An Ontological Modelling of Reason-Based Preferences

Daniele Porello[0000−0003−3655−0218]

University of Genoa, Via Balbi 4, 1626, Genoa, Italy
daniele.porello@unige.it
http://www.springer.com/gp/computer-science/lncs

Abstract. We present an ontological framework for the reason-based model of individual preferences introduced by F. Dietrich and C. List. According to this perspective, an agent prefers x to y if and only if the importance of the reasons motivating x outweighs the importance of the reasons motivating y. Firstly, we represent motivating reasons as concepts in Description Logic, to enable a rich ontological theory that provides a clear and shareable semantics of reasons. Secondly, we present a model to express preferences on combinations of reasons. Finally, we discuss how preferences on alternatives depend on preferences on motivating reasons. We present the framework in a knowledge-dependent way, meaning that the ontological background constrains the definable preferences on alternatives and reasons.

Keywords: property-based preferences · motivating reasons · knowledge-dependence · ontologies

1 Introduction

The theory of preference relations plays a fundamental role in various branches of economics, including game theory, decision theory, and social choice theory. Preferences are also a fundamental concept in computer science, particularly in multiagent systems and Artificial Intelligence. They contribute to defining the very notion of an agent —an entity typically endowed with beliefs, goals, desires, intentions, and preferences— and help understand and model scenarios such as coordination, negotiation, and competition among agents, including artificial ones.

The foundation of the theory of preference relations can be traced back to the work of Kenneth Arrow [2], where preferences are formalised as binary relations satisfying a number of desiderata, and to Amartya Sen [15], where preferences are inferred from observable choices. The debate on the nature of preferences and the question of 'where do preferences come from?' is, of course, too broad to be summarised here. We directly introduce a recent important contribution to the theory of preferences, that is at the origin of the proposed approach: the work of Dietrich and List, which construes preferences as based on motivating reasons, see [7] and [6]. A related logic-based view models preferences as grounded on agents' beliefs about the alternatives or on the properties of options, e.g. [9], [12].

The reason-based approach is interesting because it is quite general, it clearly (axiomatically) defines what it means for a preference to be reason-based, and it provides necessary and sufficient conditions that characterise preferences based on motivating reasons.

Therefore, this theory has interesting applications to the foundations of the problems *rationalisability* and *explicability* of individual preferences.

The modelling of reasons and preferences of Dietrich and List is semantical: preference are relations on a set of alternatives X, while reasons are properties of alternatives, that is, subsets of X. In fact, no formal language is explicitly introduced to model reasons and preferences. Moreover, the possible logical connections or conflicts between reasons —or, more generally, the meanings of reasons— remain somehow implicit in the applications of the approach.

The objective of this work is to develop the reason-based approach to preferences by leveraging the techniques of Knowledge Representation and Applied Ontology. In particular, we shall define a formal language to express reasons based on Description Logics ($\mathscr{DL}s$). To provide the information which is required to give contents to reasons, we shall introduce a methodology to place the reason-based approach within a rich ontological setting.

The semantic approach of Dietrich and List can be construed as if it were based on a single intended model of a theory. This option is often too demanding, as it forces the crisp agreement on any piece of information. We shall base our restatement of [7] on the models enabled by a formal theory. That is, the definitions that we shall develop (e.g. the satisfaction of a motivating reason) are to be intended as *knowledege-dependent*, in the sense explored in [8]. We also investigate how to express preferences on possible combinations of reasons. As we shall see, this task is related to the problem of defining preferences in combinatorial domains. To this task, we will adapt the setting in [16, 17] to the case of concepts in $\mathscr{DL}s$. Then, we shall model how the preferences defined on combinations of reasons affect the preferences on the alternatives that satisfy those reasons.

A delicate aspect that we encounter when defining preference on combination of reasons is the possible logical dependency or conflict between reasons, which constraint the definable preferences.

Two articles are specifically related to the proposed approach, they are in the area of multi-attribute decision making. In [14], weighted concepts in $\mathscr{DL}$ are used to define preferences on formulas. The main difference with the present approach is that we shall work with the class of models of a Knowledge Base, rather than with single models, besides, we are specifically interested in expanding the reason-based approach in an ontological setting. Secondly, in [1] weighted $\mathscr{DL}s$ are used to model preferences and utility functions on options depending on the attributes that they satisfy. They do work with models of a Knowledge Base, the main difference with this works is that we are interested specifically in the reason-based approach of [7], and in discussing the compatibility between preferences and a Knowledge Base.

The remainder of this paper is organised as follows. In Section 2, we review the reason-based approach to preference proposed by Dietrich and List, [7]. In Section 3, we present our modelling of reason-based preferences in $\mathscr{DL}s$, specifically in $\mathscr{ALC}$. Then, we introduce a method to define utility functions and preferences on combinations of reasons. In Section 4, we approach the dependence of preference on motivating reasons and we discuss which preferences can be construed as reason-based, consistently with a Knowledge Base. Section 5 suggests applications, indicates future work, and concludes.

2 A model of reason-based preferences

In [7], the reason-based approach to preferences is presented "semantically", i.e. in terms of sets and relations. We rephrase it by means of a formal language. We refer here to a general predicative language $\mathscr{L}$ (e.g. a fragment of first-order logic), while in the next sections we will instantiate $\mathscr{L}$ with the language of Description Logics, i.e. $\mathscr{ALC}$.

Let X be a finite set of alternatives. As usual, a preference relation on X is a binary relation $\succeq \subseteq X \times X$ that is reflexive, transitive, and complete. The indifference relation is defined by $x \sim y$ iff $x \succeq y$ and $y \succeq x$.

