Representing Preferences Among Sets

Gerhard Brewka
University of Leipzig
Department of Computer Science
Augustusplatz 10-11
D-04109 Leipzig, Germany
brewka@informatik.uni-leipzig.de

Mirosław Truszczyński
University of Kentucky
Depertment of Computer Science
Lexington, KY 40506-0046, USA
mirek@cs.uky.edu

Stefan Woltran
Vienna University of Technology
Institute of Information Systems
Favoritenstr. 9-11
A-1040 Vienna, Austria
woltran@dbai.tuwien.ac.at

Abstract

We study methods to specify preferences among subsets of a set (a *universe*). The methods we focus on are of two types. The first one assumes the universe comes with a preference relation on its elements and attempts to *lift* that relation to subsets of the universe. That approach has limited expressivity but results in orderings that capture interesting general preference principles. The second method consists of developing formalisms allowing the user to specify "atomic" improvements, and generating from them preferences on the powerset of the universe. We show that the particular formalism we propose is expressive enough to capture the lifted preference relations of the first approach, and generalizes propositional CP-nets. We discuss the importance of domain-independent methods for specifying preferences on sets for knowledge representation formalisms, selecting the formalism of argumentation frameworks as an illustrative example.

Introduction

Preferences play a fundamental role in many AI applications. They are needed whenever agents have to choose among alternatives in decision making, or when they have to select a coherent set of beliefs based on possibly conflicting information. For this reason, languages for representing and reasoning with preferences have been developed. One of the best-known examples are CP-nets (Boutilier et al. 2004).

Our interest in this paper lies in preferences among *sets*. Given a *finite* universe A, we want to be able to express succinctly preferences on the powerset 2^A of A, and use them to select from some space of subsets of A an optimal one. The problem arises in a variety of contexts. Under the assumption that objects are "attributed" (are tuples of attribute/feature values) it was considered by desJardins and Wagstaff (2005) and by the CP-net community (Brafman et al. 2006). However, in many important contexts where that problem needs to be addressed, there is no such structure to assume.

Thus, we are interested in the problem of preference among sets in a more abstract setting, when individual objects are "atomic", but can possibly be related to each other by one or more relations. A well-studied area where such preferences arise is nonmonotonic reasoning and, in particular, answer-set programming (Brewka, Niemelä, and Truszczyński 2009); another one concerns monotonic preferences over sets of goods (Bouveret, Endriss, and Lang 2009).

The area that we will use as an example is *argumentation*. In argumentation one commonly uses *argumentation frameworks* (AFs) to represent available arguments and conflicts among them (Dung 1995). Dung defined several semantics (e.g., stable, preferred and others), assigning to an AF a collection of acceptable sets of arguments, called *extensions*. Preferences can serve as an important mechanism for selecting those extensions which are most desirable, say, have a strongest support. Thus, they have received considerable attention (Amgoud and Cayrol 1998; Bench-Capon 2002; Modgil 2009). However, existing approaches are not fully satisfactory as they are all based on a modification of the original AF. As demonstrated in (Dimopoulos, Moraitis, and Amgoud 2009; Amgoud and Vesic 2009) this can lead to unintended results and new methods are called for. We argue in the paper that the solution is to *select* among the extensions rather than to modify the AF. However, in order to select, adequately represented preferences on subsets of arguments are needed.

Thus, we investigate first domain-independent ways to represent preferences among sets. Due to the combinatorial nature of the problem stating preferences among the subsets directly is out of the question. One possibility is to start from a preference ordering on A (when such an ordering is given), and to lift it to the ordering on 2^A. The first part of this paper investigates this approach. We consider the well-known orderings due to Hoare and Smyth, and propose modifications, which appear better suited for applications in AI.

We then move on to a second approach. We propose a general language to specify preference orderings on 2^A based on incremental improvements. The idea is to use simple preference rules specifying elementary modifications which, under certain conditions, transform the original set to a new one that is at least as good. A set $E \subseteq A$ is then at least as good as a set $E' \subseteq A$ if there is a sequence of elementary modifications leading from E' to E.

This approach shares motivation and philosophy with CP-nets, and with their extensions (Wilson 2004; Brafman, Domshlak, and Shimony 2006). First, there is a precise correspondence between objects of interest — subsets of A and configurations on A (a subset S of A can be viewed as a configuration in a CP-net[1]). Second, our sequences of incremental improvements correspond closely to sequences of improving flips in CP-nets. However, our language goes significantly beyond the formalism of CP-nets in several important aspects. It includes an explicit operation of *trading* an element for another better one (not available in CP-nets) and the possibility of conditioning modifications on properties of elements *gained*, *lost* and *traded* so far (taking into account the entire improving sequence generated so far and not the last state only, as CP-nets do). While both features can be simulated in CP-nets, the encodings are cumbersome and require significant extensions of the vocabulary of a CP-net.

