$. Therefore, for each point of view $i \in N$, $(y_i, \ldots, y_i, a_i^{(1)}, \ldots, a_i^{(m)})$ is a permutation of $(x_i, \ldots, x_i, b_i^{(1)}, \ldots, b_i^{(m)})$. In particular, for any parameter $\omega \in \Omega_\Sigma$, $n\omega_i(y_i) + \omega_i(a_i^{(1)}) + \cdots + \omega_i(a_i^{(1)}) = n\omega_i(x_i) + \omega_i(b_i^{(1)}) + \cdots + \omega_i(b_i^{(1)})$, so $n\phi_{(x,y)}(\omega) = \sum_{j=1}^m \phi_{(a^{(j)},b^{(j)})}(\omega)$. Now, if ω is consistent, $(\mathcal{P} \cup \mathcal{D}) \subset \mathcal{R}_\omega$ and $\phi_{(a^{(j)},b^{(j)})}(\omega) \geq 0$. Thus $n\phi_{(x,y)}(\omega)$ is nonegative as the sum of m nonnegative terms, and x is necessarily preferred to y under the assumption of an additive value model. □

Proof of $\mathcal{N}_{\mathcal{P},\mathcal{D}} \supset \mathcal{R}^\star_{\Omega_\Sigma}$.
Let $(x, y) \in \mathcal{R}^\star_{\Omega_\Sigma}$. For any parameter $\omega \in \Omega_\Sigma$ such that $\forall s \in (\mathcal{P} \cup \mathcal{D}), \phi_s(\omega) \geq 0$, $\phi_{(x,y)}(\omega) \geq 0$. This property concerns linear forms in $\mathbb{R}^{\mathbb{X}}$, which is a vector space of infinite dimension, but also holds in $\mathbb{R}^{\widehat{\mathbb{X}}}$.Indeed, $\widehat{\mathbb{X}} \subset \mathbb{X}$, is of finite dimension, and any additive parameter function $\widehat{\omega} : \widehat{\mathbb{X}} \to \mathbb{R}$ such that $(\mathcal{P} \cup \mathcal{D}) \cap \widehat{\mathbb{X}}^2 \subset \mathcal{R}_{\widehat{\omega}}$ can be extended into an additive function $\omega : \mathbb{X} \to \mathbb{R}$ describing a consistent relation $\mathcal{R}_{\sum \omega}$. By Farkas' lemma, the linear form $\phi_{(x,y)}$ is a conical combination of the $\langle \phi_s \rangle_{s \in (\mathcal{P} \cup \mathcal{D}) \cap \widehat{\mathbb{X}}^2}$. As the coefficients of all these linear forms are integers—they are, indeed, in $\{-1, 0, 1\}$—the coefficients of the conical combinations can be chosen rational, and by multiplying by the lesser common multiple of their denominators, yield an identity: $n\ \phi_{(x,y)} = \sum_{s \in (\mathcal{P} \cup \mathcal{D}) \cap \widehat{\mathbb{X}}^2} m_s\ \phi_s$, with a positive integer n and nonnegative integer coefficients $\langle m_s \rangle_{s \in (\mathcal{P} \cup \mathcal{D}) \cap \widehat{\mathbb{X}}^2}$. We claim the tuple $\langle s : m_s \rangle_{s \in (\mathcal{P} \cup \mathcal{D}) \cap \widehat{\mathbb{X}}^2}$ is a syntactic cancellative explanation of length n of the pair (x, y), with statements in $\mathcal{P} \cup \mathcal{D}$, thus $(x, y) \in \mathcal{N}_{\mathcal{P} \cup \mathcal{D}}$. □

3.4 Efficient Inference Procedures

The necessary relation assuming an additive value model $\mathcal{R}^\star_{\Omega_\Sigma}$ is defined and studied by [Greco *et al.*, 2008]. In particular, Greco et al. propose a linear program permitting to solve the decision problem corresponding to our research question concerning inference: given a pair of alternative, decide if it is in the inferred preference relation. This linear program is expressed in the primal space $\mathbb{R}^{\widehat{X}}$ of the values $\omega_i(x_i)$ given to each relevant value of the attributes. These values are of little interest concerning our cancellative argument schemes, so we propose to formulate the dual problem.

