Reasoning about Interest-Based Preferences

Wietske Visser, Koen V. Hindriks, and Catholijn M. Jonker

Man-Machine Interaction Group, Delft University of Technology
Delft, The Netherlands
{wietske.visser,k.v.hindriks,c.m.jonker}@tudelft.nl

Abstract. In decision making, negotiation, and other kinds of practical reasoning, it is necessary to model preferences over possible outcomes. Such preferences usually depend on multiple criteria. We argue that the criteria by which outcomes are evaluated should be the satisfaction of a person's underlying interests: the more an outcome satisfies his interests, the more preferred it is. Underlying interests can explain and eliminate conditional preferences. Also, modelling interests will create a better model of human preferences, and can lead to better, more creative deals in negotiation. We present an argumentation framework for reasoning about interest-based preferences. We take a qualitative approach and provide the means to derive both ceteris paribus and lexicographic preferences.

1 Introduction

We present an approach to qualitative, multi-criteria preferences that takes underlying interests explicitly into account. Reasoning about interest-based preferences is relevant in decision making, negotiation, and other types of practical reasoning. Since our long-term goal is the development of a negotiation support system, the motivations and examples in this paper are mainly taken from the context of negotiation, but the main ideas apply equally well in other contexts.

The goal of a negotiation support system is to help a human negotiator reach a better deal in negotiation. The quality of a deal is determined for a large part by the user's personal preferences. A deal generally consists of multiple issues. For example, when applying for a new job, some issues are the position, the salary, and the possibility to work part-time. For a complete deal, negotiators have to agree on the value for every issue. The satisfaction of a negotiator with a possible outcome depends on his preferences.

Since the number of possible outcomes is typically very large (exponential in the number of issues), it is not feasible to have the user express his preferences over all possible outcomes directly. It is common to compute or derive preferences over possible outcomes from preferences over the possible values of issues and a weighing or importance ordering of the issues. One of the best-known approaches is multi-criteria utility theory [1], a quantitative approach where preferences are expressed by numeric utilities. Since such quantities are hard for humans to provide, qualitative approaches have been proposed too, e.g. [2]. Our approach is also of a qualitative nature.

In this paper we argue that issues alone are not enough to derive outcome preferences. Instead, we will focus on modelling underlying interests and their relation to issues.

J. Filipe and A. Fred (Eds.): ICAART 2011, CCIS 271, pp. 115–130, 2013.

There are several reasons for taking interests into account. First, underlying interests can explain and eliminate conditional preferences. Consider the following example. If it rains, I prefer to take my umbrella, but if it doesn't, I prefer not to take it. This is a conditional preference; my preference over taking my umbrella depends on the circumstance of rain. Underlying interests can explain such conditional preferences: I prefer to take my umbrella when it rains because I do not want to get wet, and I prefer not to take it when it's dry because I don't want to carry things unnecessarily. If we take such interests as criteria on which to base preference, we can eliminate conditional preferences entirely. We will get back to this in more detail later. Second, interest-based negotiation is said to lead to better outcomes than position-based negotiation [3,4]. By understanding one's own and the other party's reasons behind a position and discussing these interests, people are more likely to find more creative options in a negotiation and by that reach a mutually acceptable agreement more easily. A well-known example is that of the two sisters negotiating about the division of an orange. They both want the orange, and end up splitting it in half. Had they known each other's underlying interests, they would have reached a better deal: one sister only needed the peel to make a cake and would gladly have let the other sister have all of the flesh for her juice. Third, thinking about underlying interests is a very natural, human thing to do. Interests are what really matters to people, they are what drive them in their decisions and opinions. Taking underlying interests explicitly into account will result in a better model of human preferences. Such a model is also suited for explanation of the reasoning and advice of a support system.

This last point brings us to the motivation for using argumentation to reason about interest-based preferences. Reasoning by means of arguments is a very human type of reasoning. People often base their decisions on (mental) lists of arguments in favour of and against certain decisions. Therefore argumentation is suitable for explanation of a system's reasoning to a human user. Another advantage of argumentation is that it is a kind of defeasible reasoning. It is able to reason with incomplete, uncertain and contradictory information. Finally, argumentation can be used to (try to) persuade the opponent during negotiation (but this is outside the scope of this paper).

The paper is organised as follows. In Section 2 we introduce and discuss the most important concepts that we will use throughout the paper. Then, in Section 3, we give an overview of existing approaches to preferences and underlying interests. We give some more details about qualitative multi-criteria preferences in Section 4. In Section 5 we motivate the explicit modelling of underlying interests, illustrated with examples. Our own approach is presented in Section 6. Finally, Section 7 concludes the paper.

2 Concepts

Before we go on, we will clarify some important concepts that we will use. In negotiation, *issues* are the matters which are under negotiation. An issue is a concrete, negotiable aspect such as monthly salary or number of holidays. Every issue has a set or range of possible values. The value of an issue in a given instance can be objectively determined (e.g. €2400, 30 days). Issues and their possible values typically depend on the domain. Besides the issues under negotiation, there may be other properties of a deal that influence preferences. For example, the location of the company that you are applying to work for can be very important, because it determines the duration of your daily commute, but it is hardly negotiable. Still, such properties are important in negotiation. If, for example, you already got an offer from another company near your home, you will only consider offers that are better taking the location into account.

A *possible outcome* or possible deal has a specified value for every issue. All bids made during a negotiation are possible outcomes. For example, a possible outcome could be a job contract for the position of programmer, with a salary of €3000 gross per month, with 25 holidays, for the duration of one year with the possibility of extension. Any other assignment to the issues would constitute a different outcome. It is the user's preferences over such possible outcomes that we are interested in.

