Property-based Preferences in Abstract Argumentation

Richard Booth[1], Souhila Kaci[2], and Tjitze Rienstra[1,2]

[1] Université du Luxembourg
6 rue Richard Coudenhove-Kalergi, Luxembourg
richard.booth/tjitze.rienstra@uni.lu
[2] LIRMM (CNRS/Université Montpellier 2)
161 rue Ada, Montpellier, France
souhila.kaci@lirmm.fr

Abstract. Many works have studied preferences in Dung-style argumentation. Preferences over arguments may be derived, e.g., from their relative specificity, relative strength or from values promoted by the arguments. An underexposed aspect in these models is change of preferences. We present a dynamic model of preferences in argumentation, centering on what we call property-based AFs. It is based on Dietrich and List's model of property-based preference and it provides an account of how and why preferences in argumentation may change. The idea is that preferences over arguments are derived from preferences over properties of arguments, and change as the result of moving to different motivational states. We also provide a dialogical proof theory that establishes whether there exists some motivational state in which an argument is accepted.

Keywords: argumentation, preferences, property-based, dialogue, Dung

1 Introduction

Dung's theory of abstract argumentation [1] plays a central role in many approaches to reasoning and decision making in AI. It is based on the concept of an *argumentation framework* (AF, for short), i.e., a set of abstract *arguments* and a binary *attack relation* encoding conflict between arguments. The outcome of an AF is a set of justifiable points of view on the acceptability of its arguments, represented by extensions and computed under a given semantics, different semantics corresponding to different degrees of skepticism or credulousness.

Many works have recognized the importance of *preferences* in this setting. Preferences over arguments may be derived, e.g., from their relative specificity or from the relative strength of the beliefs with which they are built. On the abstract level preferences can be represented by *preference-based* AFs, which instantiate AFs with a preference relation over the set of arguments [2, 3]. An attack of an argument x on y then *succeeds* only if y is not strictly preferred over x. *Value-based* AFs provide yet another account of how preferences are derived [4]. The idea here is that arguments promote certain *values* and that different *audiences* have different preferences over values, from which the preferences over arguments are derived.

An underexposed aspect in these models is change of preferences [5, 6]. Preferences are usually assumed to be fixed and no account is provided of how or why they may change. We address this aspect by applying Dietrich and List's recently introduced model of *property-based preference* [7, 8]. In this model, preferences over alternatives are derived from preferences over sets of properties satisfied by the alternatives. Furthermore, agents are assumed to have a motivational state, consisting of the properties on which the agent focuses in a given situation, when forming preferences over alternatives. The authors present an axiomatic characterization of their model, in terms of a number of reasonable constraints on the relationship between motivational states and preferences.

Our contribution is a new, dynamic model of preferences in argumentation, centering on what we call *property-based* AFs. It is based on the model of Dietrich and List and provides an account of how and why preferences in argumentation may change. Our model generalizes preference-based AFs as well as value-based AFs, if properties are used to represent values. We look at two types of acceptance, called *weak* and *strong* acceptance (i.e., acceptance in *some* or *all* motivational states). We also provide a dialogical proof theory that establishes whether an argument is weakly accepted. It is based on the grounded game [9] and extends it with dialogue moves consisting of properties.

The outline of this paper is as follows. We start in section 2 with some preliminaries concerning abstract argumentation theory. In section 3 we first give a brief outline of preference-based and value-based abstract argumentation. Then we give in section 4 an overview of the relevant parts of Dietrich and List's model of property-based preferences. We move on to our own work in section 5, where we present our model of property-based AFs, followed by a dialogical proof procedure for weak acceptance in section 6. We discuss some related work in section 7 and we conclude in section 8.

2 Preliminaries

We start out with some preliminaries concerning Dung's model of abstract argumentation [1]. We assume that argumentation frameworks are finite.

Definition 1. *An* argumentation framework *(AF for short) is a pair* $AF = (A, \rightarrow)$ *where* A *is a finite set of* arguments *and* $\rightarrow \subseteq A \times A$ *an* attack relation*.*

Given an AF $(A, \rightarrow)$ we say that x *attacks* y and also write $x \rightarrow y$ instead of $(x, y) \in \rightarrow$. The outcome of an AF consists of possible sets of arguments, called *extensions.* A *semantics* embodies a set of conditions that an extension must satisfy. The most studied ones are defined as follows:

Definition 2. *Let* $AF = (A, \rightarrow)$*. An* extension *of* AF *is a set* $E \subseteq A$*. We say that* E *is* conflict-free *iff* $\nexists x, y \in E$ *s.t.* $x \rightarrow y$*; that it* defends *an argument* $x \in A$ *iff* $\forall y \in A$ *s.t.* $y \rightarrow x$, $\exists z \in E$ *s.t.* $z \rightarrow y$*; and we define* $Def(E)$ *by* $Def(E) =$

$\{x \in A \mid E$ *defends* $x\}$. *An extension* $E \subseteq A$ *is said to be:* admissible *iff* E *is conflict free and* $E \subseteq Def(E)$, complete *iff* E *is conflict free and* $E = Def(E)$, stable *iff* E *is admissible and* $\forall x \in A \setminus E$, $\exists y \in E$ *s.t.* $y \rightarrow x$, preferred *iff* E *is maximal (w.r.t. set inclusion) among the set of admissible extensions of* AF *and* grounded *iff* E *is minimal (w.r.t. set inclusion) among the set of complete extensions of* AF.

