Computational Models of Argument
C. Reed et al. (Eds.)

doi:10.3233/FAIA240332

Value-Based Reasoning in ASPIC$^+$

Johannes P. WALLNER [a], Adam WYNER [b,1] and Tomasz ZUREK [c]
[a] *Institute of Software Technology, Graz University of Technology*
[b] *Department of Computer Science, Swansea University*
[c] *Informatics Institute, University of Amsterdam*
ORCiD ID: Johannes P. Wallner https://orcid.org/0000-0002-3051-1966, Adam Wyner https://orcid.org/0000-0002-2958-3428, Tomasz Zurek https://orcid.org/0000-0002-9129-3157

Abstract. In Value-based Argumentation Frameworks (VAFs), values are ascribed to abstract arguments and ordered one to another to reflect an audience's preferences. An attack of one argument on another is successful only if the audience does not prefer the value of the attacked argument to the value of the attacking argument. Audiences can disagree about admissible arguments relative to their value preferences. Complementary to VAFs, this paper presents a novel integration of Value-Based Reasoning Frameworks (VBFs) with instantiated argumentation, specifically we focus on the structured argumentation approach of ASPIC$^+$. Agents associate literals with social values and weight of values; together, these are used to filter the literals compatible with their values. Such a set of literals is used to construct agent-relative ASPIC$^+$ knowledge bases, agent-relative instantiated arguments, and argumentation frameworks (AFs). Agents can attack one another's arguments. VAF and VBF present complementary perspectives on values on arguments. VBF contributes a new, formal, articulated view of agreement and disagreement amongst agents, which is grounded in their values. In addition, VBF helps us understand how different agents choose what to argue from out of a pool of common resources.

Keywords. abstract argumentation, instantiated argumentation, value-based argumentation

1. Introduction

In this paper, we incorporate Value-based Reasoning Frameworks (VBFs) [1,2] into ASPIC$^+$ [3], a prominent approach to structured argumentation [4]. The paper is motivated by two observations. First, from a common pool of information, agents may propose different arguments. Yet, what justifies each agent's selection of propositions and rules for their arguments? Second, agents may propose the same argument, yet justify them in different, perhaps contrasting, ways. To address these, we provide a theoretical framework wherein each agent in a multi-agent system has a knowledge-base (KB) derived from a resource of positive literals shared amongst all agents, yet relativised to their values. In addition, arguments in ASPIC$^+$ are created from a relativised KB for each agent. We explain these observations in relation to an agent's values. Finally, we explore how one agent's arguments attack another agent's arguments, where underlyingly the attacks relate to each agent's values.

[1]Corresponding Author: Adam Wyner, a.z.wyner@swansea.ac.uk. This research was funded in whole or in part by the Austrian Science Fund (FWF) P35632.

As discussed in Section 5, Value-based Argumentation Frameworks (VAFs) [5] ascribe values to abstract arguments and order values according to an audience's (agent's) preferences. Argument attack is successful only if the audience does not prefer the value of the attacked argument to the value of the attacking argument. Audiences can disagree about admissible arguments relative to their value preferences. In this approach, there is one argumentation framework to which an audience's value preferences apply.

Each proposal relates to Perelman's typology of agents' arguments [6], where "They discuss ..., the ends to be considered, ..., the interpretation and characterisation of facts.", VAF relates to the "ends" and VBF to the "characterisation of facts". Thus, VAF and VBF are complementary approaches to how agents argue relative to their values.

The integration of VBF with ASPIC$^+$ is the main novel contribution of this work, which individualises conflicts based on values. The integration introduces agent-based, distinct argumentation theories and frameworks such that each agent can have distinct complete extensions. Consequently, while intra-agent extensions can be conflict-free, inter-agent extensions might not be. We also identify collectively acceptable arguments. Value-based Reasoning Frameworks (VBFs) are presented in Section 2. Section 3 reviews ASPIC$^+$. VBF and ASPIC$^+$ are integrated in Section 4, which includes examples and formal results on collective decision-making. Related work in Section 5 compares the integration to prior approaches. We conclude with some discussion in Section 6.