With *criteria* we mean the features on which a preference between outcomes is based. It is common to base preferences directly on the negotiated issues; in that case the issues are the criteria. In this paper we argue that not issues, but underlying interests should be used as criteria.

Many terms are used for what we consider to be *underlying interests*, such as fundamental objectives, values, concerns, goals and desires. In our view, an interest can be any kind of motivation that leads to a preference. Essentially, a preference depends on how well your interests are met in the outcomes to be compared. The degree to which interests are met is influenced by the issues, but there is not necessarily a one-to-one relation between issues and interests. For example, an applicant with childcare responsibilities will have the interest that the children are taken care of after school. This interest can be met by various different issues, for example part-time work, the possibility to work from home, a salary that will cover childcare expenses, etc. One issue may also contribute to multiple interests. Many issues that deal with money do so, because the interests different people have for using the money will be diverse.

3 Related Work

Existing literature about preferences is abundant and very diverse. In this section we briefly discuss the approaches that are most closely related to our interests.

Interest-based negotiation is discussed in [4]. However, this approach has a particular view on negotiation as an allocation of indivisible and non-sharable resources. The resources are needed to carry out plans to reach certain goals. Even though the goals can be seen as underlying interests, it is hard to model e.g. negotiation about a job contract as an allocation of resources. Salary might be an allocation of money, but other issues, like position or start date, cannot be translated as easily into resources.

Argumentation about preferences has been studied extensively in the context of *decision making* [5,6,7,8]. The aim of decision making is to choose an action to perform. The quality of an action is determined by how well its consequences satisfy certain criteria. For example, [5] present an approach in which arguments of various strengths in favour of and against a decision are compared. However, it is a two-step process in which argumentation is used only for epistemic reasoning. In our approach, we combine reasoning about preferences and knowledge in a single argumentation framework.

Within the context of argumentation, an approach that is related to underlying interests is *value-based argumentation* [9,10]. Values are used in the sense of 'fundamental social or personal goods that are desirable in themselves' [10], and are used as the basis for persuasive argument in practical reasoning. In value-based argumentation, arguments are associated with values that they promote. Values are ordered according to importance to a particular audience. An argument only defeats another argument if it attacks it and the value promoted by the attacked argument is not more important than the value promoted by the attacker. We will illustrate this with a little example. Consider two job offers a and b. a offers a higher salary, but b offers a better position. We can construct two mutually attacking preference arguments, A: 'I prefer job offer a over job offer b because it has a higher salary', and B: 'I prefer job offer b over job offer a because it has a better position'. In Dung-style argumentation frameworks [11], there is no way to choose between two mutually attacking arguments (unless one is defended and the other is not). In value-based argumentation, we could say that preferring a over b promotes the value of wealth (w), and preferring b over a promotes the value of status (s), and e.g. wealth is considered more important than status. In this case A defeats B, but not the other way around.

In this framework, every argument is associated with only one value, while in many cases there are multiple values or interests at stake. [12] define so-called *value-specification argumentation frameworks*, in which arguments can support multiple values, and preference statements about values can be given. However, the preference between arguments is not derived from the preference between the values promoted by the arguments. Besides, there is no guarantee that a value-specification argumentation framework is consistent, i.e., some sets of preference statements do not correspond to a preference ordering on arguments.

In value-based argumentation, we cannot argue about what values are promoted by the arguments or the ordering of values; this mapping and ordering are supposed to be given. But these might well be the conclusion of reasoning, and might be defeasible. Therefore, it would be natural to include this information at the object level. [13] describe some argument schemes regarding the influence of certain perspectives on values. However, for the aggregation of multiple values, they assume a given order on sets of values, whereas we want to derive such an order from an order on individual values.

4 Qualitative Multi-criteria Preferences

Regardless of whether we take issues or interests as criteria, we need to be able to model multiple criteria. In any realistic setting, preferences are determined by multiple criteria and the interplay between them. Therefore we shortly introduce two well-known approaches to multi-criteria preferences which we will use in our framework.

One approach is *ceteris paribus* ('all else being equal') comparison. One outcome is preferred to another ceteris paribus, if it is better on some criteria and the same on all other criteria. This approach has been widely used since [14]. Also [15] derive preferences from sets of goals in a ceteris paribus way. In [16], ceteris paribus comparison is combined with conditional preferences in a graphical preference language called CP-nets. The preference order resulting from ceteris paribus comparison is not complete; an

Table 1. Satisfaction of issues and interests

	a. Issues				**b.** Interests		
	high salary	high position	full-time		wealth	status	family time
a	✓	✓	✓	*a*	✓	✓	✗
b	✓	✓	✗	*b*	✓	✓	✓
c	✓	✗	✓	*c*	✓	✓	✗
d	✓	✗	✗	*d*	✓	✓	✓
e	✗	✓	✓	*e*	✗	✓	✗
f	✗	✓	✗	*f*	✗	✓	✓
g	✗	✗	✓	*g*	✗	✗	✗
h	✗	✗	✗	*h*	✗	✗	✓

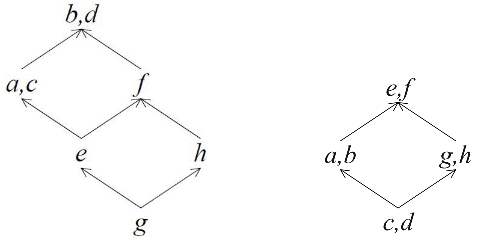
outcome satisfying criterion G but not H cannot be compared to an outcome satisfying H but not G.