Corollary 2 (Polytime inference via conical decomposition)**.** *For all pairs of alternatives* $(x, y) \in \mathbb{X}^2$*,* $(x, y) \in \mathcal{N}_{\mathcal{P},\mathcal{D}}$ *if,*

and only if, the following linear program is feasible:
find nonnegative real numbers $\langle \lambda_s \rangle_{s \in (\mathcal{P} \cup \mathcal{D}) \cap \widehat{X}^2}$ *such that* $\phi_{(x,y)} = \sum_{s \in (\mathcal{P} \cup \mathcal{D}) \cap \widehat{X}^2} \lambda_s \phi_s$.

Would the decision variables λ be integers, they could directly be interpreted as a multiset $\langle s : \lambda_s \rangle$ serving as a syntactic cancellative explanation for (x, y). As the elliptic scheme tells us that the coefficients corresponding to the dominance statements are eventually irrelevant, this formulation ought to be further streamlined. Unfortunately, the conical span of the dominance statements is not easy to characterize in the dual base (ϵ_{i,x_i}). This obstacle can be lifted by representing the preference statements in an alternative decomposition, that focuses on differences of values, rather than values.

Definition 8. *Given a finite binary relation* $A \subset \mathbb{X}^2$*, for all points of view* $i \in N$ *we denote* $\{\widehat{x}_{i,1} \precsim \widehat{x}_{i,2} \precsim \cdots \precsim \widehat{x}_{i,|\widehat{\mathbb{X}}_i|}\} = \widehat{\mathbb{X}}_i$ *. For any integer* $k, 1 \leq k < |\widehat{\mathbb{X}}_i|$*, let* $\delta_{\widehat{\mathbb{X}},i,k} := \epsilon_{i,\widehat{x}_{i,k+1}} - \epsilon_{i,\widehat{x}_{i,k}}$.

Lemma 3. *For any statement* $(x, y) \in A$ *and any point of view* $i \in N$*,*

$$\epsilon_{i,x_i} - \epsilon_{i,y_i} = \begin{cases} 0 & \text{if } x_i \sim_i y_i \\ \sum_{k : x_i \leq \widehat{x}_{i,k} < y_i} (-1) \cdot \delta_{\widehat{\mathbb{X}},i,k}, & \text{if } x_i \prec_i y_i \\ \sum_{k : y_i \leq \widehat{x}_{i,k} < x_i} (+1) \cdot \delta_{\widehat{\mathbb{X}},i,k}, & \text{if } x_i \succ_i y_i \end{cases}$$

This lemma has two important consequences:

i) Corollary 2 can be expressed in terms of $\langle \delta_{\widehat{X}^2} \rangle$ rather than $\langle \epsilon \rangle$; and

ii) dominance statements in A are exactly the conical span of the $\langle \delta_A \rangle$.

This leads to a leaner reformulation of the inference problem.

Definition 9. *For all* $x, y \in \mathbb{X}$*,* $i \in N$ *and* $k \in \mathbb{N} : \ 1 \leq k < |\widehat{\mathbb{X}}_i|$*, let*

$$\varphi^{(i,k)}_{(x,y)} := \begin{cases} -1, & \text{if } x_i \precsim_i \widehat{x}_{i,k} \prec_i y_i; \\ 0, & \text{if } x_i \sim_i y_i; \text{ or} \\ +1, & \text{if } y_i \precsim_i \widehat{x}_{i,k} \prec_i x_i. \end{cases}$$

Theorem 2 (Inference via LP)**.** *For all pairs of alternatives* $(x, y) \in \mathbb{X}^2$*,* $(x, y) \in \mathcal{N}_{\mathcal{P},\mathcal{D}}$ *if, and only if, the following linear program is feasible:*
find nonnegative real numbers $\langle \lambda_s \rangle_{s \in \mathcal{P}}$ *such that the inequality* $\varphi^{(i,k)}_{(x,y)} \geq \sum_{s \in \mathcal{P}} \lambda_s \varphi^{(i,k)}_s$ *holds for every indices* $i \in N$ *and* $1 \leq k < |\widehat{\mathbb{X}}_i|$.

3.5 Repetition of the Conclusion

The presence of repetition of the conclusion makes the explanation scheme cumbersome. One may wonder whether it is possible to get rid of the repetition of the conclusion.