Note that the grounded extension is unique and always exists, and represents the most skeptical viewpoint on the acceptability of the arguments in the AF. Although the concepts we introduce in this paper can be applied generally to all semantics, we will focus in this paper on the grounded semantics.

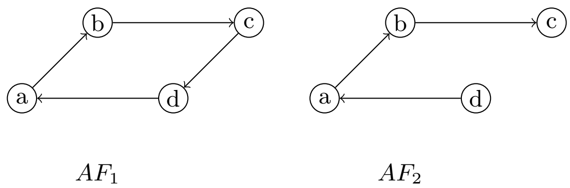

Fig. 1. Two argumentation frameworks.

Example 1. Consider the AF AF_1 shown in figure 1 (nodes represent arguments and arrows represent attacks). The AF has three complete extensions, namely $\emptyset$, $\{a, c\}$ and $\{b, d\}$. The extension $\emptyset$ is also the grounded extension, while $\{a, c\}$ and $\{b, d\}$ are also stable and preferred extension. The AF AF_2 has a single complete extension namely $\{d, b\}$. This extension is thus also a grounded, stable and preferred extension.

3 Preferences and values in argumentation

Preference-based AFs [2] extend AFs with a preference relation over arguments, used to represent the relative strength of arguments. The idea is that an attack of an argument x on y succeeds only if y is not strictly preferred over (i.e., not stronger than) x. A preference-based AF *represents* a unique AF $(A, \rightarrow)$, where the attack relation $\rightarrow$ consists only of the attacks that succeed [10]. The extensions of a preference-based AF are those of the AF that it represents. Formally:

Definition 3. *A preference-based AF (PAF for short) is a triple* $PAF = (A, \rightsquigarrow, \preceq)$ *where* A *is a finite set of arguments,* $\rightsquigarrow$ *an attack relation and* $\preceq$ *a partial pre-order (i.e., a reflexive and transitive relation) or a total pre-order (i.e., a reflexive, transitive and complete relation) over* A. *A PAF* $(A, \rightsquigarrow, \preceq)$ represents *the AF* $(A, \rightarrow)$ *where* $\rightarrow$ *is defined by* $\forall x, y \in A, x \rightarrow y$ *iff* $x \rightsquigarrow y$ *and not* $(x \prec y)$.

Example 2. Consider the PAF $(A, \rightsquigarrow, \preceq)$ where A and $\rightsquigarrow$ are as in AF_1 in example 1 and $\preceq$ is a total pre-order defined by $x \preceq y$ iff $x \in \{b, c\}$ or $y \in \{a, d\}$. We have that $(A, \rightsquigarrow, \preceq)$ represents AF_2, shown in figure 1. This AF has one complete, grounded, stable and preferred extension, namely $\{d, b\}$.

Preference-based AFs give—at least at the abstract level—no account of how preferences over arguments are formed. Bench-Capon's [4] model of *value-based* AFs does. In a value-based AF, the idea is that arguments may promote certain *values* and that different *audiences* have different preferences over values, from which the preferences over arguments are derived. An *audience specific* value-based AF encodes a single audience's preferences over values.

Definition 4. *A* value-based AF *(VAF for short) is a 5-tuple* $(A, \rightsquigarrow, V, val, U)$*, where A is a set of arguments, $\rightsquigarrow$ an attack relation, V a set of* values*, $val : A \to V$ a mapping from arguments to values and U a set of* audiences*. An* audience specific value-based AF *(aVAF for short) is a 5-tuple* $(A, \rightsquigarrow, V, val, <_a)$ *where $a \in U$ is an audience and $<_a$ a partial order (i.e. an irreflexive and transitive relation) over V.*

An $aVAF$ represents a unique PAF [10]:

Definition 5. *An $aVAF$ $(A, \rightsquigarrow, V, val, <_a)$* represents *the PAF $(A, \rightsquigarrow, \preceq)$, where $\preceq$ is defined by $\forall x, y \in A, x \preceq y$ iff $val(x) <_a val(y)$ or $val(x) = val(y)$.*

Since a PAF represents a unique AF, an $aVAF$ also represents a unique AF. The extensions of an $aVAF$ are the extensions of this AF.

Example 3. Consider the $aVAF$ $(A, \rightsquigarrow, V, val, <_a)$ where A and $\rightsquigarrow$ are as in example 1, $V = \{blue, red\}$, $val(a) = val(d) = blue, val(b) = val(c) = red$ and $<_a$ is defined by $x <_a y$ iff $x = red$ and $y = blue$. It can be checked that this $aVAF$ represents the $PAF$from example 2 and thus the AF AF_2 shown in figure 1.

4 Dietrich and List's model of property-based preference

Dietrich and List's model of *property-based preference* [7, 8] aims at giving an account of rational choice that explains how preferences are formed and how they may change. This is opposed to traditional models that assume an agent's preferences over alternatives to be given and fixed. In this model, every alternative $x \in X$ is associated with a set $P(x)$ of *properties* satisfied by x, each $P(x)$ being a subset of a set $\mathcal{P}$ of possible properties. Furthermore, a set $\mathcal{M} \subseteq 2^{\mathcal{P}}$ of *motivational states* encodes sets of properties on which an agent may focus in a given situation. That is, if $M \in \mathcal{M}$ is the agent's state then only the properties in M matter to the agent when forming preferences over X. Change of preferences can then be understood as being caused by moving from one motivational state to another. Note that $\mathcal{M}$ may coincide with $2^{\mathcal{P}}$ but in general this need not be the case, as certain combinations of properties may be deemed inconsistent.