Another well-known approach is the *lexicographic* preference ordering (see e.g. [2], where it is denoted #). Here, preferences over outcomes are based on a set of relevant criteria, which are ranked according to their importance. The importance ranking of criteria is defined by a total preorder $\succeq$, which yields a stratification of the set of criteria into importance levels. Each importance level consists of criteria that are equally important. The lexicographic preference ordering first considers the highest importance level. If some outcome satisfies more criteria on that level than another, then the first is preferred over the second. If two outcomes satisfy the same number of criteria on this level, the next importance level is considered, and so on. Two outcomes are equally preferred if they satisfy the same number of criteria on every level.

We use a slightly more abstract definition of preference that covers both ceteris paribus and lexicographic preferences. Let $\mathcal{C}$ be a set of binary criteria, ordered according to importance by a preorder $\succeq$. If $P \succeq Q$ and not $Q \succeq P$, we say that P is strictly more important than Q and write $P \succ Q$. If $P \succeq Q$ and $Q \succeq P$, we say that P is equally important as Q and write $P \approx Q$. $\mathcal{C}$ can be divided into equivalence classes induced by $\approx$, which we call importance levels. An importance level L is said to be more important than L' iff the criteria in L are more important than the criteria in L'. Let $\mathcal{O}$ be a set of outcomes, and *sat* a function that maps outcomes $a \in \mathcal{O}$ to sets of criteria $\mathcal{C}_a \in 2^{\mathcal{C}}$. If $P \in sat(a)$, we say that a satisfies P.

Definition 1. (Preference). An outcome a is *strictly preferred* to another outcome b if it satisfies more criteria on some importance level L, and for any importance level L' on which b satisfies more criteria than a, there is a more important level on which a satisfies more criteria than b. An outcome a is *equally preferred* as another outcome b if both satisfy the same number of criteria on every importance level.

The least specific importance order possible is the identity relation, in which case the importance levels are all singletons and no importance level is more important than any other. In this case, the preference definition is equivalent to ceteris paribus preference (if a is preferred to b ceteris paribus, there are no criteria that b satisfies but a does not). If the importance order is a total preorder, the definition is equivalent to lexicographic



(a) For Mark. (b) For Jones.

Fig. 1. Ceteris paribus preference orderings (arrows point towards more preferred outcomes)

preference. In general, the more information about the relative importance of interests is known, the more preferences can be derived. We note that lexicographic preferences subsume ceteris paribus preferences in the sense that if one outcome is preferred to another ceteris paribus, it is also preferred lexicographically, regardless of the importance ordering on criteria.

5 Modelling Interests

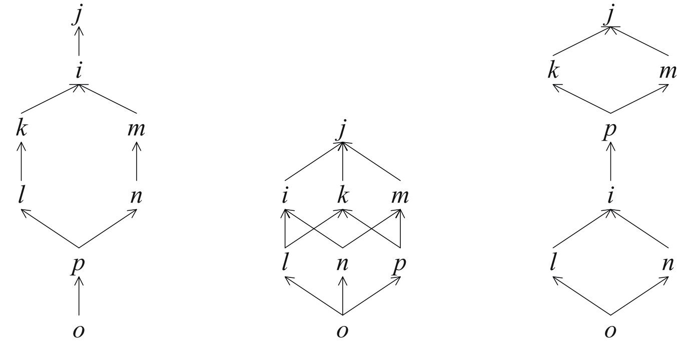
We will illustrate the ideas presented in this paper by means of an example. Mark has applied for a job at a company called Jones. After the first interview, they are ready to discuss the terms of employment. There are three issues on the table: the salary, the position, and whether the job is full-time or part-time. All possible outcomes are listed in Table 1a. After some thought, Mark has determined that the interests that are at stake for him are wealth, status, and time with his family. A high position will give status. A high salary will provide both wealth and status. A part-time job will give him time to spend with his family. Table 1b shows which interests each of the outcomes satisfies.

All information is encoded in a knowledge base, which consists of three parts.

- *Facts* about the properties of the outcomes to be compared. When comparing offers in negotiation, these may be the values for each issue, or any other relevant properties. Facts are supposed to be objectively determined.
- A set of *interests* of a negotiator. Underlying interests are personal and subjective, although they can sometimes be assumed by default. Interests may vary according to importance. If no importance ordering is given, the ceteris paribus principle can be used to derive preferences. The more information about the relative importance of interests is known, the more preferences can be derived. If there is a total preorder of interests according to importance, a complete preference ordering over possible outcomes can be derived using the lexicographic principle.
- *Rules* relating issues and other outcome properties to interests. These rules can be very subjective, e.g. some people consider themselves very wealthy if they earn €3000 gross salary per month, while for others this may be a pittance. Even so, there can still be default rules that apply in general, e.g. that a high salary promotes wealth for

Table 2. Outcomes in the evening dress example

	a. Issues			**b.** Interests	
	jacket	pants	shirt		good combi-nation
i	b	b	w	*i*	✗
j	b	b	r	*j*	✓
k	b	w	w	*k*	✓
l	b	w	r	*l*	✗
m	w	b	w	*m*	✓
n	w	b	r	*n*	✗
o	w	w	w	*o*	✗
p	w	w	r	*p*	✓



(a) Preference graph induced by CP-net.

(b.) Ceteris paribus ordering with interests.

(c) Ceteris paribus ordering with interests, good combination most important.

Fig. 2. Preference orderings (arrows point towards more preferred outcomes)

the employee. The relation between issues and interests does not have to be one-to-one. There may be multiple issues that can satisfy an interest, some issues may satisfy multiple interests at once, or a combination of issues may be needed to fulfill an interest. As is common in defeasible reasoning, there may be exceptions to rules. For example, one might say that a high position ensures status in general, but this effect is cancelled out if the job is badly paid.