Theorem 3. *It is not possible to cover all possible inferences obtained by the robust additive model by restricting the cancellation explanation schema with* $n = 1$ *repetition.*

Proof. We provide a counter-example with $|N| = 6$ features, $\mathbb{X} = \{0, 1\}^6$ and $1 \succsim_i 0$ (statement d_i) for all $i \in N$. The preference information is:

$$\pi_1 : \big((0, 0, 1, \cdot, \cdot, \cdot), (1, 1, 0, \cdot, \cdot, \cdot) \big)$$
$$\pi_2 : \big((0, \cdot, \cdot, 0, 1, \cdot), (1, \cdot, \cdot, 1, 0, \cdot) \big)$$
$$\pi_3 : \big((\cdot, 0, \cdot, \cdot, 1, 0), (\cdot, 1, \cdot, \cdot, 0, 1) \big)$$
$$\pi_4 : \big((\cdot, \cdot, 1, 0, \cdot, 0), (\cdot, \cdot, 0, 1, \cdot, 1) \big)$$

We can infer from the preference information that $(0, 0, 1, 0, 1, 0)$ is preferred to $(1, 1, 0, 1, 0, 1)$ (statement π_{C}). One can readily see that $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 2\, \pi_{\mathrm{C}}$.

Assuming by contradiction that there exist Farkas coefficients with coefficient 1 associated to π_{C}: $\pi_{\mathrm{C}} = \sum_{i=1}^{6} \lambda_i \pi_i + \sum_{i=1}^{6} \mu_i \, d_i$, where $\lambda_i, \mu_i \in \mathbb{N}$ leads to an infeasible linear system.

□

One could also want to trim down the potential complexity of the explanations by limiting the number of premises. Unfortunately, this might lead to loss of transitivity for the inferred relation.

4 Explanations for Valid Preference Statements

It seems reasonable to believe that an explanation is easier to process by a cognitive agent—'simpler'—when it is short. In the case of cancellative explanations, the actual cognitive burden mainly comes from three factors: the number of points of view $|N|$, that we consider as mostly exogenous; the length m of the premises; and the number n of repetitions of the conclusion. Without any experimental evidence, we consider the problem of finding an explanation for a given pair $(x, y) \in \mathcal{N}_{\mathcal{P}}$ which is as simple as possible as a bi-objective integer linear minimization problem:

$$\min_{n,m \in \mathbb{N}^\star} (n, m) \quad \text{such that} \quad \begin{cases} n \, \varphi_{(x,y)} \geq \displaystyle\sum_{\pi \in \mathcal{P}} \ell_\pi \varphi_\pi; & \text{and} \\ m \geq \displaystyle\sum_{\pi \in \mathcal{P}} \ell_\pi. \end{cases} \tag{1}$$

Integer linear programs offer a powerful language permitting to describe difficult combinatorial problems. These formulations can be given wholesale to dedicated solvers, that eschews the need for developing a dedicated piece of software and benefits from state-of-the-art refinements in the solving of such problems. Nevertheless, it would be unwise to delegate the search for a short explanation of a given pair of alternatives to such a solver, if this search were not, intrinsically, a difficult combinatorial problem. The following theorem addresses this issue.

Theorem 4. *The problem of deciding, for a given input* $(x, y, n, m) \in \mathbb{X} \times \mathbb{X} \times \mathbb{N}^\star \times \mathbb{N}^\star$ *if there is an elliptic cancellative explanation of the pair* (x, y) *of length at most* m *with at most* n *repetitions is NP-complete. This remains true even if the number* n *of repetitions is set to one.*

Proof. Membership to NP is ensured, as checking the validity of an elliptic scheme is polyomial in the number of distinct premises, which is upper bounded by the cardinality of $\mathcal{P}$.

Hardness can be established e.g. by reduction from VERTEX COVER [Karp, 1972]. Formally, a vertex cover V' of an undirected graph $G = (V, E)$ is a subset of V such that $uv \in E \Rightarrow u \in V' \vee v \in V'$, that is to say it is a set of vertices V' where every edge has at least one endpoint in the vertex cover V'. The VERTEX COVER problem consists in, given an instance (G, k) where $G = (V, E)$ is a graph and k a positive integer, to *decide* whether G has a vertex cover of size at most k, or not. Given an instance of VERTEX COVER, we map it to a gadget instance of our problem:

- the set of points of view is $N = V \cup E$;
- an alternative is a subset of N;
- each point of view is evaluated on a binary scale, with presence preferred to absence;
- the preference information contains all statements of the form $(\{(u, v)\}, \{u, v\})$—any edge is preferred to the set of its endpoints—for all edges $(u, v) \in E$.

Any elliptic cancellative explanation without repetition of the pair (E, V)—the pros are the edges, the cons are the vertices—of length k is a subset of E that forms a vertex cover of size k of of the graph G, and reciprocally.