Every state $M \in \mathcal{M}$ gives rise to a preference order (i.e., a total pre-order) $\preceq_M$ over X representing the agent's preferences in the state M. There is thus a family $(\preceq_M)_{M\in\mathcal{M}}$ of preference orders over X. Strict and indifference relations $\prec_M$ and $\sim_M$ are defined as usual.

According to the model of property-based preference, preferences over X are formed using an underlying *weighing relation* $\leq$ over combinations of properties. This relation can be thought of as a 'betterness' relation, i.e., if $S \leq S'$ then the set of properties S' is at least as good as the set of properties S.

Definition 6. *A family* $(\preceq_M)_{M\in\mathcal{M}}$ *of preference orders is called* property-based *if there is a* weighing relation $\leq \subseteq 2^{\mathcal{P}} \times 2^{\mathcal{P}}$ *such that, for every* $M \in \mathcal{M}$ *and* $x, y \in X$, $x \preceq_M y$ *iff* $P(x) \cap M \leq P(y) \cap M$.

The authors present an axiomatic characterization of their model, in terms of two constraints on the relationship between motivational states and preferences.

Theorem 1. *[An axiomatic characterization [7]] Let* $(\preceq_M)_{M\in\mathcal{M}}$ *be a family of preference orders. Consider the following axioms:*

Axiom 1 $\forall x, y \in X$, $\forall M \in \mathcal{M}$, *if* $P(x) \cap M = P(y) \cap M$, *then* $x \sim_M y$.

Axiom 2 $\forall x, y \in X$, $\forall M, M' \in \mathcal{M}$ *s.t.* $M \subseteq M'$, *if* $P(x) \cap (M' \setminus M) = P(y) \cap (M' \setminus M) = \emptyset$ *then* $x \preceq_M y \leftrightarrow x \preceq_{M'} y$.

It holds that if $\mathcal{M}$ *is intersection-closed (i.e.* $M, M' \in \mathcal{M}$ *implies* $M \cap M' \in \mathcal{M}$*) then a family of preference orders* $(\preceq_M)_{M\in\mathcal{M}}$ *satisfies axioms 1 and 2 iff it is property-based.*

Axiom 1 says that the preference relation is indifferent on pairs of alternatives that have the same properties that are at the same time motivational, while axiom 2 says that preferences on pairs of alternatives change only if additional properties become motivational that are satisfied by at least one of the alternatives. A third axiom, strengthening the second and concerned with the class of *separable* weighing relations may be considered as well. The reader is referred to Dietrich and List [7] for details.

5 Property based AFs

The value-based AF model gives an account of where an agent's (or audience's) preferences over arguments come from, namely the relative importance of the values they promote. However, it gives no account of how or why they may change. This motivates us to apply the model of property-based preference in argumentation, giving rise to what we call *property-based AFs*. In a property-based AF, each argument is associated with a set of properties that it satisfies. Among the types of properties we may consider are values promoted by the argument.

Furthermore, a property-based AF consists of a set of motivational states $\mathcal{M}$ and a weighing relation $\leq$ over sets of properties. The idea is as before: $\leq$ encodes the agent's preferences over sets of properties but only properties in the agent's state $M \in \mathcal{M}$ matter when forming preferences over arguments.

Definition 7. *A* property-based AF *is a 6-tuple* $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ *where* A *is a set of arguments,* $\rightsquigarrow$ *an attack relation,* $\mathcal{P}$ *is a set of properties,* $P : A \to 2^{\mathcal{P}}$ *a mapping of arguments to sets of properties,* $\mathcal{M} \subseteq 2^{\mathcal{P}}$ *is an intersection-closed set of motivational states and* $\leq \subseteq 2^{\mathcal{P}} \times 2^{\mathcal{P}}$ *a reflexive, transitive and complete weighing relation.*

Note that there are cases where $\leq$ does not need to be transitive and complete over all sets of properties. For simplicity, however, we assume that it is. The reader is referred to Dietrich and List [7, Remark 1] for details.

If we focus on values as properties then the weighing relation can be understood as encoding the relative importance that an agent associates with different combinations of values, and the motivational state as consisting of the values of which an agent is aware in a given situation.

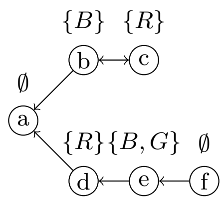

Fig. 2. An argumentation framework

Given a property-based AF, each motivational state $M \in \mathcal{M}$ represents a unique *PAF* which we denote by PAF_M. Preferences in PAF_M are formed by comparing sets of properties satisfied by the arguments, that are at the same time motivational. The AF according to which the agent determines the extensions in the motivational state M, denoted by AF_M, is the AF represented by PAF_M.

Definition 8. *Given a property-based AF* $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ *and a motivational state* $M \in \mathcal{M}$ *we say that:*

- M *represents the* $PAF_M = (A, \rightsquigarrow, \preceq)$*, where* $\preceq$ *is defined by* $\forall x, y \in A, x \preceq y$ *iff* $P(x) \cap M \leq P(y) \cap M$.
- M *represents the AF* $AF_M = (A, \rightarrow_M)$*, which is the AF represented by* PAF_M.

Given an attack $x \rightsquigarrow y$ *and state* $M \in \mathcal{M}$*, we say that* $x \rightsquigarrow y$ *is* enabled *(otherwise* disabled*) in* M *iff* $x \rightarrow_M y$.