2. Value-based Reasoning Frameworks

This section is an overview of VBF [1,2]. We assume a set of agents $\texttt{Agents} = \{\alpha_1, \ldots, \alpha_n\}$. Agents may uphold different values, denoted by $\texttt{Values}$, towards propositions $\texttt{Prop}$. Values may be considered along the lines of [7]. Both $\texttt{Values}$ and $\texttt{Prop}$ are (distinct) sets, the former denotes abstract expressions of concepts, such as freedom, security, etc, and the latter a set of positive literals (atoms aka atomic propositions).

Agents associate, for a given value, a weight to some atoms in $\texttt{Prop}$, or are indifferent towards others. Formally, we totally order a set of scalar elements $\texttt{Scale}$, e.g., the whole numbers, and define $\texttt{Weight}$ to be $\texttt{Scale} \cup \{?\}$, with "?" denoting indifference.[2]

The function $\texttt{ValLimit} : (\texttt{Agents} \times \texttt{Values}) \to \texttt{Weight}$ associates, for a specific agent and a value, a degree of importance of that value to the agent. Formally, this is represented as a "limit" that represents the lowest weight for a specific value that is deemed relevant for the current discourse or is acceptable in principle.

The function $\texttt{ValProp} : (\texttt{Agents} \times \texttt{Values} \times \texttt{Prop}) \to \texttt{Weight}$ associates a weight for a specific agent, value, and atom. For instance, an agent α may assign $1 \in \texttt{Weight}$ to atom $p \in \texttt{Prop}$ according to the value $v \in \texttt{Values}$, while having a "threshold" $\texttt{ValLimit}(\alpha, v) = 2$. Intuitively, this signals that p is, according to agent α and value v, deemed potentially to be included in the discourse, at least in principle[3].

Definition 1. *We define a value-based framework (VBF) as a tuple $V = (\texttt{Agents}, \texttt{Prop}, \texttt{Values}, \texttt{Scale})$, with $\texttt{Agents}$ a set of agents, $\texttt{Prop}$ a set of positive literals, $\texttt{Values}$*

[2]Associating a value to a literal is an auxiliary characterisation or assessment by an agent. Each agent constructs arguments based on their value-based assessment of literals; consequently, not every true literal (or axiom) need be used in the construction of every argument by every agent.

[3]There are issues to develop in associating a value with a negative literal. For instance, while *not flying* might be positively associated with a value of *ecology*, *taking a cruise liner*, which is an instance of not flying, is *not positively* associated with that value.

a set of values, and $\mathtt{Scale}$ *a totally ordered set. Associated to each VBF* V *are two functions* $\mathtt{ValLimit}_V$ *and* $\mathtt{ValProp}_V$*, as defined above.*

We can omit the subscript V for $\mathtt{ValLimit}_V$ and $\mathtt{ValProp}_V$, unless not clear from the context. Based on a VBF, each agent "filters" Prop, according to their values.

Definition 2. *Let* $V = (\mathtt{Agents}, \mathtt{Prop}, \mathtt{Values}, \mathtt{Scale})$ *be a VBF, and* $\alpha \in \mathtt{Agents}$*:*

$$\mathtt{PropBaseClean}_\alpha = \{p \in \mathtt{Prop} \mid \forall v \in \mathtt{Values} : \mathtt{ValProp}(\alpha, v, p) \not< \mathtt{ValLimit}(\alpha, v)\}.$$

In words, $\mathtt{PropBaseClean}_\alpha$ contains all atoms that pass *all* value filters for agent α.[4] "?" is not ordered in Scale, so if either $\mathtt{ValProp}(\alpha, v, p) = ?$ or $\mathtt{ValLimit}(\alpha, v) = ?$, the inequality in Definition 2 holds (and the atom "passes" the filter).

Example 1. *Suppose two agents* $\alpha, \beta \in \mathtt{Agents}$*, two values* $v_a, v_b \in \mathtt{Values}$*,* $\mathtt{Scale} = \{0, 1, 2, 3\}$*, and* $\mathtt{Prop} = \{p_1\}$*. Agent* α *has a higher limit for* v_a *than for* v_b*, i.e.,* $\mathtt{ValLimit}(\alpha, v_a) = 2$ *and* $\mathtt{ValLimit}(\alpha, v_b) = 1$*. Agent* β *treats these values differently:* $\mathtt{ValLimit}(\beta, v_a) = 1$ *and* $\mathtt{ValLimit}(\beta, v_b) = 3$*. The weights are* $\mathtt{ValProp}(\alpha, v_a, p_1) = 3$*,* $\mathtt{ValProp}(\alpha, v_b, p_1) = 2$*,* $\mathtt{ValProp}(\beta, v_a, p_1) = 2$*, and* $\mathtt{ValProp}(\beta, v_b, p_1) = 2$*.*