A motivating reason is viewed by Dietrich and List as a property of the alternatives; hence, it is intended semantically as a subset of X. Here, we model properties as unary predicates and the preference relation as a binary predicate of the alphabet of $\mathscr{L}$. The preference relation is then supposed to satisfy the axioms corresponding to reflexivity, transitivity, and completeness (cf. [13]) to individuate the intended models. Abusing the notation, we stil use $\succeq$ for the relational symbol of the language.

We assume a (finite) designated set of unary predicates $\mathscr{R} = \{P_1, \dots, P_m\}$ for representing reasons. We decide a model (X, I) to interpret the predicates, by selecting the set of alternatives X as domain. Thus, the predicates of $\mathscr{R}$ are interpreted as subsets of X: $I(P_i) \subseteq X$. A subset $M \subseteq \mathscr{R}$ represents a set of *motivating reasons* of a certain agent, cf. [7].

Definition 1 (Coherence of reasons). *A set M of reasons is* coherent *iff* $\bigcap_{P_i \in M} I(P_i) \neq \emptyset$ *(or, equivalently, iff* $P_1 x \wedge \dots \wedge P_n x$ *is satisfiable in* X, I*).*[1]

We denote by $\mathscr{M} \subseteq 2^{\mathscr{R}}$ the set of all sets of motivating reasons. In [7], $\mathscr{M}$ is intended to abstractly represent the sets of motivating reasons of an agent in various circumstances. By setting $\mathscr{M} \subseteq 2^{\mathscr{R}}$, we may exclude implausible combinations of reasons, e.g. incoherent sets of reasons, if agents exhibit a modicum of rationality.

We assume a set of individual constants to refer to alternative in X. By slightly abusing the notation, we shall use the same letters for denoting the alternatives in X and the individual constants, we also endorse the unique-name assumption. We write $I \models Pa$ to state that the formula Pa is true in the model X, I.

We mildly restate the definition of property-based preference relations, cf. [7], in this setting.

Definition 2 (Property-based preference relations). *A family of preference relations* $(\succeq_M)_{M \in \mathscr{M}}$ *is* property-based *iff there exists a relation* $\succeq$ *defined on the set of (coherent*

[1] In [7], this property is termed consistency. We term it coherence, although it refers here to a single model, the one where P_is are interpreted. By contrast, the notion of coherence of a concept in $\mathscr{DL}s$ refers to the existence of a model where the concept is instantiated.

sets of) reasons ($\geq \subseteq \mathscr{M} \times \mathscr{M}$) such that, for every a, b in X and for any motivating set of reasons $M \in \mathscr{C}$, the following equivalence holds:

$$a \succeq_M b \text{ iff } \{P \in M \; s.t. \; I \models Pa\} \succeq \{P \in M \; s.t. \; I \models Pb\} \tag{1}$$

That is, a is preferred to b according to the motivating reasons M iff the reasons in M that a satisfies are "better" than the reasons in M that b satisfies (according to the relation $\succeq$). The relation $\succeq$ is termed a *weighing* relation, it compares the importance of every pair of sets of motivating reasons.[2]

In [7], property-based preferences are constrained by the following two axioms.

Axiom 1: If $\{P \in M \; s.t. \; I \models Pa\} = \{P \in M \; s.t. \; I \models Pb\}$ then $a \sim_M b$.

Axiom 2: For any a, b in X and any set M, M' of reasons in $\mathscr{M}$ with $M \subseteq M'$, if no P in M' is true of a and b, then $a \succeq_M b$ iff $a \succeq_{M'} b$.

Axiom 1 states that, if the properties of a are the same as the properties of b, then a is indifferent to b. That is, the only way to distinguish between a to b in terms of preferences is by proposing motivating reasons. Axiom 2 states that the reasons that do not apply to a nor to b cannot decide the ranking of a and b.

Dietrich and List proved that, if Axioms 1 and 2 hold, then it is possible to associate a single weighing relation $\succeq$ defined $\mathscr{M} \times \mathscr{M}$ to the family of preferences $(\succeq_M)_{\mathscr{M}}$. The weighing relation expresses the relevance of the combinations of reasons for the preference (cf. Theorem 1 in [7]) and every preference relation $\succeq_M$ can be generated by means of a weighing order on sets of reasons. We could say that any $\succeq_M$ is *rationalised* by a set of reasons M.

We conclude this section, by illustrating the reason-based setting by means of a toy example.

Example 1. Consider the purchase of a bicycle. Suppose that the motivationally salient reasons are, in times of inflation, "being a cheap bike" (represented by C) and "being a durable bike" (represented by D). Thus, we select a set of predicates for reasons $\mathscr{R} = \{C, D\}$. We select a domain X, consisting of four alternative bikes. $X = \{cd, c\bar{d}, \bar{c}d, \bar{c}\bar{d}\}$, where, e.g., cd denotes the element of the domain X which satisfies C and D, and $\bar{c}\bar{d}$ denotes an element of X which does not satisfy C nor D. By abusing the notation, we use the same symbols for the individual constants. Thus, we are interpreting C and D in X, so that we know which bikes are cheap and which bikes are durable: $I(C) = \{cd, c\bar{d}\}$ and $I(D) = \{cd, \bar{c}d\}$.

The possible sets of motivating reasons are then $2^{\mathscr{R}} = \{\emptyset, \{C\}, \{D\}, \{C, D\}\}$. Out of them, we can select the set $\mathscr{M} = \{\{C\}, \{D\}, \{C, D\}\}$, i.e. we exclude $\emptyset$, which amounts to assuming that preferences are always based on some combination of reasons. Notice that each set of reason is coherent in this case. The preference relations, based the various sets of motivating reasons, might be as follows.