Let us illustrate the definitions with an example.

Example 4. Consider the property-based AF $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ where A and $\rightsquigarrow$ and the properties assigned by P to the arguments are as shown in figure 2. Furthermore, $\mathcal{P} = \{R, G, B\}$, $\mathcal{M} = 2^{\mathcal{P}}$ and $\leq$ is defined via a weight function $w : \mathcal{P} \rightarrow \mathbb{Z}$ with $w(R) = w(G) = 1$ and $w(B) = -2$ as follows: $X \leq X'$ iff $\sum_{x \in X} w(x) \leq \sum_{x \in X'} w(x)$. This gives rise to the weighing relation $\{B\} < \{R, B\} = \{G, B\} < \{R, G, B\} = \emptyset < \{R\} = \{G\} < \{R, G\}$, where $<$ is the strict counterpart of $\leq$.

Figure 3 shows the AFs represented by all possible motivational states. We have, e.g., that in $PAF_{\{G\}}$ the argument e is strictly preferred over f, so that the attack from f to e is disabled $AF_{\{G\}}$. On the other hand, in $PAF_{\emptyset}$ and $PAF_{\{B,G\}}$ the argument e is not preferred over f. Here, the attack from f to e succeeds and is therefore enabled in $AF_{\emptyset}$ and $AF_{\{B,G\}}$.

Arguments in the AFs in figure 3 that are a member of the grounded extension of the respective AFs are shown black. We can see, e.g., that a is accepted only in the motivational state $\{R, G\}$.

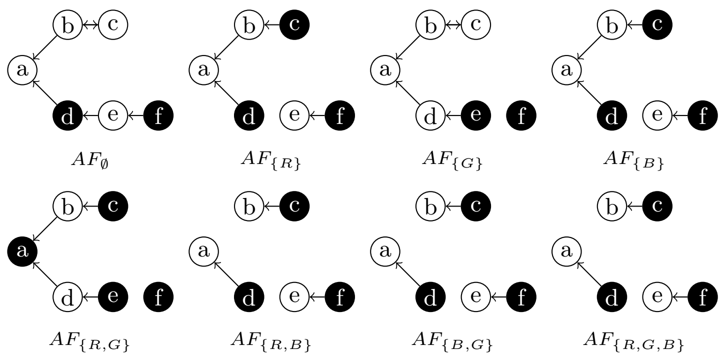

Fig. 3. AFs represented by all motivational states in example 4.

We should remark that in many systems of argumentation, arguments have (in)formal 'logical content'. As a result, conflicts between arguments cannot generally be disregarded, on pain of inconsistency of the AF's outcome. This can be taken into account by requiring, for example, the relation $\rightsquigarrow$ to be symmetric, representing a conflict relation over two arguments, i.e. both arguments cannot be accepted together. In this way one attack between a pair of arguments always remains enabled.

Apart from looking at acceptance of arguments in a given motivational state, we can look at acceptance of arguments in some or all possible states. We will say that an argument is *weakly* (resp. *strongly*) accepted iff it is a member of the grounded extension given some (resp. all) motivational states. Weak acceptance thus means that the agent may accept an argument, namely when she moves to the right motivational state, whereas strong acceptance means that an agent accepts an argument regardless of her motivational state.

Definition 9. *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF and $x \in A$ an argument. We say that x is* weakly accepted (*resp.* strongly accepted) *iff x is a member of the grounded extension of AF_M for some (resp. all) $M \in \mathcal{M}$.*

Example 5 (Continued from example 4). All arguments except b are weakly accepted. Only f is strongly accepted.

The following properties follow directly from theorem 1.

Proposition 1. *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF. We have:*

Property 1 *$\forall x, y \in A$ s.t. $x \rightsquigarrow y$, $\forall M \in \mathcal{M}$ s.t. $P(x) \cap M = P(y) \cap M, x \rightarrow_M y$.*

Property 2 *$\forall x, y \in A$, $\forall M, M' \in \mathcal{M}$ s.t. $M \subseteq M'$, if $P(x) \cap (M' \setminus M) = P(y) \cap (M' \setminus M) = \emptyset$ then $x \rightarrow_M y$ iff $x \rightarrow_{M'} y$.*

Property 1 states that an attack $x \rightsquigarrow y$ is enabled in a motivational state M if x and y have the same set of properties that are also motivational in M, while property 2 states that an attack between two arguments x and y changes only if additional properties become motivational that are satisfied either by x or by y.

6 A dialogical proof theory for weak acceptance

In this section we present a proof procedure to establish weak acceptance of an argument in a property-based AF. It is a dialogical proof procedure because it is based on generating dialogues where two players (PRO and OPP) take alternating turns in putting forward attacks according to a certain set of rules. This is similar in spirit to the *grounded game*, a dialogical proof procedure that establishes an argument's membership of the grounded extension [9]. In the grounded game, PRO repeatedly puts forward arguments (either as an initial claim or in defence against OPP's attacks) and OPP can initiate different disputes by putting forward possible attacks on the arguments put forward by PRO. PRO wins iff it can end every dispute in its favor according to a "last-word" principle.

By contrast, the proof procedure we present simply generates dialogues won by PRO. Such dialogues represent proofs that the initial argument is weakly accepted, and are structured as single sequences of moves where PRO and OPP put forward attacks and, in addition, PRO puts forward properties. If the procedure generates no dialogues then the argument is not weakly accepted.