In such a case $\mathtt{PropBaseClean}_\alpha$ *will contain the atom* p_1*, because it passes the filter of both values, while* $\mathtt{PropBaseClean}_\beta$ *does not contain* p_1*, because it fails to pass a filter of value* v_b*, i.e., (*$\mathtt{ValProp}(\beta, v_b, p_1) < \mathtt{ValLimit}(\beta, v_b)$*).*

3. Background on ASPIC$^+$

We recap the relevant elements of the ASPIC$^+$ [8,3] framework as used in this paper. We assume a knowledge base of so-called ordinary premises and axioms. Informally, the former signal defeasible assumptions, while the latter are interpreted as assumptions that hold strictly. Both sets are based on a formal language $\mathscr{L}$, a set of literals.

Definition 3. *A knowledge base is a set* $\mathscr{K}_n \cup \mathscr{K}_p = \mathscr{K} \subseteq \mathscr{L}$*, with two disjoint sets* $\mathscr{K}_n$ *(axioms) and* $\mathscr{K}_p$ *(ordinary premises).*

Contraries of an atom are defined using a contrary function $^-$ that reflects symmetric contraries. In this paper, $\mathscr{L}$ is composed of atoms p and their negations $\neg p$. Complementary literals are contrary to each other, formally for each $l \in \mathscr{L}$ its (unique) contrary $\bar{l}$ is $\neg p$ if $l = p$ and p if $l = \neg p$. We assume that $\mathscr{L}$ is closed under negation (i.e., if $l \in \mathscr{L}$ then so is $\bar{l} \in \mathscr{L}$).

Rules in ASPIC$^+$ are defined as rules over $\mathscr{L}$. We denote the set of all rules as $\mathscr{R}$. ASPIC$^+$ allows for both strict $\mathscr{R}_s$ and defeasible rules $\mathscr{R}_d$. A defeasible rule has the form $a_1, \ldots, a_n \Rightarrow b$ while a strict rule has the form $a_1, \ldots, a_n \rightarrow b$. When we do not distinguish between strict or defeasible rules, we write $a_1, \ldots, a_n \rightsquigarrow b$. A partial function $n : \mathscr{R}_d \rightarrow \mathscr{L}$ gives names to defeasible rules. For a rule $r = a_1, \ldots, a_n \rightsquigarrow b$, we denote its head by $head(r) = b$ and its body by $body(r) = \{a_1, \ldots, a_n\}$.

Definition 4. *An argumentation theory (AT) is a tuple* $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$*, with* $\mathscr{L}$ *a set of literals,* $\mathscr{R}$ *a set of strict and defeasible rules,* $n : \mathscr{R}_d \rightarrow \mathscr{L}$ *a partial function,* $^-$ *a contrary function, and* $\mathscr{K}$ *a knowledge base.*

[4] As discussed in [1,2], this is a simplifying assumption which can be relaxed to represent other scenarios.

An AT gives rise to argument structures, which are derivations that "start off" at elements in the knowledge base and use rules to derive a conclusion. Ordinary premises and axioms are arguments themselves.

Definition 5. *Given an AT $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$, the set of arguments in T is inductively defined as the smallest set satisfying the following.*

- *If $x \in \mathscr{K}$, then $A = x$ is an argument in T with $\texttt{Conc}(A) = x$.*
- *If $A_1, \dots, A_n$ are arguments in T, $x_i = \texttt{Conc}(A_i)$ for $1 \leq i \leq n$, and $(x_1, \dots, x_n \rightsquigarrow x) \in \mathscr{R}$, then $A = A_1, \dots, A_n \rightsquigarrow x$ is an argument in T with $\texttt{Conc}(A) = x$.*

As defined by Modgil and Prakken [3], we consider only arguments of finite size (i.e., with finitely many rule "applications").