[2] To avoid proliferation of symbols, we denote by $\succ_M$ the ordering of the alternatives, while $\succ$, with no index, indicates the ordering on sets of reasons.

$$\begin{array}{ll} M = \{C,D\} & cd \succeq_M c\bar{d} \succeq_M \bar{c}d \succeq_M \bar{c}\bar{d} \\ M = \{C\} & cd \sim_M c\bar{d} \succeq_M \bar{c}d \succeq_M \bar{c}\bar{d} \\ M = \{D\} & cd \sim_M \bar{c}d \succeq_M c\bar{d} \succeq_M \bar{c}\bar{d} \end{array}$$

Many preference relations can depend on those sets of reasons, the one above is just a case. However, the family of preference relations $\succeq_M$, for $M \in \mathscr{M}$, is *property-base*, cf. Definition 2. A weighing relation exists and meets the condition of Equation (1):

$$\{C,D\} \succeq \{C\} \succeq \{D\}$$

3 Ontologies for reason-based preferences

We rephrase the reason-based approach within an ontological setting. The motivations are essentially two and they provide the two objectives of this work. Firstly, we aim at introducing a semantical understanding of reasons. In Example 1, agents are supposed to have access to a single "intended" model, while usually agents' information is represented by (the models of) a Knowledge Base. The definitions of Section 2 were indeed phrased wrt. a single "intended" model. Moreover, Example 1 conveys a lot of implicit information about the alternatives and the motivating reasons. Bikes are physical objects, in particular, they are artifacts, usually produced by some factory, marketed by some company, etc. Durability is a property of physical object (in a technical ontological jargon, it may be construed as a *quality* or a *disposition*). Cheapness is also a property, which is related to having a low cost.

Secondly, we wish to explore the reason-based model of [7] in three directions: *i*) by proposing a general mechanism to express preferences on combination of reasons, *ii*) by proposing a general strategy for computing the weighing relation between sets of motivating reasons, *iii*) by investigating the dependence between the preferences definable on combinations of reasons and the logical connections between reasons.

The first task can be approached by selecting a suitable background theory —a Knowledge Base— where reasons are embedded in a network of constraints. That is, reasons are placed within a Tbox (a *terminological box*, i.e. a set of axioms), which makes the agent's information explicit. To provide a general language to express high-level ontological definitions (such as "objects", "events", "artifacts", "qualities", "social objects", etc.), the proposal is to include in the Tbox a *foundational ontology*, such as DOLCE [10, 4]. For an overview and a comparison of the main foundational ontologies and their modelling choices, see [5]. Usually foundational ontologies are expressed in rich logical language (e.g. first-order modal logics), while here we shall limit their expressive power by deploying their decidable counterparts in OWL, cf. [5].

The second task is approached by adapting the modelling of preferences in combinatorial domains of [16, 17] to $\mathscr{DL}s$.

While we do not enter the details of how to implement the first task, we stress that the next sections constraint the possibility of defining the reason-based approach in a knowledge-dependent way, that is, by considering the models of a given ontological theory. We shall delve into the task of extending the reason-based model with combination of reasons in the subsequent paragraphs. As we shall see, the possible preferences that are defined on sets of reasons shall depend on the intended semantics of reasons, i.e. on the Knowledge Base.

3.1 Description logics

We use *Description Logics* ($\mathscr{DL}s$) as they are fundamental languages for representing *concepts*, i.e. unary predicates. Therefore, motivating reasons are here represented as concepts in DLs. We briefly introduce $\mathscr{ALC}$; for an exhaustive introduction, see [3]. We work with $\mathscr{ALC}$, however nothing prevents to apply the following definitions to richer $\mathscr{DL}s$. The syntax of $\mathscr{ALC}$ is based on three disjoint sets N_I, N_C, and N_R of *individual names*, *concept names*, and *role names*, respectively. The set of $\mathscr{ALC}$ *concepts* is generated by the following grammar, where $A \in N_C$ and $R \in N_R$.

$$C ::= A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \forall R.C \mid \exists R.C$$

A *TBox* $\mathscr{T}$ is a finite set of *general concept inclusions* (GCIs) of the form $C \sqsubseteq D$ where C and D are concepts of $\mathscr{ALC}$. The TBox is used to store general terminological (semantic) knowledge about concepts and roles. An *ABox* is a finite set of formulas of the form Ca and Rab, which express knowledge about particular objects. A *Knowledge Base* $\mathscr{K}$ consists of a $\mathscr{T}$ and an $\mathscr{A}$.

The semantics of $\mathscr{ALC}$ is defined by means of *interpretations* $I = (\Delta^I, \cdot^I)$, where Δ^I is a non-empty set, the *domain*, and $\cdot^I$ is a function mapping every individual name in N_I to an element of Δ^I, each concept name in N_C to a subset of the domain, and each role name in N_R to a binary relation on the domain. Then, I extends from concepts in N_C to the full set of $\mathscr{ALC}$-concepts inductively, cf. [3].

We say that the interpretation I is a *model* of the TBox $\mathscr{T}$ ($I \models \mathscr{T}$) iff I satisfies all the GCIs in $\mathscr{T}$, i.e. for each $C \sqsubseteq D \in \mathscr{T}$, $C^I \subseteq D^I$. An interpretation I is *a model* of an ABox $\mathscr{A}$ ($I \models \mathscr{A}$) iff I satisfies every formula in $\mathscr{A}$, i.e. if $Ca \in \mathscr{A}$, then $a^I \in C^I$ and if $Rab \in \mathscr{A}$, $(a^I, b^I) \in R^I$. An interpretation I *is a model* of a Knowledge Base, $I \models \mathscr{K}$ iff $I \models \mathscr{T}$ and $I \models \mathscr{A}$. We say that $\mathscr{K}$ is *consistent* if there exists an interpretation I that is a model of $\mathscr{K}$.