Dialogical proof procedures make it possible to relate a semantics to a stereotypic pattern of dialogue. It has been shown, e.g., that the grounded and preferred credulous semantics can be related to persuasion and socratic style dialogue [11, 12]. Dialogues generated by our procedure can also be thought of as persuasion dialogues, where PRO has the additional freedom to change the motivational state of the players by putting forward properties. Intuitively, this may benefit PRO in two ways: PRO can enable attacks necessary to put up a successful line of defence, and disable attacks put forward by the opponent from which PRO cannot defend its own arguments. PRO thus persuades OPP to accept an argument, where PRO decides which properties become motivational. Dialogues are structured as follows.

Definition 10. *Let* $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ *be a property-based AF. A* dialogue *is a sequence* $S = (m_1, \ldots, m_n)$*, where each* m_i *is either:*
- *an* attack move *"***OPP:** $x \rightsquigarrow y$*", where* $x, y \in A$ *and* $x \rightsquigarrow y$*,*
- *a* defence move *"***PRO:** $x \rightsquigarrow y$*", where* $x, y \in A$ *and* $x \rightsquigarrow y$*,*
- *an* enabling property move *"***PRO:** $P+$*", where* $P \subseteq \mathcal{P}$*,*
- *a* disabling property move *"***PRO:** $P-$*", where* $P \subseteq \mathcal{P}$*,*
- *a* conceding move *"***OPP: ok***",*
- *a* success claim move *"***PRO: win***".*

We denote by $S \cdot S'$ *the concatenation of* S *and* S' *and we say that* S *is a* subsequence *of* S' *iff* $S' = S'' \cdot S \cdot S'''$ *for some* S'', S'''*, and that* S *is a* proper *subsequence of* S' *iff* $S' = S'' \cdot S \cdot S'''$ *for nonempty* S'' *or* S'''*.*

Definition 11. *Let* $S = (m_1, \ldots, m_n)$ *be a dialogue. We denote the* motivational state in S at index i *by* M_i^S*, defined recursively by:*

$$M_i^S = \begin{cases} \emptyset & \text{if } i = 0, \\ M_{i-1}^S \cup P & \text{if } m_i = \textbf{PRO:}\, P+ \ \text{ or } m_i = \textbf{PRO:}\, P-, \\ M_{i-1}^S & \text{otherwise.} \end{cases}$$

We now define a set of production rules that generate *weak x-acceptance dialogues*. Note that AFs containing cycles may generate infinite sequences of moves. We prevent this by requiring dialogues to be finite.

Definition 12 (Weak acceptance dialogue). *Let* $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ *be a property-based AF and let* $x \in A$*.*
- *A* weak x-acceptance dialogue *is a finite sequence*

$$S_1 \cdot (\textbf{PRO: win})$$

  where S_1 *is an* x*-attack sequence.*
- *An* x-attack sequence *is a sequence*

$$(\textbf{OPP:}\, y_1 \rightsquigarrow x) \cdot S_1 \cdot \ldots \cdot (\textbf{OPP:}\, y_n \rightsquigarrow x) \cdot S_n \cdot (\textbf{OPP: ok})$$

  where $\{y_1, \ldots, y_n\} = \{y \mid y \rightsquigarrow x\}$ *and each* S_i *is a* y_i*-defence sequence.*
- *An* x-defence sequence *is either:*
  - *a* regular x-defence sequence

$$(\textbf{PRO:}\, y \rightsquigarrow x) \cdot S_1$$

    for some $y \in A$ *s.t.* $y \rightsquigarrow x$*, where* S_1 *is a* y*-attack sequence,*
  - *an* enabling property defence sequence

$$(\textbf{PRO:}\, P+) \cdot S_1$$

    for some $P \subseteq \mathcal{P}$*, where* S_1 *is a regular* x*-defence sequence,*

- a disabling property defence sequence

$$(\textbf{\textit{PRO:}}\, P-)$$

for some $P \subseteq \mathcal{P}$.

Intuitively, a disabling property move can be interpreted as saying "the preceding move is invalid considering the properties P." An enabling property move, on the other hand, says "the following move is valid considering the properties P." Not every weak x-acceptance dialogue, generated by the production rules in definition 12, will follow this interpretation. We need to impose a number of additional constraints to ensure that property moves make sense.

Definition 13 (Property-consistency). *Let* $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ *be a property-based AF and* $S = (m_1, \ldots, m_n)$ *a sequence. We say that* S *is* property-consistent *iff for all* $i \in [1, \ldots, n]$*, we have:*

1. $M_i^S \in \mathcal{M}$
2. *If* $m_i = \textbf{\textit{PRO:}}\, x \rightsquigarrow y$ *then for all* $j \in [i, \ldots, n]$*,* $x \rightarrow_{M_j^S} y$*,*
3. *If* $m_i = \textbf{\textit{PRO:}}\, P-$ *and* $m_{i-1} = \textbf{\textit{OPP:}}\, x \rightsquigarrow y$ *then for all* $j \in [i, \ldots, n]$*,* $x \not\rightarrow_{M_j^S} y$*.*

Condition 1 ensures that property moves are valid in the sense that they actually lead to a new motivational state $M \in \mathcal{M}$. Conditions 2 and 3 ensure that a property move does not undermine preceding property moves. That is, condition 2 ensures that attacks put forward by PRO remain enabled in subsequent states and condition 3 ensures that disabled attacks remain disabled.

Example 6 (Continued from example 4). Consider the following two property-consistent weak acceptance dialogues for the argument a.