□

5 Discussion and Perspectives

In the current quest for "explainable A.I.", the additive value (i.e. linear) model might be seen as occupying the very end of the spectrum –an obviously interpretable model [Ribeiro *et al.*, 2016]. Even though recent advances have been made towards providing "simpler" models, e.g. [Ustun and Rudin, 2016], most of these approaches ignore the perspective of the decision maker [Miller, 2019], and the need to provide her with a way of challenging the decision [Kroll *et al.*, 2017].

Several works have explored the interplay between argumentation and decision aiding. In [Amgoud and Prade, 2009], argumentation is used as a mean make a decision and justify it, while in [Zhong *et al.*, 2019], it is shown that the outcomes of a simple decision model are similar to the extension of the corresponding argumentation framework.

Here, the preference information is considered exogenous. It might have been obtained through dialog, by considering domain knowledge—reference cases, jurisprudence, or inferred by some means—learning from similar situations, or previous interactions with the user.

5.1 Contributions

We introduced the notion of *cancellative explanations*, based on the accrual of premises to obtain a conclusion. We studied this explanative framework in the light of the principles stated in introduction. This contrasts with approaches in decision theory [Fishburn, 1970; Gonzales, 2000], where cancellation is seen as a property of the preference relation, not a mean to infer new preference statements and justify them. Our main contributions are as follows:

Completeness. Every preference statement that can be skeptically inferred from the preference information and the way of reasoning corresponding to the additive value model is supported by a cancellative explanation.

Soundness. Every preference statement that is supported by a cancellative explanation can be skeptically inferred from the preference information and the way of reasoning corresponding to the additive value model;

Simplicity. We provided several ways of presenting cancellative explanations, in the form of tables, diagrams, or argument schemes, and proposed to ground them on a syntactic check, or alternatively to keep implicit the information tied to dominance, which can easily be restored by the recipient, in the spirit of *enthymemes*. We provided formalizations that lend themselves to an efficient implementation. We proposed an intuitive partial ordering of explanations according to their alleged complexity, and formulated the problem of finding explanations as simple as possible.

Computation. Remarkably, while adjudicating necessary preference is a polynomial problem, explaining it concisely is NP-complete.

5.2 Perspectives

Providing an argument scheme along with the result of a comparative statement opens the possibility to discuss or challenge this result. This is made possible through what is called critical questions [Walton, 1996], a tool associated with argument schemes representing attacks or criticisms that, if not answered adequately, falsify the argument fitting the scheme.

In our setting, the criticism may point out (implicitly or explicitly) elements perceived as missing or wrong in the reasoning steps. Indeed, for instance, the decision maker (DM) may challenge the fact that a preference between two alternatives is not the right one. The consequence is that either it is possible to derive a new conclusion with this new information, or the DM's statements express conflicting preferences. Thus, the challenge of finding a principled way to deal with inconsistency in an accountable manner, needs to be addressed. Several promising approaches have been proposed: considering maximally consistent subsets of statements [Mousseau *et al.*, 2003]; relaxing the aggregation model until a model sufficiently expressive to accommodate for the preference information is found [Ouerdane, 2011; Greco *et al.*, 2014]; or using a numerical estimation of inconsistency such as a belief function [Destercke, 2018].

Another situation is that the DM's reasoning is incompatible with the principles and properties underlying the preference model. For instance, expressing a preference dependency may defeat the fundamental feature (ceteris paribus) of an additive model [Fisher, 1892]. In this situation, relaxing the preference model could be a solution [Ouerdane, 2011]. Many models account for interactions between the influence of the points of view, such as Generalized additive models (GAI) [Fishburn, 1967]. An underlying question that has been less investigated (for notable exceptions, see e.g. [Labreuche, 2011] and [Cailloux and Endriss, 2014]), and remains difficult [Procaccia, 2019], is the question of the accountability of recommendations based on an induced model.

Acknowledgements

This work is partially supported by the ANR project 14-CE24-0007-01- CoCoRICo-CoDec.

References

[Amgoud and Prade, 2009] Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173:413–436, 2009.

[Amor *et al.*, 2016] Nahla Ben Amor, Didier Dubois, Hela Gouider, and Henri Prade. Graphical models for preference representation: An overview. In *Proceedings of the 10th International Conference on SUM*, pages 96–111, 2016.

[Bouyssou *et al.*, 2009] Denis Bouyssou, Didier Dubois, Marc Pirlot, and Henri Prade. *Decision Making Process: Concepts and Methods*. Wiley ISTE, 2009.