With the inference scheme of defeasible modus ponens (see scheme 1 in Table 4), arguments can be constructed that derive statements about what interests are satisfied by possible outcomes, based on their issue values and the rules relating issues to interests. The conclusions from these arguments are summarized in Table 1b. If we compare the possible outcomes ceteris paribus, we can construct a partial preference order for Mark, with *b* and *d* being the most preferred options, and *g* the least preferred (see Figure 1a). This preference order is not complete. To determine Mark's preference between *a* and *c* on the one hand and *f* on the other hand, we need to know whether wealth or family time is more important to him. If wealth is more important, Mark will prefer *a* or *c*.

If family time is more important, he will prefer f. Similarly, to determine a preference between e and h, we need to know whether status or family time is more important.

The company Jones has two major interests: it needs a manager and it has to cut back on expenses. These interests relate directly (one-to-one) to high position and low salary. The ceteris paribus preference ordering for Jones is displayed in Figure 1b.

The Added Value of Interests. It may seem that using interests next to issues just introduces an extra layer in reasoning. From the issues and the relations between issues and interests, we derive the interests that are met by outcomes, and from that we derive preferences. Would it not be easier to derive the preferences directly from the issues? We could just state that Jones has the interests of high position and low salary, optionally with an ordering between them, and we would be able to derive Jones' preferences from that. This is because in this case there is a one-to-one relation between interests and issues: every interest is met by exactly one issue, and every (relevant) issue meets exactly one interest.

There are good reasons, however, why this approach is not always a good solution. Consider for example Mark's preferences. A high salary satisfies both wealth and status, and status can be satisfied by either a high salary or a high position. Because of this, the (partial) preference ordering we determined for Mark cannot be defined as a ceteris paribus ordering if the issues are taken as criteria. This is because high position as criterion is dependent on high salary: if the salary is not high, then high position is a distinguishing criterion, but if the salary is high, high position is not relevant anymore, since the only interest that it serves, status, is already satisfied by high salary. So with a fixed set of issues as criteria, ceteris paribus or lexicographic models cannot represent every preference order. In many cases, this can be solved intuitively by taking underlying interests into account.

There are other approaches to deal with this matter. Instead of assuming independence of the criteria, one can also model conditional preferences, where criteria may be dependent on other criteria. A well-known approach to represent conditional preferences is CP-nets [16], which is short for conditional ceteris paribus preference networks. A CP-net is a graph where the nodes are variables (comparable to our notion of issues). Every node is annotated with a conditional preference table, which lists a user's preferences over the possible values of that variable. If such preferences are conditional (dependent on other variables), each condition has a separate entry in the table, and the variables that influence the preference are parent nodes of this variable in the graph. In [16], an example of conditional preference is given regarding an evening dress. A man unconditionally prefers black to white as a colour for both the jacket and the pants. His preference between a white and a red shirt is conditioned on the combination of jacket and pants. If they have the same colour, he prefers a red shirt (for a white shirt will make his outfit too colourless). If they are of different colours, he prefers a white shirt (because a red shirt will make his outfit too flashy). The complete assignments (outcomes in our terminology) are listed in Table 2a. The preference graph induced by the CP-net for this example is displayed in Figure 2a.

We propose to replace the variables the preferences over which are conditional with underlying interests – the reason for the dependency. In the evening dress example, the underlying interest is that the colours of jacket, pants and shirt make a good

Table 3. The knowledge base for the example

$highsal(c)$	$I_M(wealth)$	$highsal(x) \Rightarrow wealth(x)$
$\neg highpos(c)$	$I_M(status)$	$highsal(x) \Rightarrow status(x)$
$full\text{-}time(c)$	$I_M(family)$	$highpos(x) \Rightarrow status(x)$
$\neg highsal(f)$		$\neg full\text{-}time(x) \Rightarrow family(x)$
$highpos(f)$	$I_J(manager)$	$highpos(x) \Rightarrow manager(x)$
$\neg full\text{-}time(f)$	$I_J(cutback)$	$\neg highsal(x) \Rightarrow cutback(x)$

combination, which in this case is defined by being neither too colourless nor too flashy. The satisfaction of this interest by the different outcomes is listed in Table 2b. The variables jacket and pants are unconditional, so they can remain as criteria. If we take jacket, pants, and good combination as criteria, we can construct the preference graph in Figure 2b, using the ceteris paribus principle. The difference with the preferences induced by the CP-net is that in the CP-net case, outcome i is more preferred than k and m, and p is less preferred than l and n, while in the interest-based case they are incomparable. This is due to the fact that in CP-nets, conditional preferences are implicitly considered less important than the preferences on the variables they depend on ([16], p. 145). In fact, if we would specify that both jacket and pants are more important than a good combination, our preference ordering would be the same as in Figure 2a. But the interest approach is more flexible; it is possible to specify any (partial) importance ordering on interests. For example, we could also state that a good combination is more important than either the jacket or the pants, which results in the preference ordering in Figure 2c. In our opinion, there is no a priori reason to attach more importance to unconditional variables as is done in the CP-net approach.

6 Argumentation Framework

In this section, we present an argumentation framework (AF) for reasoning about qualitative, interest-based preferences. An abstract AF in the sense of Dung [11] is a pair $\langle \mathcal{A}, \rightarrow \rangle$ where $\mathcal{A}$ is a set of arguments and $\rightarrow$ is a defeat relation (informally, a counterargument relation) among those arguments. To define which arguments are justified, we use Dung's [11] preferred semantics.