Index	Move	State
1	**OPP:** $b \rightsquigarrow a$	$\emptyset$
2	**PRO:** $c \rightsquigarrow b$	$\emptyset$
3	**OPP:** $b \rightsquigarrow c$	$\emptyset$
4	**PRO:** $\{R\}-$	$\{R\}$
5	**OPP: ok**	$\{R\}$
6	**OPP:** $d \rightsquigarrow a$	$\{R\}$
7	**PRO:** $\{G\}+$	$\{R, G\}$
8	**PRO:** $e \rightsquigarrow d$	$\{R, G\}$
9	**OPP:** $f \rightsquigarrow e$	$\{R, G\}$
10	**PRO:** $\emptyset-$	$\{R, G\}$
11	**OPP: ok**	$\{R, G\}$
12	**OPP: ok**	$\{R, G\}$
13	**PRO: win**	$\{R, G\}$

Index	Move	State
1	**OPP:** $b \rightsquigarrow a$	$\emptyset$
2	**PRO:** $c \rightsquigarrow b$	$\emptyset$
3	**OPP:** $b \rightsquigarrow c$	$\emptyset$
4	**PRO:** $\{R, G\}-$	$\{R, G\}$
5	**OPP: ok**	$\{R, G\}$
6	**OPP:** $d \rightsquigarrow a$	$\{R, G\}$
7	**PRO:** $e \rightsquigarrow d$	$\{R, G\}$
8	**OPP:** $f \rightsquigarrow e$	$\{R, G\}$
9	**PRO:** $\emptyset-$	$\{R, G\}$
10	**OPP: ok**	$\{R, G\}$
11	**OPP: ok**	$\{R, G\}$
12	**PRO: win**	$\{R, G\}$

Explanation: In the dialogue shown on the left, the initial exchange of attacks consists of $b \rightsquigarrow a, c \rightsquigarrow b$ and $b \rightsquigarrow c$. PRO must end this line of argument by making a disabling property to disable the attack $b \rightsquigarrow c$. PRO moves **PRO:** $\{R\}-$

and as a result, the motivational state of the dialogue becomes $\{R\}$. OPP's next attack is $d \rightsquigarrow a$. PRO cannot move $e \rightsquigarrow d$ because this attack is disabled in the current motivational state. PRO moves **PRO:** $\{G\}+$, changing the motivational state of the dialogue to $\{R, G\}$, so that $e \rightsquigarrow d$ is enabled. To OPP's attack $f \rightsquigarrow e$ PRO responds with an empty disabling move, as $f \rightsquigarrow e$ is already disabled in the current motivational state. The dialogue on the right is similar with the exception that PRO immediately moves both R and G when making a disabling property move on line 4. As a result, no enabling property move is needed on line 7 because the attack $d \rightsquigarrow e$ is already enabled.

The existence of a property-consistent weak x-acceptance dialogue implies weak acceptance of x, i.e., it is a sound proof procedure:

Lemma 1 (Soundness). *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF and $x \in A$. If there exists a property-consistent weak x-acceptance dialogue $S = (m_1, \ldots, m_n)$ then x is a member of the grounded extension of the AF represented by M_n^S. Hence x is weakly accepted.*

Proof (of lemma 1). Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property based AF, $x \in A$ and S a property-consistent weak x-acceptance dialogue. A subsequence S' of S that is a y-attack sequence (for some $y \in A$) will be called a *y-attack subsequence.* We denote the *depth* of an attack subsequence S' by $D(S')$ and define it by $D(S') = 0$, if $S' = (\textbf{OPP: ok})$ and $1 + k$ otherwise, where $k = max(\{D(S'') \mid S'' \in T\})$, where T is the set of attack sequences that are proper subsequences of S'. Furthermore from hereon we denote the grounded extension of $(A, \rightarrow_{M_n^S})$ by G. We show that for every y-attack subsequence S' it holds that $y \in G$. We prove this by strong induction on the depth of S'. Let the induction hypothesis $H(k)$ stand for "if S' is a y-attack subsequence with depth k then $y \in G$."

- Base case ($H(0)$): Here $S' = (\textbf{OPP: ok})$, thus y has no attackers in $(A, \rightsquigarrow)$, hence no attackers in $(A, \rightarrow_{M_n^S})$. It follows that $y \in G$.
- Induction step: Assume $H(0), \ldots, H(k-1)$ holds. We need to prove $H(k)$. It can be checked that for every z s.t. $z \rightsquigarrow y$, either:
  - There is a z'-attack sequence S'' that is a proper subsequence of S'. Thus $D(S'') < k$ and $z' \rightsquigarrow z$. From $H(D(S''))$ and the fact that S is property-consistent it follows that z is attacked by G.
  - S' contains a disabling property move. Hence $z \not\rightarrow_{M_n^S} y$.

  This means that for every z such that $z \rightarrow_{M_n^S} y$, G attacks z, hence $y \in G$.

By the principle of strong induction it follows that if there is a y-attack subsequence then $y \in G$. Thus we have $x \in G$, hence x is weakly accepted. □

Conversely, if x is weakly accepted then a property-consistent weak x-acceptance dialogue exists:

Lemma 2 (Completeness). *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF and $x \in A$ be weakly accepted. There exists a weak x-acceptance dialogue S that is property-consistent.*

Proof. Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF and $x \in A$ be weakly accepted. Then there is some $M \in \mathcal{M}$ s.t. x is a member of the grounded extension of $(A, \rightarrow_M)$. From hereon we use M to refer to any such motivational state and G to refer to the grounded extension of $(A, \rightarrow_M)$.