Most importantly though, our formalism has a *predicate* variant supporting the use of variables ranging over elements in A, which allows conditions to refer to properties of elements expressed in terms of relations the set A may be endowed with (a fundamental extension over CP-net languages).

Preferences on Subsets of Ordered Sets

Explicit specifications of preference relations on 2^A are infeasible as the cardinality of 2^A is exponential in the cardinality of A. One workaround is to define preferences on 2^A indirectly in terms of properties of elements of A. We discuss it now under the assumption that A comes endowed with a preference relation. The goal is to lift that relation to 2^A.

A most common type of a preference relation is a *preorder*, a relation that is reflexive and transitive. Thus, let $\geq$ be a preorder on A. Two well-known orderings lifting $\geq$ to 2^A are:

1. (*Hoare*) $E_1 \succeq^h E_2$ if for every $y \in E_2$ there is $x \in E_1$ such that $x \geq y$.
2. (*Smyth*) $E_1 \succeq^s E_2$ if for every $x \in E_1$ there is $y \in E_2$ such that $x \geq y$.

The relations $\succeq^h$ and $\succeq^s$ are preorders.[2] We note that if a set E contains low-quality elements, it still may be of high quality wrt $\succeq^h$, as long as it also contains some high-quality elements. The situation is different for $\succeq^s$. Low-quality elements make a set to be of low quality. Moreover, just adding some high-quality elements does not necessarily improve the quality of the set. In other words, the Hoare preorder is determined by the *stars*, the Smyth preorder by the *weakest* links.

While Hoare and Smyth orderings capture some natural criteria for comparing subsets of a set endowed with a preorder, they have drawbacks as preference preorders. First, if $E_1 \subseteq E_2$ then, by the reflexivity of $\geq$, $E_1 \succeq^s E_2$ ($E_2 \succeq^h E_1$, respectively). In other words, the Hoare (Smyth) preorder necessarily extends the superset (subset) relation, leaving the user with no say as to whether supersets (subsets) are to be preferred. Second, when comparing two sets, both preorders take into account elements that are *common* to the two sets. This does not seem intuitive. When deciding which of the two sets to prefer, the elements that distinguish them (those which belong to exactly one of the two sets) should be especially important (for a simple example, in many card games having the ace and the king in the same suit is better than having the ace and the queen).

Thus, we will now propose two variants of the basic relations $\succeq^h$ and $\succeq^s$ that aim to address these issues, while preserving the essence of the Hoare and Smyth principles. First, for every $E_1, E_2 \subseteq A$, we set:

1. $E_1 \succeq^h_1 E_2$ if for every $y \in E_2 \setminus E_1$ there is $x \in E_1 \setminus E_2$ such that $x \geq y$.
2. $E_1 \succeq^s_1 E_2$ if for every $x \in E_1 \setminus E_2$ there is $y \in E_2 \setminus E_1$ such that $x \geq y$.

The relations $\succeq^h_1$ and $\succeq^s_1$ depend only on the elements that distinguish the sets under comparison. However, it still holds that $\succeq^h_1$ extends $\supseteq$, and $\succeq^s_1$ extends $\subseteq$. To address this issue, given $E_1, E_2 \subseteq A$, we let:

1. $E_1 \succeq^h_2 E_2$ if $E_1 = E_2$, or if $E_2 \setminus E_1 \neq \emptyset$ and for every $y \in E_2 \setminus E_1$ there is $x \in E_1 \setminus E_2$ such that $x \geq y$.
2. $E_1 \succeq^s_2 E_2$ if $E_1 = E_2$, or if $E_1 \setminus E_2 \neq \emptyset$ and for every $x \in E_1 \setminus E_2$ there is $y \in E_2 \setminus E_1$ such that $x \geq y$.

As with $\succeq^h_1$ and $\succeq^s_1$, when two sets are compared under $\succeq^h_2$ and $\succeq^s_2$, elements common to both sets play no role. In addition, $\succeq^h_2$ and $\succeq^s_2$ no longer contain the relations $\supseteq$ and $\subseteq$, respectively. Moreover, while for both $\succeq^h_2$ and $\succeq^s_2$, $\emptyset$ and A are incomparable to any other subset of A and so, neither one is entirely free of $\geq$-independent ramifications, the ramifications concern special comparisons only that can easily be addressed separately.

Since, in general, the two sets of relations we introduced are not transitive, they are not preorders. Thus, we use their transitive closures as preference orderings.