Definition 2. (Preferred Semantics) . A *preferred extension* of an AF $\langle \mathcal{A}, \rightarrow \rangle$ is a maximal (w.r.t. $\subseteq$) set $S \subseteq \mathcal{A}$ such that: $\forall A, B \in S : A \not\rightarrow B$ and $\forall A \in S$: if $B \rightarrow A$ then $\exists C \in S : C \rightarrow B$. An argument is credulously (sceptically) *justified* w.r.t. preferred semantics if it is in some (all) preferred extension(s).

Informally, a preferred extension is a coherent point of view that can be defended against all its attackers. In case of contradictory information, there will be multiple preferred extensions, each advocating one point of view. The contradictory conclusions will be credulously, but not sceptically justified.

We instantiate an abstract AF by specifying the structure of arguments and the defeat relation.

Table 4. Inference schemes

$$1\ \frac{L_1,\ldots,L_k,\sim L_l,\ldots,\sim L_m \Rightarrow L_n \quad L_1 \quad \ldots \quad L_k \quad \sim L_l \quad \ldots \quad \sim L_m}{L_n}\ \mathit{DMP}$$

$$2\ \frac{}{\sim L}\ \mathit{asm}(\sim L)$$

$$3\ \frac{L}{\mathit{asm}(\sim L)\text{ is inapplicable}}\ \mathit{asm}(\sim L)\mathit{uc}$$

$$4\ \frac{}{\mathit{sat}(a,[P]_\alpha,0)}\ \mathit{count}(a,[P]_\alpha,\varnothing)$$

$$5\ \frac{P_1(a) \quad \ldots \quad P_n(a) \quad P_1 \approx_\alpha \ldots \approx_\alpha P_n \quad I_\alpha(P_1) \quad \ldots \quad I_\alpha(P_n)}{\mathit{sat}(a,[P_1]_\alpha,n)}\ \mathit{count}(a,[P_1]_\alpha,\{P_1,\ldots,P_n\})$$

$$6\ \frac{P_1(a) \quad \ldots \quad P_n(a) \quad P_1 \approx_\alpha \ldots \approx_\alpha P_n \quad I_\alpha(P_1) \quad \ldots \quad I_\alpha(P_n)}{\mathit{count}(a,[P_1]_\alpha,S \subset \{P_1,\ldots,P_n\})\text{ is inapplicable}}\ \mathit{count}(a,[P_1]_\alpha,S)\mathit{uc}$$

$$7\ \frac{\mathit{sat}(a,[P]_\alpha,n) \quad \mathit{sat}(b,[P']_\alpha,m) \quad P \approx_\alpha P' \quad n > m}{\mathit{pref}_\alpha(a,b)}\ \mathit{prefinf}(a,b,[P]_\alpha)$$

$$8\ \frac{\mathit{sat}(a,[Q]_\alpha,n) \quad \mathit{sat}(b,[Q']_\alpha,m) \quad Q \approx_\alpha Q' \succ_\alpha P \quad n < m}{\mathit{prefinf}(a,b,[P]_\alpha)\text{ is inapplicable}}\ \mathit{prefinf}(a,b,[P]_\alpha)\mathit{uc}$$

$$9\ \frac{\mathit{sat}(a,[P]_\alpha,n) \quad \mathit{sat}(b,[P']_\alpha,m) \quad P \approx_\alpha P' \quad n = m}{\mathit{eqpref}_\alpha(a,b)}\ \mathit{eqprefinf}(a,b,[P]_\alpha)$$

$$10\ \frac{\mathit{sat}(a,[Q]_\alpha,n) \quad \mathit{sat}(b,[Q']_\alpha,m) \quad Q \approx_\alpha Q' \quad n \neq m}{\mathit{eqprefinf}(a,b,[P]_\alpha)\text{ is inapplicable}}\ \mathit{eqprefinf}(a,b,[P]_\alpha)\mathit{uc}$$

Arguments. Arguments are built from formulas of a logical language, that are chained together using inference steps. Every inference step consists of premises and a conclusion. Inferences can be chained by using the conclusion of one inference step as a premise in the following step. Thus a tree of chained inferences is created, which we use as the formal definition of an argument (cf. e.g. [17]).

Definition 3. (Argument). An *argument* is a tree, where the nodes are inferences, and an inference can be connected to a parent node if its conclusion is a premise of that node. Leaf nodes only have a conclusion (a formula from the knowledge base), and no premises. A subtree of an argument is also called a *subargument*. `inf` returns the last inference of an argument (the root node), and `conc` returns the conclusion of an argument, which is the same as the conclusion of the last inference.

Definition 4. (Language). Let $\mathcal{P}$ be a set of predicate names with typical elements P,Q; $\mathcal{O}$ a set of outcome names with typical elements a,b; α an audience; and n a non-negative integer. The *input language* $\mathcal{L}_{KB}$ and full *language* $\mathcal{L}$ are defined as follows.

Table 5. Example arguments

$$A\quad \dfrac{\dfrac{\dfrac{\mathit{highsal}(c)\quad \mathit{highsal}(x)\Rightarrow \mathit{wealth}(x)}{\mathit{wealth}(c)}\quad I_M(\mathit{wealth})}{\mathit{sat}(c,[\mathit{wealth}]_M,1)}\quad \dfrac{}{\mathit{sat}(f,[\mathit{wealth}]_M,0)}\quad \mathit{wealth}\approx_M \mathit{wealth}\quad 1>0}{\mathit{pref}_M(c,f)}\ \alpha$$

$$B\quad \dfrac{\dfrac{\dfrac{\neg \mathit{full\text{-}time}(f)\quad \neg \mathit{full\text{-}time}(x)\Rightarrow \mathit{family}(x)}{\mathit{family}(f)}\quad I_M(\mathit{family})}{\mathit{sat}(f,[\mathit{family}]_M,1)}\quad \dfrac{}{\mathit{sat}(c,[\mathit{family}]_M,0)}\quad \mathit{family}\approx_M \mathit{family}\quad 1>0}{\mathit{pref}_M(f,c)}\ \beta$$