Given two concepts C and D, we say that C is *subsumed* by D w.r.t. $\mathscr{K}$, $C \sqsubseteq_{\mathscr{K}} D$, iff $C^I \subseteq D^I$, for every model I of $\mathscr{K}$. We write $C \equiv_{\mathscr{K}} D$ when $C \sqsubseteq_{\mathscr{K}} D$ and $D \sqsubseteq_{\mathscr{K}} C$. A knowledge base $\mathscr{K}$ entails an Abox formula ϕ, $\mathscr{K} \models \phi$ iff, for every interpretation I that is a model of $\mathscr{K}$, I is a model of ϕ.

3.2 Concept bases of $\mathscr{ALC}$

The approach developed in [16, 17] was designed to compactly represent utility functions over (finite) combinations of goods. In fact, goods are represented there by finite set of literals of a propositional language. To represent utility functions, the notion of a *goal base* is introduced: it is a set of weighted formulas of propositional logic (interpreted as goals) that allows for generating utility functions from sets of literals (representing goods) to (real) numbers.

In particular, we adapt [16, 17] for concepts of $\mathcal{ALC}$, following the approaches in [14] and [1] for $\mathcal{DL}s$ and in [11] for first-order logic.

Let $\mathscr{R} = \{D_1, \ldots, D_m\} \subseteq N_C$ be a finite set of concept names of $\mathcal{ALC}$ (we exclude $\top$ and $\bot$) and assume that a Knowledge Base $\mathcal{K}$ is given. Let $\mathcal{ALC}_{\mathscr{R}}$ be the set of concepts of $\mathcal{ALC}$ constructed out of concept names in $\mathscr{R}$. $\mathscr{R}$ represents the set of motivating reasons, discussed in Section 2, while $\mathcal{K}$, and in particular its TBox, shall represent the background theory that constraints the meaning of the concepts in $\mathscr{R}$. We shall assume throughout the paper that the Tbox $\mathcal{T}$ is acyclic, cf. [3], and when we use a Knowledge Base $\mathcal{K}$ that includes $\mathcal{T}$, we assume that $\mathcal{K}$ is consistent.

A *weighted concept* is a pair (C, w) where C is a concept of $\mathcal{ALC}_{\mathscr{R}}$ and $w \in \mathbb{W}$ is a value in a suitable set of values (usually real numbers, but this is not important here).

We are ready to introduce the definition of a *concept base*, which is directly inspired by the notion of a goal base in [16, 17].

Definition 3 (Concept base). *A* concept base $\mathtt{C}$ *is a finite set of weighted concepts*

$$\mathtt{C} = \{(C_1, w_1), \ldots, (C_m, w_m)\}$$

The concept base allows for expressing preferences over possible combinations of reasons in $\mathscr{R}$. An example is $\mathtt{C} = \{(D_1 \sqcap D_2, w_1), (D_1, w_2)\}$, where $w_1 > w_2$. In this case, an agent evaluates the conjunction of reasons $D_1 \sqcap D_2$ as more important than D_1 alone, while possibly not caring at all about D_2 alone, see also Example 2 below.

For each goal base $\mathtt{C}$, we can define a (utility) function on any possible combination of reasons. That is, for every $\mathtt{C}$, we can define a function $u_{\mathtt{C}} : 2^{\mathscr{R}} \to \mathbb{W}$, that takes any set of reasons $M \in 2^{\mathscr{R}}$ and returns a value. To define $u_{\mathtt{C}}$, we shall identify any set of reasons $\{D_1 \ldots, D_n\} \in 2^{\mathscr{R}}$ with their conjunction $\bigsqcap_{i \in \{1,\ldots,n\}} D_i$. In particular, we put $\bigsqcap \emptyset = \top$[3]. To compute the values of $u_{\mathtt{C}}$, we take here the sum of the weights of the concepts in $\mathtt{C}$ that are entailed by M.[4] Summing up, $u_{\mathtt{C}}$ is defined as follows.

$$u_{\mathtt{C}}(\{D_1, \ldots, D_m\}) = \sum \{w \mid (C, w) \in \mathtt{C} \text{ and } D_1 \sqcap \cdots \sqcap D_m \sqsubseteq_{\mathcal{T}} C\} \tag{2}$$

As the following example shows, the value of $u_{\mathtt{C}}$ depends on $\mathcal{T}$. In fact, the adequate notation for such functions is $u_{\mathtt{C}}^{\mathcal{T}}$. Since we shall consider one $\mathcal{T}$ at a time, we omit the superscript.[5]

Example 2. Let $\mathtt{C} = \{(D_1 \sqcap D_2, w_1), (D_1, w_2)\}$. And assume that $\mathscr{R} = \{D_1, D_2\}$. The value $u_{\mathtt{C}}(M)$ for $M \in 2^{\mathscr{R}}$ depends on $\mathcal{K}$, and in particular on $\mathcal{T}$. Consider $\mathcal{T} = \emptyset$. In this case, subsumptions $C \sqsubseteq_{\emptyset} D$ are assessed wrt. all possible models. Then, the graph of $u_{\mathtt{C}}(M)$ is as follows.

---

3 In a lattice, $\top$ is the infimum of the empty set.

4 In this paper, we assume that weights are aggregated by means of the sum, but other choices are possible (e.g. products, [17]).

5 This definition of the values returned by the concept base is termed "implication based" in [14].

M	$u_{\mathtt{C}}(M)$
$\emptyset$	0
$\{D_1\}$	w_2
$\{D_2\}$	0
$\{D_1, D_2\}$	$w_1 + w_2$

The first line returns 0, because neither $\top \sqsubseteq_{\emptyset} D_1 \sqcap D_2$ nor $\top \sqsubseteq_{\emptyset} D_1$ hold.[6] The second line returns w_2 since $D_1 \sqsubseteq_{\emptyset} D_1$. The third line returns 0, because $D_2 \sqsubseteq D_1$ does not hold in every model. The fourth line returns $w_1 + w_2$, because $D_1 \sqcap D_2 \sqsubseteq_{\emptyset} D_1$ and $D_1 \sqcap D_2 \sqsubseteq_{\emptyset} D_1 \sqcap D_2$.