[Cailloux and Endriss, 2014] Olivier Cailloux and Ulle Endriss. Eliciting a suitable voting rule via examples. In *Proceedings of the 21st ECAI'14* , pages 183–188, 2014.

[Destercke, 2018] Sébastien Destercke. A generic framework to include belief functions in preference handling and multi-criteria decision. *International Journal of Approximate Reasoning*, 98:62 – 77, 2018.

[Dubois *et al.*, 2008] Didier Dubois, Hélène Fargier, and Jean-Fraccois Bonnefon. On the qualitative comparison of decisions having positive and negative features. *J. Artif. Intell. Res.*, 32:385–417, 2008.

[Fishburn, 1967] Peter C. Fishburn. Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review*, 8(3):335–342, 1967.

[Fishburn, 1970] Peter C. Fishburn. *Utility Theory for Decision Making*. J. Wiley & Sons, 1970.

[Fisher, 1892] Irving Fisher. *Mathematical investigations in the theory of value and prices, and appreciation and interest*. 1892.

[Gonzales, 2000] Christophe Gonzales. Two factor conjoint measurement with one solvable component. *Journal of Mathematical Psychology*, 44(2):285–309, 2000.

[Greco *et al.*, 2008] Salvatore Greco, Vincent Mousseau, and Roman Słowiński. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *EJOR*, 191(2):416–436, 2008.

[Greco *et al.*, 2014] Salvatore Greco, Vincent Mousseau, and Roman Słowiński. Robust ordinal regression for value functions handling interacting criteria. *EJOR*, 239(3):711–730, 2014.

[Hammond *et al.*, 1998] John Hammond, Ralph Keeney, and Howard Raiffa. Even Swaps: a rational method for making trade-offs. *Harvard Business Review*, pages 137–149, 1998.

[Jacquet-Lagrèze and Siskos, 1982] Eric Jacquet-Lagrèze and Yanis Siskos. Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *EJOR*, 10:151–164, 1982.

[Karp, 1972] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations: Proceedings of a symposium on the Complexity of Computer Computations*, pages 85–103. 1972.

[Krantz *et al.*, 1971] David H. Krantz, Duncan R. Luce, Patrick Suppes, and Amos Tversky. *Foundations of measurement*, volume 1: Additive and Polynomial Representations. Academic Press, 1971.

[Kroll *et al.*, 2017] Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. Accountable algorithms. *University of Pennsylvania Law Review*, 165, 2017.

[Labreuche, 2011] Christophe Labreuche. A general framework for explaining the results of a multi-attribute preference model. *AIJ*, 175:1410–1448, 2011.

[Miller, 2019] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, 267:1–38, 2019.

[Mitchell, 1982] Tom M. Mitchell. Generalization as search. *Artificial Intelligence*, 18(2):203–226, 1982.

[Mousseau *et al.*, 2003] Vincent Mousseau, Luis C. Dias, José Figueira, Carlos Gomes, and João N. Clímaco. Resolving inconsistencies among constraints on the parameters of an MCDA model. *EJOR*, 147(1):72–93, 2003.

[Ouerdane, 2011] Wassila Ouerdane. Multiple criteria decision aiding: a dialectical perspective. *4OR*, 9:429–432, 2011.

[Procaccia, 2019] Ariel D. Procaccia. Axioms should explain solutions. In Laslier, Moulin, Sanver, and Zwicker, editors, *Future of Economic Design*. 2019.

[Ribeiro *et al.*, 2016] Marco T. Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD*, pages 1135–1144, 2016.

[Toulmin, 1958] Stephen Toulmin. *The uses of Arguments*. Cambridge University Press, 1958.

[Ustun and Rudin, 2016] Berk Ustun and Cynthia Rudin. Supersparse linear integer models for optimized medical scoring systems. *Mach. Learn.*, 102(3):349–391, 2016.

[Wakker, 1989] Peter Wakker. *Additive Representations of Preferences: A New Foundation of Decision Analysis*. Theory and Decision Library C. 1989.

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

[Wilson, 2009] Nic Wilson. Efficient inference for expressive comparative preference languages. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 961–966, 2009.

[Zhong *et al.*, 2019] Qiaoting Zhong, Xiuyi Fan, Xudong Luo, and Francesca Toni. An explainable multi-attribute decision model based on argumentation. *Expert Systems with Applications*, 117:42–61, 2019.