First some notation: The *characteristic function* $C : 2^A \rightarrow 2^A$ of an AF $(A, \rightarrow)$ is defined by $C(X) = \{x \in A \mid x \text{ is defended by } X\}$. It is well known that G coincides with the least fixed point of C [1]. We define the *degree* $Deg(x)$ of any $x \in G$ as the smallest positive integer s.t. $x \in C^n(\emptyset)$.

We now prove, by strong induction over the degree of an argument $y \in G$ that there exists a property consistent weak y-acceptance dialogue. Let $H(k)$ stand for "If $y \in G$ and $Deg(y) = k$ then there exists a property consistent weak y-acceptance dialogue."

- Base case ($H(0)$): If $y \in G$ and $Deg(y) = 0$ then there is no $z \in A$ s.t. $z \rightarrow_M y$ and we can define S by $(\textbf{OPP:}\, z_1 \rightsquigarrow y) \cdot S' \cdot \ldots \cdot (\textbf{OPP:}\, z_n \rightsquigarrow y) \cdot S' \cdot (\textbf{OPP: ok}) \cdot (\textbf{PRO: win})$, where $\{z_1, \ldots, z_n\} = \{z' \mid z' \rightsquigarrow y\}$ and $S' = (\textbf{PRO:}\, M-)$. It can be checked that S is a property consistent weak y-acceptance dialogue.
- Induction step: Assume $H(0), \ldots, H(k-1)$ holds. Thus if $y' \in G$ and $Deg(y') < k$ then there exists a property consistent weak y'-acceptance dialogue. We denote this dialogue by $S(y')$. We need to prove $H(k)$.
  Assume that $y \in G$ and $Deg(y) = k$. It follows that for every $z \in A$ s.t. $z \rightarrow_M y$, there exists an argument which we denote by $def(z, y)$ such that $def(z, y) \in G$ and $def(z, y) \rightarrow_M z$. Furthermore from the fixpoint construction it follows that $Deg(def(z, y)) < k$, so that $S(def(z, y))$ is well defined. Now, for every $z \in A$ s.t. $z \rightsquigarrow y$ we define $T_y(z)$ by (1) $T_y(z) = (\textbf{OPP:}\, z \rightsquigarrow y) \cdot (\textbf{PRO:}\, M-)$, if $z \not\rightarrow_M y$ and (2) $T_y(z) = (\textbf{OPP:}\, z \rightsquigarrow y) \cdot (\textbf{PRO:}\, M+) \cdot S'$, if $z \rightarrow_M y$—where S' is defined by $S(def(z, y)) = S' \cdot (\textbf{PRO: win})$. It can be checked that $T_y(z_1) \cdot \ldots \cdot T_y(z_i) \cdot (\textbf{OPP: ok}) \cdot (\textbf{PRO: win})$ (where $\{z_1, \ldots, z_i\} = \{z' \mid z' \rightsquigarrow y\}$) is a property consistent weak y-acceptance dialogue.

By the principle of strong induction it follows that for every $y \in G$, there exists a property consistent weak y-acceptance dialogue. Hence, there exists a property consistent weak x-acceptance dialogue. □

Notice that in the fourth move of in the second dialogue in example 6, PRO puts forward both R and G in a disabling property move. However, it suffices to put forward just R, as in the first dialogue, because G is not relevant with respect to disabling the attack $b \rightsquigarrow c$. We call a dialogue in which property moves are relevant a *property-relevant* dialogue. Property moves in a property-relevant dialogue consist only of properties satisfied by one of the arguments involved in the attack that is enabled or disabled.

Definition 14 (Property-relevance). *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF and $S = (m_1, \ldots, m_n)$ a weak acceptance dialogue. We say that S is* property-relevant *iff for all $i, j \in [1, \ldots, n]$ s.t. $j = i + 1$, we have:*

1. *If $m_i =$ **OPP:** $x \rightsquigarrow y$ and $m_j =$ **PRO:** $P-$ then $P \subseteq P(x) \cup P(y)$.*

*2. If $m_i =$ **PRO:** $P+$ and $m_j =$ **PRO:** $x \rightsquigarrow y$ then $P \subseteq P(x) \cup P(y)$.*

Note that in example 6 the first dialogue is property-relevant, whereas the second one is not. Focusing on property-relevant dialogues can be used to optimize the algorithm. Furthermore, it makes sense intuitively: when persuading an opponent to accept an argument, one does not refer to properties not relevant to this objective.

As a final result we show that weak acceptance of an argument implies the existence of a property-consistent weak x-acceptance dialogue that is, in addition, property relevant. However, this requires that $\mathcal{M}$ is sufficiently rich to ensure that PRO is not forced to put forward irrelevant properties. This can be achieved by assuming that $\mathcal{M} = 2^{\mathcal{P}}$, but note that there are cases where a weaker assumption is sufficient.