Definition 1 *The first and second* marginal *Hoare and Smyth preorders $\succeq^{h,*}_1$ and $\succeq^{s,*}_1$, and $\succeq^{h,*}_2$ and $\succeq^{s,*}_2$ are the transitive closures of $\succeq^h_1$ and $\succeq^s_1$, and $\succeq^h_2$ and $\succeq^s_2$, respectively.*

The properties of the relations $\succeq^h_i$ and $\succeq^s_i$, $i = 1, 2$, which we discussed above, suggest that the marginal preorders they determine (especially, the second ones) are better suited as preference relations than the original Hoare and Smyth preorders.

There are other similar ways to lift a preorder $\geq$ on A to the powerset of A (we do not discuss them due to space limits). In each case, they capture some general principle by which two sets could be compared. While the principles are often quite intuitive, ultimately the set of choices is limited. And this is where the main shortcoming of this approach is. The class of preference relations that can be represented that way is narrow!

[1]Everywhere in the paper, whenever we refer to CP-nets, we mean *generalized CP-nets* as studied in (Goldsmith et al. 2008), which are binary.

[2]Sometimes, the *Plotkin* preorder, the intersection of $\succeq^h$ and $\succeq^s$, is also of interest. Much of our discussion can be extended to the Plotkin order, too. Hoare, Smyth and Plotkin preorders were studied in many contexts, for instance, in the area of possibilistic logic (Benferhat, Lagrue, and Papini 2004).

The Incremental-Improvement Approach

Due to limitations of the method discussed above, a more flexible and expressive way of specifying preferences is needed. We will now present such a method based on *elementary modifications*. We refer to it as the *incremental-improvement* approach (IIA).

The idea, first proposed as the basis of the CP-net formalism, is that the user specifies situations when a small change (or "elementary modification") to an object yields an object that is at least as good. In our case, objects are sets $E \subseteq A$ and we consider elementary modifications of the following three types:

Gaining an element $a \in A$, written $g(a)$. If $a \notin E$, the result of $g(a)$ applied to E is $E \cup \{a\}$ (otherwise, $g(a)$ does not apply).

Losing an element $a \in A$, written $l(a)$. If $a \in E$, the result of $l(a)$ applied to E is $E \setminus \{a\}$ (otherwise, $l(a)$ does not apply).

Trading an element $a \in A$ for another element $b \in A$, written $t(a, b)$. If $a \in E$ and $b \notin E$, the result of $t(a, b)$ applied to E is $(E \setminus \{a\}) \cup \{b\}$ (otherwise, $t(a, b)$ does not apply).

To define a preference relation on 2^A, one might select a set of elementary modifications, each of which once applied results in a set that is deemed at least as good.

Definition 2 *Given a set $\mathcal{I}$ of elementary modifications (*incremental *improvements), a sequence of sets $s = \langle E_0, \ldots, E_k \rangle$ is a* basic improving sequence *wrt $\mathcal{I}$ if for every i, $0 \le i \le k-1$, there is an operation $O \in \mathcal{I}$ applicable to E_i whose application results in E_{i+1}. We define $E \succeq_{\mathcal{I}} E'$ if there is a basic improving sequence wrt $\mathcal{I}$ that starts in E' and ends in E.*

Clearly, $\succeq_{\mathcal{I}}$ is reflexive and transitive. Thus, it is a preorder.

Defining preorders on sets in terms of incremental improvements, while capable of capturing some interesting preorders, is not expressive enough yet as it does not allow the user to specify *conditional* preferences. If $g(a) \in \mathcal{I}$, then for every E such that $a \notin E$, $E \cup \{a\}$ is preferred to E. What if we only wanted that to be the case if $b \notin E$, or if when constructing an improving sequence we already lost c?

To address that problem, we define (propositional) *conditional preference statements* (or simply, preference statements) to be expressions $\epsilon : \varphi$, where ϵ is an elementary modification, and φ is a *condition*. An intuitive reading of $\epsilon : \varphi$ is: if φ holds, ϵ is an incremental improvement, that is, if applicable and applied, it results in a set that is at least as good. Formally, we define a condition to be a propositional formula over atoms of the form $in(a)$, $g(a)$, $l(a)$, $t(a, b)$, for $a, b \in A$.

Let α be an atom and $s = \langle E_0, \ldots, E_k \rangle$ be a sequence of sets such that each E_{i+1} is the result of an elementary modification applied to E_i (this modification is determined by E_i and E_{i+1}). We define the satisfaction relation $s \models \alpha$ as follows (everywhere below $a, b \in A$):

1. $s \models in(a)$ if $a \in E_k$
2. $s \models g(a)$ ($l(a), t(a, b)$, respectively), if a was gained (lost, traded, respectively) before, that is, if for some $i < k$, $E_{i+1} = E_i \cup \{a\}$ ($E_{i+1} = E_i \setminus \{a\}$, $E_{i+1} = (E_i \setminus \{a\}) \cup \{b\}$, respectively)

The relation extends in the standard way to arbitrary formulas in the language.[3] When φ is a tautology, we simplify the notation and write ϵ instead of $\epsilon : \varphi$.