$$C\quad \dfrac{\dfrac{\dfrac{\mathit{highsal}(c)\quad \mathit{highsal}(x)\Rightarrow \mathit{wealth}(x)}{\mathit{wealth}(c)}\quad I_M(\mathit{wealth})}{\mathit{sat}(c,[\mathit{wealth}]_M,1)}\quad \dfrac{}{\mathit{sat}(f,[\mathit{wealth}]_M,0)}\quad \mathit{wealth}\succ_M \mathit{family}\quad 1\neq 0}{\beta \text{ is inapplicable}}$$

$$D\quad \dfrac{\dfrac{\dfrac{\neg \mathit{full\text{-}time}(f)\quad \neg \mathit{full\text{-}time}(x)\Rightarrow \mathit{family}(x)}{\mathit{family}(f)}\quad I_M(\mathit{family})}{\mathit{sat}(f,[\mathit{family}]_M,1)}\quad \dfrac{}{\mathit{sat}(c,[\mathit{family}]_M,0)}\quad \mathit{family}\succ_M \mathit{wealth}\quad 1\neq 0}{\alpha isinapplicable}$$

$$\varphi \in \mathcal{L}_{KB} ::= L \mid I_\alpha(P) \mid P \succ_\alpha Q \mid P \approx_\alpha Q \mid \\ L_1,\ldots,L_k, \sim L_l,\ldots,\sim L_m \Rightarrow L_n$$

where $L_i = P(a)$ or $\neg P(a)$.

$$\psi \in \mathcal{L} ::= \varphi \in \mathcal{L}_{KB} \mid\ \sim L \mid \mathit{sat}(a,[P]_\alpha,n) \mid \\ \mathit{pref}_\alpha(a,b) \mid \mathit{eqpref}_\alpha(a,b)$$

We make a distinction between an input and full language. A knowledge base, which is the input for an argumentation framework, is specified in the input language. The input language allows us to express facts about the criteria that outcomes (do not) satisfy, statements about interests of an audience and their importance ordering, and defeasible rules. The knowledge base for the job contract example (the facts restricted to outcomes c and f) is displayed in Table 3. Other formulas of the language that are not part of the input language, e.g. expressing a preference between two outcomes, can be derived from a knowledge base using inference steps that build up an argument (such formulas are not allowed in a knowledge base because they might contradict derived statements).

Inferences. Table 4 shows the inference schemes that are used. The first inference scheme is called defeasible modus ponens. It allows to infer conclusions from defeasible rules. The next two inference rules define the meaning of the weak negation $\sim$. According to inference rule 2, a formula $\sim \varphi$ can always be inferred, but such an argument will be defeated by an undercutter built with inference rule 3 if φ is the case. Inference schemes 4 and 5 are used to count the number of interests of equal importance (according to audience α) as some interest P_1 that outcome a satisfies. This type of inference is inspired by accrual [18], which combines multiple arguments with the same conclusion into one accrued argument for the same conclusion. Although our application is different, we use a similar mechanism. Inference scheme 4 can be used when an outcome satisfies no interests. It is possible to construct an argument that does not count all interests that are satisfied, a so-called non-maximal count. But we want all interests to be counted, otherwise we would conclude incorrect preferences. To ensure that only maximal counts are used, we provide an inference scheme to construct arguments that undercut non-maximal counts (inference scheme 6). An argument of this type says that any count which is not maximal is not applicable. Inference scheme 7 says that an outcome a is preferred over an outcome b if the number of interests of a certain importance level that a satisfies is higher than the number of interests on that same level that b satisfies. Inference scheme 8 undercuts scheme 7 if there is a more important level than that of P on which a and b do not satisfy the same number of interests. Finally, inference schemes 9 and 10 do the same as 7 and 8, but for equal preference.

Defeat. The most common type of defeat is rebuttal. An argument rebuts another argument if its conclusion contradicts conclusion of the other argument. Conclusions contradict each other if one is the negation of the other, or if they are preference or importance statements that are incompatible (e.g. $pref_\alpha(a,b)$ and $pref_\alpha(b,a)$, or $pref_\alpha(a,b)$ and $eqpref_\alpha(a,b)$). Defeat by rebuttal is mutual. Another type of defeat is undercut. An undercutter is an argument for the inapplicability of an inference used in another argument. Undercut works only one way. Defeat is defined recursively, which means that rebuttal can attack an argument on all its premises and (intermediate) conclusions, and undercut can attack it on all its inferences.

Definition 5. (Defeat) An argument A *defeats* an argument B ($A \rightarrow B$) if $\texttt{conc}(A)$ and $\texttt{conc}(B)$ are contradictory (*rebuttal*), or $\texttt{conc}(A) =$'$\texttt{inf}(B)$ is inapplicable' (*undercut*), or A defeats a subargument of B.

Let us return to the example. With the information from the knowledge base, the arguments A and B in Table 5 can be formed. A advocates a preference for c, based on the interest wealth. B advocates a preference for f, based on the interest family. Without an ordering on these interests, no decision between these arguments can be made. But if $wealth \succ_M family$ is known, argument C can be made, which undercuts B. Similarly, with $family \succ_M wealth$, argument D can be made, which undercuts A.

Validity. If some conditions in the input knowledge base (KB) hold, it can be shown that the proposed argumentation framework models ceteris paribus and lexicographic preference. In the following, we consider a single audience and leave out the subscript α.