Consider the case of $\mathscr{T} = \{D_1 \sqsubseteq D_2\}$. In this case, the values of $u_{\mathtt{C}}$ are as follows.

M	$u_{\mathtt{C}}(M)$
$\emptyset$	0
$\{D_1\}$	$w_1 + w_2$
$\{D_2\}$	0
$\{D_1, D_2\}$	$w_1 + w_2$

In the second line, since $D_1 \sqsubseteq D_2 \in \mathscr{T}$, $\mathscr{T}$ also entails that $D_1 \sqsubseteq D_1 \sqcap D_2$.

Notice that the notion of coherence of a set of reasons M in $2^{\mathscr{R}}$ (cf. Section 2, Definition 1) depends, in our ontological rendering, on $\mathscr{T}$. Intuitively, the elements of $\mathscr{R}$ are concept names in N_C, thus, they are just simple symbols and, if they are not logically connected by some axiom in $\mathscr{T}$, they cannot clash.

However, consider the case where $\mathscr{R} = \{D_1, D_2, D_3\}$ and $\mathscr{T} = \{D_1 \equiv \neg D_2\}$. In this case, the set $M_1 = \{D_1, D_2\}$ is not coherent, as there cannot be a $d \in D_1^I \cap D_2^I$, in any interpretation I that makes the TBox true. By contrast, $\{D_2, D_3\}$ is coherent (there are models I of $\mathscr{T}$ with a $d \in D_1^I \cap D_2^I$).

If $M \in 2^{\mathscr{R}}$ is incoherent wrt. $\mathscr{T}$, then $v_{\mathtt{C}}(M) = \sum\{w \mid (C, w) \in \mathtt{C} \text{ and } \bigsqcap M \sqsubseteq_{\mathscr{T}} C\}$ shall return the sum of all weights occurring in $\mathtt{C}$ (by *ex falso quodlibet*).

We discuss now which functions are representable, given a concept base $\mathtt{C}$ and a Tbox $\mathscr{T}$.

Definition 4. *We say that a function $f : 2^{\mathscr{R}} \to \mathbb{W}$ is represented by a concept base* $\mathtt{C}$ *iff $f = u_{\mathtt{C}}$ (we may also say that f is generated by* $\mathtt{C}$*).*

By tinkering Theorem 3.2 in [16], we could prove that every function from $2^{\mathscr{R}} \to \mathbb{W}$ can be *represented* by a concept base $\mathtt{C}$, at least when $\mathbb{W} = \mathbb{R}$. We leave the details of the proof for a dedicated work. Notice however that the case of $\mathscr{DL}s$ is more delicate than the propositional case of [16]. The general representation result holds only for the case of an empty $\mathscr{T}$. That is, the set of functions that are representable depends on the Tbox $\mathscr{T}$, as the following example shows.

Example 3 (Representability). Let $\mathscr{R} = \{D_1, D_2, D_3\}$ and $\mathscr{T} = \{D_1 \equiv D_2\}$. Consider a function $f : 2^{\mathscr{R}} \to \mathbb{R}$ such that $f(\{D_1\}) \neq f(\{D_2\})$. We show that, in this case, f

[6] That is, it is not true that, in every model, $\Delta \subseteq D_1^I$ nor that $\Delta \subseteq (D_1 \cap D_2)^I$.

cannot be represented by any $u_{\mathtt{C}}$. By contradiction, assume that there exists a $\mathtt{C}$ such that $f = u_{\mathtt{C}}$. Thus, $u_{\mathtt{C}}(\{D_1\}) \neq u_{\mathtt{C}}(\{D_2\})$. That is,

$$\sum\{w \mid (C,w) \in \mathtt{C} \text{ and } D_1 \sqsubseteq_{\mathscr{T}} C\} \neq \sum\{w \mid (C,w) \in \mathtt{C} \text{ and } D_2 \sqsubseteq_{\mathscr{T}} C\}$$

The inequality implies that $\{C \text{ occurring in } \mathtt{C} \mid D_1 \sqsubseteq_{\mathscr{T}} C\} \neq \{C \text{ occurring in } \mathtt{C} \mid D_2 \sqsubseteq_{\mathscr{T}} C\}$. Thus, there must be a C_0 occurring in $\mathtt{C}$ such that, e.g., $D_1 \sqsubseteq_{\mathscr{T}} C_0$ and $D_2 \not\sqsubseteq_{\mathscr{T}} C_0$. That is, in every model I of $\mathscr{K}$, $D_1^I = D_2^I$ is both included in C_0^I and not included in C_0^I, which is a contradiction.

Therefore, f cannot be represented by any $\mathtt{C}$. Other examples can be envisaged, e.g. when there are sets of reasons that are incoherent wrt. $\mathscr{T}$.

We shall therefore restrict to those utility functions on combinations of reasons that are compatible with the information conveyed by $\mathscr{T}$.

Definition 5 (Compatibility). *A function $f : 2^{\mathscr{R}} \to \mathbb{W}$ is compatible with the TBox $\mathscr{T}$ iff there exists a concept base $\mathtt{C}$ such that f is representable, i.e. $f = u_{\mathtt{C}}$.*

Notice that, if $\mathscr{K}$ is consistent, then the $\mathscr{A}$ of $\mathscr{K}$ has no effect on which functions are representable. The motivation for the restriction to preferences and utility functions that are compatible with $\mathscr{T}$ is that $\mathscr{T}$ specifies the meanings of the reasons in $\mathscr{R}$, which is supposed to be acknowledged by the agents who are justifying their preferences in terms of motivating reasons.