Next, we generalize the notion of an improving sequence.

Definition 3 *Let $\mathcal{I}$ be a set of preference statements. A sequence $\langle E_0, \ldots, E_k \rangle$ of subsets of A is an* improving sequence wrt $\mathcal{I}$ *if for every i, $0 \le i < k$, there is a preference statement $\epsilon : \varphi$ in $\mathcal{I}$ such that $\langle E_0, \ldots, E_i \rangle \models \varphi$, ϵ is applicable to E_i, and the result of the application is E_{i+1}.*

If $\langle E_0, \ldots, E_k \rangle$ and $\langle E_k, \ldots, E_m \rangle$ are improving sequences, it may be that $\langle E_0, \ldots, E_k, E_{k+1}, \ldots, E_m \rangle$ is not an improving sequence (in the combined sequence, the "history" for each E_t, $k \le t \le m$, changes). Thus, the preference relation one could obtain on the basis of improving sequences is not transitive. Therefore, we define $\succeq^*_{\mathcal{I}}$ as the transitive closure of that relation. Specifically, we say that $E \succeq^*_{\mathcal{I}} E'$ if there is a sequence of sets $\langle E_0, \ldots E_k \rangle$ such that $E' = E_0$, $E = E_k$ and for every i, $0 \le i < k$, there is an improving sequence starting in E_i and ending in E_{i+1}. It is clear that $\succeq^*_{\mathcal{I}}$ is a preorder.

However, in many cases, concatenating improving sequences does result in an improving sequence.

Proposition 1 *Let $\mathcal{I}$ be a set of preference statements over a set A whose conditions contain only atoms of the form $in(a)$, where $a \in A$. Then, $\succeq_{\mathcal{I}} = \succeq^*_{\mathcal{I}}$.*

Representing Orderings in IIA

All orders discussed here so far can be expressed in the incremental-improvement approach. We start with the basic Hoare and Smyth preorders. To this end, we let

$$\mathcal{I}_h = \{g(a) \mid a \in A\} \cup \{l(b) : in(a) \mid a, b \in A,\ a > b\}$$

$$\mathcal{I}_s = \{l(a) \mid a \in A\} \cup \{g(b) : in(a) \mid a, b \in A,\ b > a\}$$

(we write $a > b$ when $a \ge b$ and $b \not\ge a$). We note that preference statements in $\mathcal{I}_h$ and $\mathcal{I}_s$ contain atoms $in(a)$, $a \in A$, only. Thus there is no need to distinguish between $E \succeq_{\mathcal{I}} E'$ and $E \succeq^*_{\mathcal{I}} E'$.

Theorem 1 *Let $E, E' \subseteq A$. Then, $E \succeq^h E'$ iff $E \succeq_{\mathcal{I}_h} E'$, and $E \succeq^s E'$ iff $E \succeq_{\mathcal{I}_s} E'$.*

The marginal Hoare and Smyth orderings of both types can also be characterized in a similar fashion. We state the results for the Hoare orderings only (the corresponding Smyth orderings have dual characterizations). In some preference statements given below we write $t(\_, \_)$ as an abbreviation for the disjunction of all atoms of the form $t(a, b)$, where $a, b \in A$, and $t(\_, a)$ as an abbreviation for the disjunction of all atoms of the form $t(b, a)$, where $b \in A$.

The characterizations are given by the sets $\mathcal{MH}_1 = \{g(a) \mid a \in A\} \cup \mathcal{U}$ and $\mathcal{MH}_2 = \{g(a) : t(\_, \_) \mid a \in A\} \cup \mathcal{U}$, where $\mathcal{U}$ is the union of:

[3]The class of conditions can be extended to take into account the order in which elements are gained, lost or traded. Due to lack of space, we do not consider this generalization here.

1. $\{l(b) : t(\_, a) \vee g(a) \mid a, b \in A,\ a > b\}$, and
2. $\{t(a, b) \mid a, b \in A,\ b > a\}$.

Theorem 2 *Let $E, E' \subseteq A$. For $i = 1, 2$, $E \succeq_i^h E'$ iff $E \succeq_{\mathcal{MH}_i} E'$ (and so, $E \succeq_i^{h,*} E'$ iff $E \succeq_{\mathcal{MH}_i}^* E'$).*

The sets of preference statements used in the last two characterizations exploit atoms of the form $g(a), l(a)$ and $t(a, b)$ that are not available in the CP-net formalism. One needs to extend A with additional elements to model them in order to construct CP-nets characterizing marginal Hoare and Smyth orderings. We discuss this technique in the next section, when reducing the incremental-improvement approach to CP-nets.