Definition 6. *Let $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$ be an AT, and A an argument in T.*

- *If $A = x \in \mathscr{K}$ then $\texttt{Sub}(A) = \{A\}$ and $\texttt{Rules}(A) = \emptyset$.*
- *If $A = A_1, \dots, A_n \rightsquigarrow x$, then*
  $\texttt{Sub}(A) = \{A\} \cup \bigcup_{i=1}^{n} \texttt{Sub}(A_i)$,
  $\texttt{TopRule}(A) = (\texttt{Conc}(A_1), \dots, \texttt{Conc}(A_n) \rightsquigarrow x)$, *and*
  $\texttt{Rules}(A) = \{\texttt{TopRule}(A)\} \cup \bigcup_{i=1}^{n} \texttt{Rules}(A_i)$.

Further, $\texttt{Prem}(A) = \texttt{Sub}(A) \cap \mathscr{K}$, $\texttt{Prem}_\texttt{d}(A) = \texttt{Prem}(A) \cap \mathscr{K}_p$, *and* $\texttt{DefRules}(A) = \texttt{Rules}(A) \cap \mathscr{R}_d$. *Moreover,* $\texttt{defPart}(A) = \texttt{Prem}_\texttt{d}(A) \cup \texttt{DefRules}(A)$. *If $A \in \mathscr{K}$, then* $\texttt{TopRule}(A)$ *is undefined.*

We extend the shorthands for a set of arguments $\mathscr{A}$ as $\texttt{Conc}(\mathscr{A}) = \{\texttt{Conc}(A) \mid A \in \mathscr{A}\}$ and $\texttt{TopRule}(\mathscr{A}) = \{\texttt{TopRule}(A) \mid A \in \mathscr{A}\}$. For each shorthand $f \in \{\texttt{Sub}, \texttt{Rules}, \texttt{DefRules}, \texttt{Prem}, \texttt{Prem}_\texttt{d}\}$ returning a set, we define $f(\mathscr{A}) = \bigcup_{A \in \mathscr{A}} f(A)$. An argument A is an immediate subargument of $B = A_1, \dots, A_n \rightsquigarrow x$ if $A \in \{A_1, \dots, A_n\}$.

We recall the notions of attacks between arguments next.

Definition 7. *Given an AT $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$ and two arguments A and B in T, argument A attacks argument B iff A undercuts, rebuts, or undermines B, where*

- *A undercuts B (on B') iff $\texttt{Conc}(A) \in \overline{n(r)}$ for some $B' \in \texttt{Sub}(B)$ such that $\texttt{TopRule}(B') = r$ is defeasible;*
- *A rebuts B (on B') iff $\texttt{Conc}(A) \in \bar{x}$ for some $B' = B_1, \dots, B_n \Rightarrow x \in \texttt{Sub}(B)$; and*
- *A undermines B (on x) iff $\texttt{Conc}(A) \in \bar{x}$ and $x \in \texttt{Prem}_\texttt{d}(B)$.*

Each AT gives rise to an argumentation framework (AF) [9], which is a pair of a set of arguments and attacks between these arguments.

Definition 8. *An argumentation framework (AF) is a pair $F = (\mathscr{A}, \mathscr{D})$ with $\mathscr{A}$ a set of arguments and $\mathscr{D} \subseteq \mathscr{A} \times \mathscr{A}$ an attack relation on F.*

Argumentation semantics drive argumentative reasoning in ASPIC$^+$ based on the associated AFs. We recall the prominent grounded semantics of AFs next.

Definition 9. *Given an AF $F = (\mathscr{A}, \mathscr{D})$, a set $\mathscr{E} \subseteq \mathscr{A}$ is conflict-free (in F) if there are no A, B in $\mathscr{E}$ such that $(A, B) \in \mathscr{D}$. The set of all conflict-free sets of F is denoted by $cf(F)$. We say that a set of arguments $\mathscr{B} \subseteq \mathscr{A}$ defends an argument $A \in \mathscr{A}$ if for each $(B, A) \in \mathscr{D}$ it holds that there is a $(C, B) \in \mathscr{D}$ with $C \in \mathscr{B}$. For an $\mathscr{E} \in cf(F)$, we define*

- *$\mathscr{E} \in adm(F)$ iff each $A \in \mathscr{E}$ is defended by $\mathscr{E}$;*
- *$\mathscr{E} \in com(F)$ iff $\mathscr{E} \in adm(F)$ and each A defended by $\mathscr{E}$ is in $\mathscr{E}$; and*
- *$\mathscr{E} \in grd(F)$ iff $\mathscr{E}$ is $\subseteq$-minimal complete in F.*

A set $\mathscr{E} \in \sigma(F)$, for $\sigma \in \{adm, com, grd\}$ is called a σ extension of F. It holds that there is a unique grounded extension in each AF [9]. Slightly abusing notation, we write $grd(F)$ to be the grounded extension.