In [16], it is possible to characterise several classes of utility functions in terms of the formulas occurring in the goal base. In the case of $\mathscr{DL}s$ this is not straightforward. E.g. if the TBox is empty and $\mathtt{C}$ contains only concept names $D \in \mathscr{R}$, then $u_{\mathtt{C}}$ is additive (i.e. $u_{\mathtt{C}}(D_1,\dots,D_n) = \sum_{i=1}^{i=n} u_{\mathtt{C}}(\{D_i\})$). However, if the TBox is not empty, then we cannot ensure the independence of each D_i. While in the propositional case of [16], it suffices to constrain the language inside the goal base, in the case of $\mathscr{DL}s$ one needs to study the correspondence between the inferential features of the TBox and the definable classes of functions. We leave this interesting point for a future dedicated work.

4 Concept bases and reason-based preferences

Given a concept base $\mathtt{C}$, by means of $u_{\mathtt{C}}$, we can always define an ordering of sets of motivating reasons, as follows.

$$M \succeq_{\mathtt{C}} M' \text{ iff } u_{\mathtt{C}}(M) \geq u_{\mathtt{C}}(M') \tag{3}$$

As $\succeq_{\mathtt{C}}$ comes from $\geq$, it is always reflexive, transitive and complete. Moreover, since $\succeq_{\mathtt{C}}$ comes from $u_{\mathtt{C}}$ it is always compatible with the TBox $\mathscr{T}$. By contrast, given an ordering $\succeq$ on $2^{\mathscr{R}}$, it is not the case that $\geq$ is always representable as $\succeq_{\mathtt{C}}$, for some $\mathtt{C}$ and $\mathscr{K}$. Consider the following example.

Example 4. Let $\mathscr{R}$ and $\mathscr{K}$ as in Example 3. Define f such that $f(\{D_1\}) > f(\{D_2\})$ and $\succ$ such that $\{D_1\} \succ \{D_2\}$. Then, there is no concept base $\mathtt{C}$ such that $\succ = \succeq_{\mathtt{C}}$.

Therefore, we shall confine to ordering on $2^{\mathscr{R}}$ which are compatible with $\mathscr{K}$, in the following sense.

Definition 6. *An order relation $\succeq \subseteq 2^{\mathscr{R}} \times 2^{\mathscr{R}}$ is compatible with $\mathscr{K}$ iff there exists a concept base* $\mathtt{C}$ *such that $\succeq = \succeq_{\mathtt{C}}$, where $\succeq_{\mathtt{C}}$ is defined by Equation 3.*

We turn now to discussing how preferences on alternatives depend on their motivating reasons. Alternatives are here interpreted as individual names $a \in N_I$. Let $X \subseteq N^I$ be a finite set of alternatives. The utility of an alternative a, in a certain situation, is given by the sum of the weights of the motivating reasons that a satisfies in that situation.

A "situation" is modelled by means of a knowledge base $\mathscr{K}$, where the TBox $\mathscr{T}$ is given and the Abox $\mathscr{A}$ describes the (some) alternatives by the (some) reasons that they satisfy, i.e. $\mathscr{A}$ is a set of formulas of the form $D_i x$ for $x \in X$ and $D_i \in \mathscr{R}$. When $M \subseteq \mathscr{R}$, we denote by $\bigsqcap M$ the concept $\bigsqcap_i D_i$, for $D_i \in M$.

Definition 7 (Utility of alternatives). *The utility of $a \in X$, given a concept base* $\mathtt{C}$*, a Knowledge Base $\mathscr{K}$, and a set of reasons $M = \{D_1, \dots, D_m\} \subseteq \mathscr{R}$ is defined as follows.*

$$u^M_{\mathtt{C}}(a) = \sum \{w \mid (C,w) \in \mathtt{C}, \bigsqcap M' \sqsubseteq_{\mathscr{T}} C, \textit{ for } M' \subseteq M, \textit{ and } \mathscr{K} \models D_i a, \textit{ for all } D_i \in M'\}$$

Thus, the utility of a given a concept base $\mathtt{C}$, when the motivating reasons are M, is the sum of the weights of the concepts C such that: *i*) $(C,w) \in \mathtt{C}$, *ii*) C is entailed by some of motivating reasons in M', and *iii*) a satisfies those motivating reasons $D_i \in M'$.

Example 5. Let $X = \{a,b\} \subseteq N_I$, $\mathscr{R} = \{D_1, D_2, D_3, D_4\}$, $\mathscr{T} = \{D_1 \sqsubseteq D_2\}$. The ABox is $\mathscr{A} = \{D_1 a, D_2 b, D_3 b, D_4 a\}$. Let $\mathtt{C} = \{(D_1 \sqcap D_2, w_1), (D_1 \sqcap D_3, w_2), (D_3, w_3), (D_4, w_4)\}$. Consider, for example, the case where $M = \{D_1, D_3\}$ are the sole motivating reason. The value $u^{\{D_1,D_3\}}_{\mathtt{C}}(a)$ is then obtained as follows. We have that $D_1 \sqsubseteq_{\mathscr{T}} D_1 \sqcap D_2$ (given that $D_1 \sqsubseteq D_2 \in \mathscr{T}$) and $D_1 a$ is in $\mathscr{A}$, so the weight w_1 is obtained. The weight w_2 is not added, because $D_1 \sqcap D_3 \sqsubseteq_{\mathscr{T}} D_1 \sqcap D_3$, but $D_3 a$ is not in $\mathscr{K}$. Note that the weight w_4 is not added, even if $D_4 a \in \mathscr{K}$, because D_4 is not a motivating reason in this case.

By Definition 7, a preference order, that depends on M and $\mathtt{C}$, on the set of alternatives can be defined.

$$a \succeq_M b \text{ iff } u^M_{\mathtt{C}}(a) \geq u^M_{\mathtt{C}}(b) \tag{4}$$

We claim that $a \succeq_M b$, for $M \in \mathscr{M}$, is a preference based on reasons, in the sense of [7]. That is, $a \succeq_M b$ satisfies Axiom 1 and 2, rephrased in this context.