Expressive Power of IIA

We show now that CP-nets can be expressed within IIA in a straightforward way and that, conversely, CP-nets can simulate IIA, but at the cost of a more complicated rewriting and additional CP-net variables.

Under a fixed enumeration of $A = \{a_1, \ldots, a_n\}$, there is a precise correspondence between subsets of A and binary outcomes $\alpha = \langle \alpha^1, \ldots, \alpha^n \rangle$, where each α^i is either a (meaning "a is in the set") or $\overline{a}$ ("a is not in the set"). *CP-statements* (over A) are expressions $a : \varphi$ and $\overline{a} : \varphi$, where $a \in A$ and φ is a formula (condition) built of elements in A (regarded as propositional atoms) that does not contain a. We note that outcomes can be viewed as propositional interpretations of conditions φ.

Following Goldsmith et al. (2008), a *CP-net over A* is a collection N of CP-statements over A. An outcome β results from an outcome α by means of an *improving flip* (wrt N) if there is j such that α and β differ only on the position j, and for some $\epsilon : \varphi$ in N, φ holds in α and either (1) $\alpha^j = a_j$ and $\epsilon = \overline{a}_j$, or (2) $\alpha^j = \overline{a}_j$ and $\epsilon = a_j$. A sequence of outcomes $\langle \alpha_0, \ldots, \alpha_k \rangle$ is *improving* if for $i = 0, \ldots, k-1$, α_{i+1} results from α_i by means of an improving flip. For outcomes α and β, we define $\alpha \succeq_N \beta$ if there is an improving sequence starting in β and ending in α. If α and β correspond to sets E_α and E_β, we will also write $E_\alpha \succeq_N E_\beta$.

A CP-statement $a : \varphi$ can be expressed as a preference statement $g(a) : in(\varphi)$ and, likewise, $\overline{a} : \varphi$ can be expressed as $l(a) : in(\varphi)$, where $in(\varphi)$ denotes the formula obtained from φ by replacing each occurrence of $b \in A$ with $in(b)$. Given a CP-net N, we denote by $\mathcal{I}_N$ the set of preference statements obtained by translating, as described above, all CP-statements in N.

Theorem 3 *Let N be a CP-net over A. Then, for all sets $E_1, E_2 \subseteq A$, $E_1 \succeq_N E_2$ iff $E_1 \succeq_{\mathcal{I}_N} E_2$.*

We will now sketch the other direction. For a set $\mathcal{I}$ of preference statements over A, we use CP-statements over atoms $in(a)$, $g(a)$, $l(a)$, $t(a, b)$ (where $a, b \in A$) and, additionally, new atoms $\gamma(a)$, $\lambda(a)$, $\theta(a, b)$ for $a, b \in A$. Atoms $\gamma(\cdot)$, $\lambda(\cdot)$, and $\theta(\cdot, \cdot)$ will be used to indicate that we currently perform a gain, loss, or trade, respectively.

We now construct a CP-net $N_\mathcal{I}$ as follows (as above, we use $\gamma(\_)$ and $\lambda(\_)$ as a shorthand for $\bigvee_{a \in A} \gamma(a)$, and respectively, $\bigvee_{a \in A} \lambda(a)$). For each trading statement $t(a, b) : \varphi$ in $\mathcal{I}$, we add the following statements to $N_\mathcal{I}$:

$$\theta(a, b) : in(a) \wedge \neg in(b) \wedge \varphi \wedge \neg\gamma(\_) \wedge \neg\lambda(\_) \wedge \neg \bigvee_{(c,d) \in A \times A \setminus \{(a,b)\}} \theta(c, d)$$

$$\overline{in(a)} : \theta(a, b); \quad in(b) : \theta(a, b); \quad t(a, b) : \theta(a, b)$$

$$\overline{\theta(a, b)} : \neg in(a) \wedge in(b) \wedge t(a, b)$$

Intuitively, the first CP-statement says that we can start the trading process *only if* a is in the current set, b is not, φ holds, and no other modification is underway (the remaining conjuncts). The start is marked by flipping $\theta(a, b)$ to true, which temporarily disallows any other modifications to commence. The next three CP-statements set $in(a)$ to false and $in(b)$ to true (execute the trade), and (if not done earlier) set $t(a, b)$ to true (record the trade). The last statement terminates the trading process and, by flipping $\theta(a, b)$ back to $\overline{\theta(a, b)}$ it allows execution of other modifications.