An AT is associated to a unique AF, as specified in the following definition.

Definition 10. *Let $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$ be an AT. An AF $F = (\mathscr{A}, \mathscr{D})$ corresponds to T if $\mathscr{A}$ is the set of all arguments in T and $\mathscr{D}$ the defeat relation based on T.*

For ease of notation, for an AT T we define $\sigma(T) = \sigma(F)$ to be the set of extensions under σ of the AF F associated to T.

Example 2. *Given AT $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$ with $\mathscr{L}$ containing $\{a, b, c, d, e, f, d_1, d_2, d_3, d_4\}$ and their negations, and $\mathscr{R}$ containing the following four defeasible rules (with their names in front): $d_1 : a, b \Rightarrow \neg c$, $d_2 : d \Rightarrow c$, $d_3 : c \Rightarrow e$, $d_4 : f \Rightarrow a$. The set of strict rules is empty, i.e., $\mathscr{R}_s = \emptyset$. Finally, let $\mathscr{K} = \mathscr{K}_p \cup \mathscr{K}_n$, with $\mathscr{K}_n = \emptyset$ and $\mathscr{K}_p = \{b, d, f\}$. On the basis of the above, we can construct an associated AF $F = (\mathscr{A}, \mathscr{D})$ with the arguments:*

- $A_1 : b$
- $A_2 : d$
- $A_3 : f$
- $A_4 : A_2 \Rightarrow c$
- $A_5 : A_4 \Rightarrow e$
- $A_6 : A_3 \Rightarrow a$
- $A_7 : A_6, A_1 \Rightarrow \neg c$

The set of attacks is then $\mathscr{D} = \{(A_7, A_4), (A_4, A_7), (A_7, A_5)\}$. It holds that argument A_7 rebuts A_4 (and vice versa), and A_7 also rebuts A_5 (on the sub argument A_4). The complete extensions of this AF are $\{A_1, A_2, A_3, A_6\}$ (the grounded extension), $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ and $\{A_1, A_2, A_3, A_6, A_7\}$.

4. Integration of VBF into ASPIC$^+$

This section presents how VBF can function as a mechanism allowing for individualisation of arguments. That is, we develop here an approach that gives rise to what we call a *subjective* AT for an agent, given an AT and VBF; it can be understood as an AT filtered by values of an agent. For brevity, we usually assume a given VBF.

Let $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$ be an AT. As stated above, we assume that $\mathscr{L}$ is a set of literals that is closed under symmetric negation, where `Prop` is a set of positive literals.

Agents can associate values to defeasible elements (ordinary premises) in an AT. We discuss issues with values on non-defeasible elements (axioms) in Example 3. Let N be the image of n (i.e., the names of defeasible rules) and $H = \{head(r) \mid r \in \mathscr{R}_d\}$ the set of heads of defeasible rules. Let $Def = N \cup H \cup \mathscr{K}_p$ be the set of defeasible elements in the given AT. We assume that `Prop` $\subseteq Def$ to ensure that agents can have a value on defeasible elements. Moreover, we assume that Def is distinct from atoms that follow from axioms or strict rules. That is, $Def \cap (\mathscr{K}_n \cup \{head(r) \mid r \in \mathscr{R}_s\}) = \emptyset$.

4.1. Subjective Knowledge Bases and ATs

We are now going to define subjective knowledge bases and ATs, intuitively signaling what an agent is "prepared" to argue on, and what the agent finds in principle (relative to their values) unacceptable and does not desire to argue on. The basic idea is that an agent's values filter the positive literals in both the knowledge base and in defeasible rules.