Condition 1. Let $\mathcal{C}$ be a set of interests to be used as criteria, with importance order $\succeq$.
(1) For all P, '$I(P)$' is in KB iff $P \in \mathcal{C}$.
(2) For all $P \in \mathcal{C}$, a, '$P(a)$' is a conclusion of a sceptically justified argument iff a satisfies P.
(3) The relative importance among interests is
(a) a total preorder,
(b) the identity relation,
and for all P, $Q \in \mathcal{C}$, '$P \succ Q$' is in KB iff $P \succ Q$, and '$P \approx Q$' is in KB iff $P \approx Q$.

Theorem 1. (i) If conditions 1.1, 1.2 and 1.3a hold, then *pref*(a,b) (resp. *eqpref*(a,b)) is a sceptically justified conclusion of the argumentation framework iff a is strictly (resp. equally) preferred over b according to the lexicographic preference ordering.
(ii) If conditions 1.1, 1.2 and 1.3b hold, then *pref*(a,b) (resp. *eqpref*(a,b)) is a sceptically justified conclusion of the argumentation framework iff a is strictly (resp. equally) preferred over b according to the ceteris paribus preference ordering.

Proof. We prove the theorem for strict preference. The same line of argument can be followed for equal preference.
(i) $\Leftarrow$: Suppose a is strictly lexicographically preferred over b. This means that there is an importance level on which a satisfies more interests (say, $P_1,\ldots,P_n$) than b (say, $P'_1,\ldots,P'_m$, $n>m$), and on all more important levels, a and b satisfy an equal number of interests. In this case, we can construct the following arguments, where the first two arguments are subarguments of the third (note that these arguments can also be built if m is equal to 0, by using the empty set count).

$$\frac{P_1(a) \quad \ldots \quad P_n(a) \quad I(P_1) \quad \ldots \quad I(P_n) \quad P_1 \approx \ldots \approx P_n}{sat(a,[P_1],n)}$$

$$\frac{P'_1(b) \quad \ldots \quad P'_m(b) \quad I(P'_1) \quad \ldots \quad I(P'_m) \quad P'_1 \approx \ldots \approx P'_m}{sat(b,[P'_1],m)}$$

$$\frac{sat(a,[P_1],n) \quad sat(b,[P'_1],m) \quad P_1 \approx P'_1 \quad n>m}{pref(a,b)}$$

We will now try to defeat this argument. Premises of the type $P(a)$ are justified by condition 1.2. Premises of the type $I(P)$ and $P_1 \approx P_2$ cannot be defeated (conditions 1.1 and 1.3a). There are three inferences we can try to undercut (the last inference of the argument and the last inferences of two subarguments). For the first count, this can only be done if there is another P_j such that $I(P_j)$ and $P_j \approx P$ and $P_j \notin \{P_1,\ldots,P_n\}$ and $P_j(a)$ is the case. However, $P_1 \ldots P_n$ encompass all interests that a satisfies on this level, so count undercut is not possible. The same argument holds for the other count. At this point it is useful to note that these two counts are the only ones that are undefeated. Any lesser count will be undercut by the count undercutter that takes all of $P_1 \ldots P_n$ (resp. $P'_1 \ldots P'_m$) into account. Such an undercutter has no defeaters, so any non-maximal count is not justified. The undercutter of *prefinf*$(a,b,[P_1])$ is based on two counts. We have seen that any non-maximal count will be undercut. If the maximal counts are used, we have $n=m$ for undercutter arguments that use $Q \succ P$, since we have that on all more important levels than $[P_1]$, a and b satisfy an equal number of interests. So the undercutter inference rule cannot be applied since $n \neq m$ is not true. For that reason, a rebutting argument with conclusion *pref*(b,a) will not be justified. This means that for every possible type of defeat, either the defeat is inapplicable or the defeater is itself defeated by undefeated arguments. This means that the argument is sceptically justified.
$\Rightarrow$: Suppose that a is not strictly lexicographically preferred over b. This means that for all importance levels $[P]$, either a does not satisfy more interests than b on that

level, or there exists a more important level where b satisfies more interests than a. This means that any argument with conclusion $\mathit{pref}(a,b)$ (which has to be of the form above) is either undercut by $\mathit{count}(b,[P],S)\mathit{uc}$ because it uses a non-maximal count, or by $\mathit{prefinf}(a,b,[P])\mathit{uc}$ because there is a more important level where a preference for b over a can be derived. This means that any such argument will not be sceptically justified.
(ii) $\Leftarrow$: Suppose a is strictly ceteris paribus preferred over b. This means that there is (at least) one interest, let us say P, that a satisfies and b does not, and there are no interests that b satisfies and a does not. In this case, we can construct the following argument.

$$\frac{\dfrac{P(a) \quad I(P)}{\mathit{sat}(a,[P],1)} \quad \dfrac{}{\mathit{sat}(b,[P],0)} \quad P \approx P \quad 1>0}{\mathit{pref}(a,b)}$$

Premise $P(a)$ is justified by condition 1.2. Premise $I(P)$ cannot be defeated (condition 1.1). Note that, since there is no importance ordering specified, counts can only include 0 or 1 interest(s). So the first count cannot be undercut, because there are no other interests that are equally important as P (condition 1.3b). The second count cannot be undercut because b does not satisfy P. Since there are no interests that b satisfies but a does not, the last inference can only be undercut by an undercutter that uses a non-maximal count and so will be undercut itself.
$\Rightarrow$: Suppose a is not strictly ceteris paribus preferred over b. This means that either there is no interest that a satisfies but b does not, or there is some interest that b satisfies and a does not. In the first case, the only arguments that derive a preference for a over b have to use non-maximal counts and hence are undercut. In the second case, any argument that derives a preference for a over b is rebut by the following argument,

$$\frac{\dfrac{Q(b) \quad I(Q)}{\mathit{sat}(b,[Q],1)} \quad \dfrac{}{\mathit{sat}(a,[Q],0)} \quad Q \approx Q \quad 1>0}{\mathit{pref}(b,a)}$$

and is not sceptically justified. □

7 Conclusions

In this paper we have made a case for explicitly modelling underlying interests when reasoning about preferences in the context of practical reasoning. We have presented an argumentation framework for reasoning about qualitative interest-based preferences that models ceteris paribus and lexicographic preference.