Statements $g(a) : \varphi$ and $l(a) : \varphi$ are reduced to CP-statements in $N_\mathcal{I}$ in a similar way.

Theorem 4 *Let $\mathcal{I}$ be a set of preference statements over A. Then, for all sets $E_1, E_2 \subseteq A$, $E_1 \succeq_\mathcal{I} E_2$ iff there is a set F_1, s.t. $F_1 \succeq_{N_\mathcal{I}} \{in(a) \mid a \in E_2\}$ and $\{a \mid in(a) \in F_1\} = E_1$.*

The extension of this result to the transitive closure $E \succeq_\mathcal{I}^* E'$ is also straightforward (we just need additional rules, to remove atoms $g(a)$, $l(b)$, $t(a, b)$ at any point).

The two reductions we presented here are constructible in polynomial time and, thus, clarify the complexity of our formalism. Specifically, many known complexity results concerning CP-nets (Goldsmith et al. 2008) carry over to IIA resulting in PSPACE-completeness for problems like dominance (decide whether $E \succeq_\mathcal{I} E'$) or consistency (check whether E can be obtained from itself using a non-empty improving sequence). However, we note that deciding $E \succeq_\mathcal{I} E'$ (dominance) restricted to sequences, in which the status of each element changes only once is in NP. The NP-hardness for that problem follows from this simple reduction from SAT to IIA. Let φ be a propositional formula over atoms V, w a new atom. Define $\mathcal{I} = \{g(a) \mid a \in A\} \cup \{g(w) : in(\varphi)\}$. Then, $A \cup \{w\} \succeq_\mathcal{I} \emptyset$ iff φ is satisfiable.

IIA — One More Generalization

So far, preference statements have been conditioned by properties of sets that can be expressed in terms of which elements they contain, and which elements they do not. In addition, we distinguish between the current set and the sets of elements that were gained and lost (directly or in trades). We do not have a way to condition preference statements with more complex properties of sets. In particular, if A has some structure captured by a set of relations on A, as of now we do not have a way to take advantage of that structure. For instance, if there is an equivalence relation s defined on r, a reasonable preference statement could say: add an element a, if a is not related wrt s to any other element in the current set (intuitively, sets containing representatives of more equivalence classes are better). For another example, if A

comes with a partial order $\geq$ (this is the case we focused on in the first part of the paper), a preference statement might be: trade any element X for any element Y that is better than any element in the current set (including X).

We will now describe a language in which such preference statements can be expressed.

We assume that A comes with a set R of relations on A. We will always assume that R contains the equality relation. To define properties of subsets of A, we will consider a language $\mathcal{L}$ of predicate calculus with the set of constants A and the set of relation symbols $\mathcal{R} \cup \{in, g, l, t\}$, where $\mathcal{R}$ is the set of names of the relations in R, in is a predicate symbol we will use to say that an element belongs to the current set, and the symbols g, l and t are designed to talk about elements that were gained, lost or traded.

Definition 4 A (general) preference statement *is an expression* $\epsilon : \varphi$*, where* ϵ *is a modification atom (that is, an atom built of a predicate symbol g, l or t) and* φ *is a formula in the language* $\mathcal{L}$ *such that every free variable of* φ *occurs in* ϵ.

The following are examples of general preference statements that capture those that we mentioned above.

1. $g(a) : \neg\exists X\ (in(X) \wedge (s(a, X) \vee s(X, a)))$. Intuitively, it is to mean that adding a that is not related by s to any element in the current set leads to a better set.
2. $t(X, Y) : \forall Z\ (in(Z) \rightarrow (Y \geq Z))$. Intuitively, it is to mean that trading X for Y results in a better set if Y is better than any element in the current set

To define a semantics of a set of preference statements $\mathcal{I}$, we first ground $\mathcal{I}$, that is, produce a set of preference statements $\mathcal{I}'$ with propositional modification atoms in the heads, by replacing variables in the heads with atoms from A. We regard $\mathcal{I}$ and $\mathcal{I}'$ as having the same meaning. We note that the condition in every preference statement in $\mathcal{I}'$ is a sentence of $\mathcal{L}$.

We will interpret preference statements in $\mathcal{I}'$ wrt the sequences $s = \langle E_0, \ldots, E_k \rangle$ of subsets of A such that each set E_{i+1} is the result of applying a modification atom to E_i (this modification atom is uniquely determined by E_i and E_{i+1}). Given such a sequence s, we define g_s (l_s, respectively) to be the set of elements that were gained (lost) at some step in the sequence s. We also define t_s to be a set of all pairs (a, b) such that a was traded for b at some step in s. Clearly, g_s and l_s are unary relations and t_s is a binary relation. Finally, we define $in_s = E_k$ (a unary relation).