Proposition 1. *Let $a \succeq_M b$, for $M \in \mathscr{M}$, defined as in Equation 4. If $\{D \in M \mid \mathscr{K} \models Da\} = \{D \in M \mid \mathscr{K} \models Db\}$, then $a \sim_M b$.*

Proof. Suppose $\{D \in M \mid \mathscr{K} \models Da\} = \{D \in M \mid \mathscr{K} \models Db\}$. For any M and $M' \subseteq M$, we have that

$$\{C \mid (C,w) \in \mathtt{C} \text{ and } \bigsqcap M' \sqsubseteq_{\mathscr{T}} C \text{ and } \mathscr{K} \models D_i a\}$$

$$=$$

$$\{C \mid (C,w) \in \mathtt{C} \text{ and } \bigsqcap M' \sqsubseteq_{\mathscr{T}} C \text{ and } \mathscr{K} \models D_i b\}$$

Therefore, $u^M_{\mathtt{C}}(a) = u^M_{\mathtt{C}}(b)$, hence $a \sim_M b$.

Proposition 2. *Let $a \succeq_M b$, for $M \in \mathscr{M}$, defined as in Equation 4. For any a, $b \in X$ and any $M_1, M_2 \in \mathscr{M}$, such that $M_1 \subseteq M_2$, if there is no $D \in M_2 \setminus M_1$ such that $\mathscr{K} \models Da$ or $\mathscr{K} \models Db$, then $a \succeq_{M_1} b$ iff $a \succeq_{M_2} b$.*

Proof. Assume that $M_1' \subseteq M_1$, $M_2' \subseteq M_2$. By assumption, $M_1 \subseteq M_2$ and there is no $D \in M_2 \setminus M_1$ such that $\mathscr{K} \models Da$ or $\mathscr{K} \models Db$. Therefore, we have that, for all $D_i \in M_2$:

$$\{C \mid (C,w) \in \mathtt{C}, \bigsqcap M_1' \sqsubseteq_{\mathscr{T}} C, \mathscr{K} \models D_i a\} = \{C \mid (C,w) \in \mathtt{C}, \bigsqcap M_2' \sqsubseteq_{\mathscr{T}} C, \mathscr{K} \models D_i a\}$$

$$\{C \mid (C,w) \in \mathtt{C}, \bigsqcap M_1' \sqsubseteq_{\mathscr{T}} C, \mathscr{K} \models D_i b\} = \{C \mid (C,w) \in \mathtt{C}, \bigsqcap M_2' \sqsubseteq_{\mathscr{T}} C, \mathscr{K} \models D_i b\}$$

Therefore $u^{M_1}_{\mathtt{C}}(a) \geq u^{M_1}_{\mathtt{C}}(b)$ iff $u^{M_2}_{\mathtt{C}}(a) \geq u^{M_2}_{\mathtt{C}}(b)$.

We can finally restate the definition of property-based preference relation (cf. Definition 2), to highlight the dependence on a $\mathscr{K}$.

Definition 8. *Let $\mathscr{M} \subseteq 2^{\mathscr{R}}$. A family of preference relations $\succeq_M$ for $M \in \mathscr{M}$ is property-based iff there exists a weighing relation $\succ \subseteq \mathscr{M} \times \mathscr{M}$ such that:*

$$a \succeq_M b \textit{ iff } \{D \in M \mid \mathscr{K} \models Da\} \succeq \{D \in M \mid \mathscr{K} \models Db\}$$

We show that $\succeq_{\mathtt{C}}$ is indeed a weighing relation, for preferences $a \succeq_M b$ defined according to Equation 4.

Proposition 3. *Given a Knowledge Base $\mathscr{K}$, for every family of preference $\succeq_M$ with $M \in \mathscr{M}$ defined as in Equation 4, $\succeq_{\mathtt{C}}$ is a weighing relation for $\succeq_M$.*

Proof. We have the following sequence of equivalences: $a \succeq_M b$ iff $u^M_{\mathtt{C}}(a) \geq u^M_{\mathtt{C}}(b)$ iff

$$\sum\{w \mid (C,w) \in \mathtt{C}, \bigsqcap M' \sqsubseteq_{\mathscr{T}} C, \text{ for } M' \subseteq M, \mathscr{K} \models D_i a, \text{ for all } D_i \in M'\}$$
$$\geq$$
$$\sum\{w \mid (C,w) \in \mathtt{C}, \bigsqcap M' \sqsubseteq_{\mathscr{T}} C, \text{ for } M' \subseteq M, \text{and } \mathscr{K} \models D_i b, \text{ for all } D_i \in M'\}$$
$$\text{iff}^{(1)}$$
$$u_{\mathtt{C}}(\{D_1,\dots,D_n\} \mid D_i \in M \text{ and } \mathscr{K} \models D_i a\}) \geq u_{\mathtt{C}}(\{D_1,\dots,D_n\} \mid D_i \in M \text{ and } \mathscr{K} \models D_i b\})$$
$$\text{iff}^{(2)}$$
$$\{D \in M \mid \mathscr{K} \models Da\} \succeq_{\mathtt{C}} \{D \in M \mid \mathscr{K} \models Db\}$$

The step (1) is justified by noticing that $\{D_1,\dots,D_n\}$ are concepts in $\bigsqcap M'$ that entail C, so the value of the sum of the weights of the Cs that are satisfied is $u_{\mathtt{C}}(\{D_1,\dots,D_n\})$. Step (2) follows by definition of $\succeq_{\mathtt{C}}$, cf. Definition 3.