In the current framework, we have only considered Boolean issues and interests. While this suffices to illustrate the main points discussed in this paper, multi-valued scales would be more realistic. Such an approach would open the way to modelling different degrees of (dis)satisfaction of an interest. For example, [5] take into account the level of satisfaction of goals on a bipolar scale. In the Boolean case, the lexicographic preference ordering is based on counting the number of interests that are satisfied by outcomes. This is no longer possible if multi-valued scales are used. In that case, we could count interests that are satisfied to a certain degree (like e.g. [5]), or compare outcomes in a pairwise fashion and count the number of interests that one outcome satisfies to a higher degree than another (like e.g. [7,13]).

Currently, we suppose that the interests and importance ordering among them are given in a knowledge base. We can make our framework more flexible by allowing such statements to be derived in a way that is similar to the derivation of statements about the satisfaction of interests.

We would also like to look into the interplay between different issues promoting or demoting the same interest. For example, a high salary and a high position both lead to status, but together they may lead to even more status. Or a low salary may promote cutback, but providing a lease car will demote it. Do these effects cancel each other out? The principles that play a role here are related to the questions posed in the context of accrual of arguments [18].

Since our long-term goal is the development of an automated negotiation support system, we plan to look into negotiation strategies that are based on qualitative, interest-based preferences as described here, as opposed to utility-based approaches currently in use. For the same reason, we plan to implement the argumentation framework for reasoning about interest-based preferences that we have presented here. Another interesting question in this context is how interest-based preferences can be elicited from a human user.

Acknowledgements. We thank Henry Prakken for useful comments on earlier drafts of this paper. This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs. It is part of the Pocket Negotiator project with grant number VICI-project 08075.

References

1. Keeney, R.L., Raiffa, H.: Decisions with multiple objectives: preferences and value tradeoffs. Cambridge University Press (1993)
2. Brewka, G.: A rank based description language for qualitative preferences. In: 16th European Conference on Artificial Intelligence (ECAI 2004), pp. 303–307 (2004)
3. Keeney, R.L.: Value-Focused Thinking: A Path to Creative Decisionmaking. Harvard University Press (1992)
4. Rahwan, I., Pasquier, P., Sonenberg, L., Dignum, F.: On the benefits of exploiting underlying goals in argument-based negotiation. In: 22nd Conference on Artificial Intelligence (AAAI 2007), pp. 116–121 (2007)
5. Amgoud, L., Bonnefon, J.F., Prade, H.: An Argumentation-Based Approach to Multiple Criteria Decision. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 269–280. Springer, Heidelberg (2005)
6. Amgoud, L., Prade, H.: Using arguments for making and explaining decisions. Artificial Intelligence 173(3-4), 413–436 (2009)
7. Ouerdane, W., Maudet, N., Tsoukiàs, A.: Argument schemes and critical questions for decision aiding process. In: Besnard, P., Doutre, S., Hunter, A. (eds.) Computational Models of Argument (COMMA 2008). Frontiers in Artificial Intelligence and Applications, pp. 285–296. IOS Press (2008)

8. Ouerdane, W., Maudet, N., Tsoukiàs, A.: Argumentation theory and decision aiding. In: Ehrgott, M., Figueira, J.R., Greco, S. (eds.) New Trends in Multiple Criteria Decision Analysis. Springer, Heidelberg (2010)
9. Bench-Capon, T.J.M.: Persuasion in practical argument using value based argumentation frameworks. Journal of Logic and Computation 13(3), 429–448 (2003)
10. Bench-Capon, T., Atkinson, K.: Abstract argumentation and values. In: Rahwan, I., Simari, G.R. (eds.) Argumentation in Artificial Intelligence, pp. 45–64. Springer, Heidelberg (2009)
11. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. Artificial Intelligence 77, 321–357 (1995)
12. Kaci, S., van der Torre, L.: Preference-based argumentation: Arguments supporting multiple values. International Journal of Approximate Reasoning 48(3), 730–751 (2008)
13. Van der Weide, T., Dignum, F., Meyer, J.J., Prakken, H., Vreeswijk, G.: Practical Reasoning using Values: Giving Meaning to Values. In: McBurney, P., Rahwan, I., Parsons, S., Maudet, N. (eds.) ArgMAS 2009. LNCS, vol. 6057, pp. 79–93. Springer, Heidelberg (2010)
14. Von Wright, G.H.: The Logic of Preference: An Essay. Edinburgh University Press (1963)
15. Wellman, M.P., Doyle, J.: Preferential semantics for goals. In: 9th National Conference on Artificial Intelligence (AAAI 1991), pp. 698–703 (1991)
16. Boutilier, C., Brafman, R.I., Domshlak, C., Hoos, H.H., Poole, D.: CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. Journal of Artificial Intelligence Research 21, 135–191 (2004)
17. Vreeswijk, G.A.W.: Abstract argumentation systems. Artificial Intelligence 90(1-2), 225–279 (1997)
18. Prakken, H.: A study of accrual of arguments, with applications to evidential reasoning. In: 10th International Conference on Artificial Intelligence and Law (ICAIL 2005), pp. 85–94 (2005)