We will use the structure $(A, R \cup \{g_s, l_s, t_s, in_s\})$ to interpret sentences of $\mathcal{L}$ wrt s. Namely, if φ is a sentence in $\mathcal{L}$, by $gr(\varphi)$ we mean a sentence obtained by replacing each existential (universal) quantifier by the disjunction (conjunction) over elements in A. We say that $s \models \varphi$ if $gr(\varphi)$ holds in $(A, R \cup \{g_s, l_s, t_s, in_s\})$. Next, we generalize the notion of an improving sequence.

Definition 5 *Let* $\mathcal{I}$ *be a set of preference statements.* $\langle E_0, \ldots, E_k \rangle$ *is an* improving sequence *wrt* $\mathcal{I}$ *if for every i,* $0 \leq i \leq k-1$*, there is a statement* $\epsilon : \varphi$ *in* $\mathcal{I}'$ *(i.e., grounded* $\mathcal{I}$*) such that* $\langle E_0, \ldots, E_i \rangle \models \varphi$*,* ϵ *is applicable to* E_i*, and the result of the application is* E_{i+1}.

As for propositional preference statements, preference relations obtained on the basis of improving sequences are not necessarily transitive, but we can define $\succeq^*_{\mathcal{I}}$ whenever appropriate. Finally, we note that if R is empty, ground preference statements are propositional preference statements as discussed earlier.

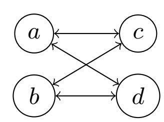
The predicate generalization of IIA brings up several interesting complexity questions that we will pursue in the future.

Applications — Argumentation

We assume some familiarity with abstract argumentation frameworks (AFs) and their semantics (Dung 1995). Approaches extending Dung style argumentation with preferences, such as (Amgoud and Cayrol 1998; Bench-Capon 2002; Modgil 2009), are based on the intuition that preferences on an AF $D = (A, att)$, can be handled by modifying D to a related AF D'. Given D and a strict partial order $>$ on A, Amgoud and Cayrol define D' as (A, att') where $att' = \{(a, b) \in att \mid b \not> a\}$. Optimal s-extensions of D, where s stands for one of the semantics developed for AFs (e.g. stable, preferred, grounded, semi-stable, ideal), are then defined as the s-extensions of D'. Bench-Capon follows a similar approach, yet derives the preference order $>$ on arguments from a more fundamental (total) preference order on values the arguments promote. Modgil's approach allows for dynamic deletion of links.

As pointed out in (Dimopoulos, Moraitis, and Amgoud 2009; Amgoud and Vesic 2009) this AF modification approach has serious problems. The following example was discussed by Dimopoulos et al.:

Example 1 *Consider the following AF D:*



Let $a > c$ *and* $b > d$*. Then, D has two stable extensions:* $\{a, b\}$ *and* $\{c, d\}$*. Given the specified preferences one would expect the former to be preferred over the latter. However, the modified framework* D' *has the same stable extensions as the original one.*

Other problems arise when the attack relation is asymmetric. The deletion of attacks based on preferences can then lead to situations where conflicting arguments are in the same extension (Amgoud and Vesic 2009).

Rather than modifying AFs, we propose a selection approach where optimal extensions are defined through a preorder on sets of arguments. We thus consider a *prioritized AF* (PAF) $D = (A, att, \geq_A)$ as an AF (A, att) equipped with a partial preorder $\geq_A$ on 2^A. We use Ext^D_s to denote the collection of s-extensions of (A, att).

Definition 6 *Let* $D = (A, att, \geq_A)$ *be a PAF. An s-extension E of D is* optimal *(*$E \in Opt^D_s$*) if E is maximal in* Ext^D_s *wrt* $\geq_A$.

Independently of the chosen preorder $\geq_A$, this approach exhibits the following arguably desirable properties:

1. $(a, b) \in att$ implies $a \notin P$ or $b \notin P$, for all $P \in Opt_s^D$.
2. $\geq_A = \emptyset$ implies $Opt_s^D = Ext_s$.
3. $Opt_s^D \subseteq Ext_s^D$.
4. $Ext_s^D \neq \emptyset$ implies $Opt_s^D \neq \emptyset$
5. $Ext_s^D = Ext_{s'}^D$ implies $Opt_s^D = Opt_{s'}^D$.
6. $Ext_s^D = Ext_s^{D'}$ implies $Opt_s^D = Opt_s^{D'}$.

(1) and a variant of (2) were discussed by Amgoud and Vesic (2009). (3) requires that no new extensions be generated. This is based on the view that the semantics defines the available choices, based on the attack relation. The preferences then are used to select among the available options. (4) is a consistency preservation principle. (5) states that the preference handling mechanism should be independent of the particular semantics chosen. Finally, (6) says that equivalent AFs should have the same preferred extensions.