To sum up, we have restricted ourselves to preference and utility functions on alternatives in X that are compatible with a Knowledge Base by construction (cf. Definition 4). As we have shown, such preferences are reason-based and $\succeq_{\mathsf{C}}$ is an adequate weighing relation.

If we start from "any" family of preference relations $\succeq_M$ on alternatives in X (i.e. not defined using Definition 4) and then we apply Theorem 1 in [7], we could determine whether $\succeq_M$ is reason-based (i.e. if it satisfies Axiom 1 and 2). However, this would not guarantee compatibility with the Knowledge Base.

5 Conclusions and future work

We have shown that the reason-based approach proposed in [7] can be adapted for the case where: *i*) reasons are expressed as concepts of $\mathscr{ALC}$, *ii*) preference relations and utility functions on sets of reasons are defined by means of a concept base (cf. Definition 3), and *iii*) preferences on alternatives depend on preferences on sets of reasons (cf. Equation 4).

The main feature of this setting is that it is knowledge-dependent, it relies on a Knowledge Base that confers meanings to the motivating reasons. To cope with the possible inconsistencies between the Knowledge Base and the preferences, we restricted to preferences and utility functions that are defined by means of concept bases, which are compatible with the Knowledge Base by design. In fact, this restriction does not limit the range of definable preferences, once we are willing to accept that an agent is rational, has a consistent Knowledge Base, and, consequently, has preferences that are compatible with it.

This framework paves the way for enriching the ontological side of the Knowledge Base with (the OWL version) of a foundational ontology, to improve the representation of motivating reasons and to enable the application to various domains of interest.

There are mainly three directions for future work. Firstly, as we mentioned previously, it is interesting to study the property of the TBox that enables defining classes of utility functions (e.g. superadditive, subadditive, *k*-additive, modular, cf. [16]).

Secondly, it is interesting to study preference aggregation based on motivating reasons. In this case, we could assume that agents agree on the Knowledge Base, i.e. on the meanings of the possible reasons, so that the ontological side of the Knowledge Base operates as a vocabulary that is shared among the agents. Agents' preferences may still be in conflict, as they are justified by different, and possibly conflicting, weighing relations.

Finally, we suggested the idea that the TBox captures the meanings of the motivating reasons. However, we depicted here a single-agent scenario. It is interesting to study reason-based preferences in the case of multiple agents, who have possibly contrasting views of reasons, namely, they might endorse possibly conflicting TBoxes. In this case, techniques of ontology integration, revision, alignment and aggregation are required to interact with the preferences defined in terms of motivating reasons.

Acknowledgements This research is partially supported by Italian National Research Project PRIN2020 2020SSKZ7R. I would like to thank the anonymous reviewers for their valuable comments on the preliminary version of this paper.

References

1. Acar, E., Fink, M., Meilicke, C., Thorne, C., Stuckenschmidt, H.: Multi-attribute decision making with weighted description logics. IfCoLog Journal of Logics and its Applications **4**(7), 1973–1996 (2017)
2. Arrow, K.: Social Choice and Individual Values. New Haven: Yale University Press. (1963)
3. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)
4. Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., Sanfilippo, E.M., Vieu, L.: Dolce: A descriptive ontology for linguistic and cognitive engineering. Applied ontology **17**(1), 45–69 (2022)
5. Borgo, S., Galton, A., Kutz, O.: Foundational ontologies in action. Applied Ontology **17**(1), 1–16 (2022). https://doi.org/10.3233/AO-220265, https://doi.org/10.3233/AO-220265
6. Dietrich, F., List, C.: A reason-based theory of rational choice. Nous **47**(1), 104–134 (2013)
7. Dietrich, F., List, C.: Where do preferences come from? International Journal of Game Theory **42**(3), 613–637 (2013)
8. Galliani, P., Kutz, O., Porello, D., Righetti, G., Troquard, N.: On knowledge dependence in weighted description logic. In: Calvanese, D., Iocchi, L. (eds.) GCAI 2019. Proceedings of the 5th Global Conference on Artificial Intelligence, Bozen/Bolzano, Italy, 17-19 September 2019. EPiC Series in Computing, vol. 65, pp. 68–80. EasyChair (2019). https://doi.org/10.29007/hjt1, https://doi.org/10.29007/hjt1
9. Liu, F., et al.: Changing for the better: Preference dynamics and agent diversity. ILLC (2008)
10. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Wonderweb deliverable d18. Tech. rep., CNR (2003)
11. Masolo, C., Porello, D.: Representing concepts by weighted formulas. In: Formal Ontology in Information Systems - Proceedings of the 10th International Conference, FOIS 2018, Cape Town, South Africa, 19-21 September 2018. pp. 55–68 (2018)
12. Osherson, D., Weinstein, S.: Preference based on reasons. The Review of Symbolic Logic **5**(01), 122–147 (2012)
13. Porello, D.: Ranking judgments in arrow's setting. Synthese **173**(2), 199–210 (2010)
14. Ragone, A., Noia, T.D., Donini, F.M., Sciascio, E.D., Wellman, M.P.: Weighted description logics preference formulas for multiattribute negotiation. In: Scalable Uncertainty Management, Third International Conference, SUM 2009, Washington, DC, USA, September 28-30, 2009. Proceedings. pp. 193–205 (2009). https://doi.org/10.1007/978-3-642-04388-8_16, https://doi.org/10.1007/978-3-642-04388-8_16
15. Sen, A.K.: Choice functions and revealed preference. The Review of Economic Studies **38**(3), 307–317 (1971)
16. Uckelman, J., Chevaleyre, Y., Endriss, U., Lang, J.: Representing utility functions via weighted goals. Mathemathical Logic Quartely **55**(4), 341–361 (2009)
17. Uckelman, J., Endriss, U.: Compactly representing utility functions using weighted goals and the max aggregator. Artificial Intelligence **174**(15), 1222–1246 (2010)