Lemma 3 (Property-relevant completeness). *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF where $\mathcal{M} = 2^{\mathcal{P}}$, and let $x \in A$ be weakly accepted. There exists a weak x-acceptance dialogue S that is property-consistent and property-relevant.*

Proof. Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF and $x \in A$ be weakly accepted. Let $S = (m_1, \ldots, m_n)$ be the property-consistent weak x-acceptance dialogue (for x a member of the grounded extension of $(A, \rightarrow_M)$) as constructed in the proof of lemma 2. That is, every property move in S is either of the form **PRO:** $M+$ or **PRO:** $M-$. Using property 1 (2) it can be checked that the dialogue S' formed by

- replacing every move $m_i =$ **PRO:** $M+$ in S by **PRO:** $M'+$, where $M' = M \cap P(x) \cup P(y)$ where x, y are defined by $m_{i+1} =$ **PRO:** $x \rightsquigarrow y$, and
- replacing every move $m_i =$ **PRO:** $M-$ in S by **PRO:** $M'-$, where $M' = M \cap P(x) \cup P(y)$ where x, y are defined by $m_{i-1} =$ **OPP:** $x \rightsquigarrow y$,

is also a property-consistent weak x-acceptance dialogue, that is in addition property-relevant. □

Summarizing, we have the following result.

Theorem 2. *Let $(A, \rightsquigarrow, \mathcal{P}, P, \mathcal{M}, \leq)$ be a property-based AF.*

- *An argument $x \in A$ is weakly accepted iff there exists a weak x-acceptance dialogue that is property-consistent.*
- *If $\mathcal{M} = 2^{\mathcal{P}}$ then an argument $x \in A$ is weakly accepted iff there exists a weak x-acceptance dialogue that is property-consistent and property-relevant.*

Proof. Follows from lemmas 1, 2 and 3. □

7 Related work

We already mentioned the relation of our model with that of preference and value-based AFs [2, 4]. Also related is a study of value-based AFs where arguments promote multiple values [10], concerned mainly with the problem of deriving a unique preference order over arguments from a preference relation over individual values. Note that in our approach, a property-based AF together with a motivational state already defines a unique preference order over arguments.

Furthermore, Bench-Capon et al. have considered dialogues in which a proponent can make moves consisting of value preferences [13]. In this approach, the outcome of a winning dialogue corresponds to the specification of an audience (i.e., a preference order over values) such that some initial set of arguments is accepted in the corresponding $aVAF$.

Also related are Modgil's model of *extended AFs*, in which arguments attack and disable attacks between other arguments [14]. Such arguments can be seen as meta-level arguments expressing preferences over object level arguments. Whereas we take the agent's state (which determines whether individual attacks are enabled) to be external to the AF, here it is part of AF itself. That is, whether an attack is enabled depends on the status of a metalevel argument.

Our work shares methodological similarities with work of Kontarinis et al. [15], who present a goal-oriented procedure to determine which attacks to disable or enable in order to make an argument accepted under a given semantics. While the procedure that they present is designed to be implemented as a term rewriting system, our procedure is defined simply by a set of production rules, amenable to implementation using e.g. PROLOG.

8 Conclusion and future work

We presented a dynamic model of preferences in argumentation, based on Dietrich and List's model of property-based preference. It provides an account of how and why preferences in argumentation may change and generalizes both preference-based AFs and value-based AFs, if properties are taken to be values. We consider a number of directions for future work. First, we plan to complete the proof-theoretic picture by looking at the problem of deciding whether an argument is strongly accepted. In addition, we will consider other semantics in addition to grounded. Second, we plan to investigate the possibility of axiomatizing property-based AFs, in the spirit of Dietrich and List's axiomatization as presented in section 4. Finally, we intend to look at connections between property-based AFs and Modgil's model of extended AFs.

Acknowledgements

Richard Booth is supported by the FNR (National Research Fund) DYNGBaT project.

References

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

2. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Ann. Math. Artif. Intell. **34**(1-3) (2002) 197–215
3. Simari, G., Loui, R.: A mathematical treatment of defeasible reasoning and its implementation. Artificial Intelligence **53** (1992) 125–157
4. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. J. Log. Comput. **13**(3) (2003) 429–448
5. Liu, F.: Reasoning about preference dynamics. Springer (2011)
6. Grüne-Yanoff, T., Hansson, S.O., eds.: Preference change: Approaches from philosophy, economics and psychology. Springer (2009)
7. Dietrich, F., List, C.: Where do preferences come from? International Journal of Game Theory (2012) 1–25
8. Dietrich, F., List, C.: A reason-based theory of rational choice. Nous **47**(1) (2013) 104–134
9. Modgil, S., Caminada, M.: Proof theories and algorithms for abstract argumentation frameworks. In: Argumentation in Artificial Intelligence. Springer (2009) 105–129
10. Kaci, S., van der Torre, L.: Preference-based argumentation: Arguments supporting multiple values. Int. J. Approx. Reasoning **48**(3) (2008) 730–751
11. Caminada, M., Podlaszewski, M.: Grounded semantics as persuasion dialogue. In: Proceedings of COMMA 2012. (2012) 478–485
12. Caminada, M.: Preferred semantics as socratic discussion. In: Proceedings of the 11th AI* IA Symposium on Artificial Intelligence. (2010) 209–216
13. Bench-Capon, T.J.M., Doutre, S., Dunne, P.E.: Audiences in argumentation frameworks. Artif. Intell. **171**(1) (2007) 42–71
14. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artif. Intell. **173**(9-10) (2009) 901–934
15. Kontarinis, D., Bonzon, E., Maudet, N., Perotti, A., van der Torre, L., Villata, S.: Rewriting rules for the computation of goal-oriented changes in an argumentation system. In Kontarinis, D., Bonzon, E., Maudet, N., Perotti, A., van der Torre, L., Villata, S., eds.: CLIMA. Lecture Notes in Computer Science, Springer (2013) Forthcoming.