This leaves us with two questions: how to represent $\geq_A$, and what are good choices for $\geq_A$?

The bulk of this paper was concerned with the first question. Any of the techniques discussed in this paper can be used, those lifting an ordering on A to an ordering on 2^A as well as those based on IIA. We stress that simply specifying an order on A (rather than an order on 2^A) does *not* provide enough information and leaves the preference handling method underspecified.

The second question is more difficult. Lifting a preorder on A to 2^A via the ordering $\succeq_2^{h,*}$ is promising. It solves the problem with Example 1 and works well in other examples we examined. However, it is unlikely that a single preorder can adequately handle all scenarios. For this reason argumentation is in need of simple and intuitive methods for specifying preferences among sets of arguments. We believe IIA is precisely such a method.

Conclusions

We studied two approaches for representing preferences among subsets of a set A, a problem that arises in various AI areas. The first approach uses, in the style of Hoare and Smyth, an ordering on A which is then lifted to sets. The second (IIA) is based on the idea to specify simple and intuitive conditional improvements. We have shown that IIA generalizes CP-nets. The other direction, namely representing IIA with CP-nets, is cumbersome and requires many additional variables. Representations of set preferences in IIA are much more concise. Moreover, IIA also comes in a predicate flavor, which offers additional modeling benefits and, in particular, can handle cases when A has additional structure.

Our future work will address in detail expressivity and computational issues. We are interested in identifying classes of preference statements with attractive computational behavior. We are especially interested in the computational complexity results for the predicate setting. We also want to study the relationship of our approach to preference logics, in particular those recently investigated in (Bienvenu, Lang, and Wilson 2010).

We will also continue to study application specific set orderings, in particular for answer set programming and argumentation. We believe that the ability provided by predicate IIA to take the structure of A into account will be crucial. In particular, in argumentation, it will allow us to refer to the attack relation in the specification of preferred sets of arguments, which is essential.

Acknowledgments

The first author was supported by *Deutsche Forschungsgemeinschaft* under grant Br 1817/3, the second author by the National Science Foundation under grant IIS-0913459, and the third author by *Wiener Wissenschafts-, Forschungs- und Technologiefonds* under grant ICT08-028.

References

Amgoud, L., and Cayrol, C. 1998. On the acceptability of arguments in preference-based argumentation. In *Proc. UAI'98*, 1–7.

Amgoud, L., and Vesic, S. 2009. Repairing preference-based argumentation frameworks. In *Proc. IJCAI'09*, 665–670.

Bench-Capon, T. J. M. 2002. Value-based argumentation frameworks. In *Proc. NMR'02*, 443–454.

Benferhat, S.; Lagrue, S.; and Papini, O. 2004. Reasoning with partially ordered information in a possibilistic logic framework. *Fuzzy Sets and Systems* 144(1):25–41.

Bienvenu, M.; Lang, J.; and Wilson, N. 2010. From preference logics to preference languages, and back. In *Proc. KR-10*, to appear.

Boutilier, C.; Brafman, R.; Domshlak, C.; Hoos, H.; and Poole, D. 2004. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *J. of Artificial Intelligence Research* 21:135–191.

Bouveret, S.; Endriss, U.; and Lang, J. 2009. Conditional importance networks: A graphical language for representing ordinal, monotonic preferences over sets of goods. In *Proc. IJCAI'09*, 67–72.

Brafman, R.; Domshlak, C.; Shimony, S.; and Silver, Y. 2006. Preferences over sets. In *Proc. AAAI'06*, 1001–1006.

Brafman, R. I.; Domshlak, C.; and Shimony, S. E. 2006. On graphical modeling of preference and importance. *J. of Artificial Intelligence Research* 25:389–424.

Brewka, G.; Niemelä, I.; and Truszczyński, M. 2009. Preferences and nonmonotonic reasoning. *AI Magazine* 29(4):69–78.

desJardins, M., and Wagstaff, K. 2005. DD-PREF: A language for expressing preferences over sets. In *Proc. AAAI-05*, 620–626.

Dimopoulos, Y.; Moraitis, P.; and Amgoud, L. 2009. Extending argumentation to make good decisions. In *Proc. ADT'09*, 225–236.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2):321–358.

Goldsmith, J.; Lang, J.; Truszczyński, M.; and Wilson, N. 2008. The computational complexity of dominance and consistency in CP-nets. *J. of Artificial Intelligence Research* 33:403–432.

Modgil, S. 2009. Reasoning about preferences in argumentation frameworks. *Artif. Intell.* 173(9-10):901–934.

Wilson, N. 2004. Extending CP-nets with stronger conditional preference statements. In *Proc. AAAI'04*, 735–741.