Say we have a positive literal $p \in \mathcal{K}_p$ as an ordinary premise in the knowledge base of a given AT. If $p \in \texttt{PropBaseClean}_\alpha$, then the agent is prepared to argue on the basis of p, and no filtering is applied. If, on the other hand, $p \notin \texttt{PropBaseClean}_\alpha$, then the agent does not argue with p, and p will be effectively removed from $\mathcal{K}_p$; and, to express α's negative expression towards p, $\neg p$ is added to $\mathcal{K}_p$ as an ordinary premise. This ensures that α actively "argues" against use of p (though possibly other agents use p).

Towards our formalisation, contrary atoms are introduced using $\overline{\texttt{PropBaseClean}_\alpha}$:

$$\overline{\texttt{PropBaseClean}_\alpha} = \texttt{Prop} \setminus \texttt{PropBaseClean}_\alpha.$$

In other words, arguments can be formed against other arguments that do not pass the filter, meaning against propositions indicative of "antagonistic" values in some way or another (weight on value or relative to propositions).

The next definition expresses complementary atoms of those that do not pass the value filter of an agent.

Definition 11. *For a given VBF and agent α, we define $\texttt{CompsProps}_\alpha = \{\neg p \mid p \in \overline{\texttt{PropBaseClean}_\alpha}\}$.*

Intuitively speaking, $\overline{\texttt{PropBaseClean}}$ filters all atoms that are deemed simply unacceptable in principle, based on an agent's values. The set $\texttt{CompsProps}$ contains the negated (complementary) literals of all unacceptable atoms. Since $\texttt{PropBaseClean}_\alpha$ contains only positive atoms, we find that $\texttt{CompsProps}_\alpha$ only contains negated literals.

Definition 12. *Given a VBF, an agent α, and a knowledge base $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$, we define the subjective knowledge base of α as $\mathcal{K}_\alpha = \mathcal{K}_n \cup \mathcal{K}_p^\alpha$ with*

$$\mathcal{K}_p^\alpha = (\mathcal{K}_p \setminus \{\overline{\neg p} \mid \neg p \in \texttt{CompsProps}\}) \cup \{\neg p \mid \neg p \in \texttt{CompsProps}\}$$

This definition removes all ordinary premises p whenever $\neg p \in \texttt{CompsProps}$ (agent α discards p because of their values) and then also introduces $\neg p$ as an ordinary premise.

The subjective knowledge base of an agent α does not modify axioms (and later on our definition of subjective ATs does not alter strict rules). We give an example of issues that may arise otherwise.

Example 3. *Let $\mathcal{K}$ be a knowledge base, with $\mathcal{K}_n = \{a,b,c\}$ and $\mathcal{K}_p = \{d,e\}$. Suppose that agents can filter axioms, and say agent α permits use of $\{a,b,d\}$ while β argues on $\{b,c,e\}$ (these are their filtered set of atoms). Following the approach above modified for axioms, the individualised knowledge bases of both agents will look as follows.*

- $\mathcal{K}_p^\alpha = \{a,b,\neg c\}$, $\mathcal{K}_n^\alpha = \{d,\neg e\}$
- $\mathcal{K}_p^\beta = \{\neg a,b,c\}$, $\mathcal{K}_n^\beta = \{\neg d,e\}$

In such a case, the axioms of the agents, if seen collectively together, are inconsistent. In ASPIC$^+$ attacks on axioms are not possible. Thus, such conflicts between agents faces the issue of having conflicts not modelled as attacks.

Subjective ATs, which in particular require a filtering of rules, require more care. We filter defeasible rules whose (i) head is filtered out or (ii) whose name is filtered out.

Definition 13. *Given a VBF, an agent α, and a set of rules $\mathscr{R}$, the subjective set of defeasible rules $\mathscr{R}_d^\alpha$ of α is defined by*

$$\{r \in \mathscr{R}_d \mid body(r) \cup \{head(r)\} \subseteq \texttt{PropBaseClean}_\alpha, n(r) \in \texttt{PropBaseClean}_\alpha\}.$$

The set of subjective rules of α is then $\mathscr{R}^\alpha = \mathscr{R}_s \cup \mathscr{R}_d^\alpha$.

Note that $\mathscr{R}_s$ is not filtered with respect to an agent's values. In this definition, we tacitly assume that all defeasible rules have names, which, however is no obstacle for usage of the ASPIC$^+$ framework. We next define subjective ATs for an agent and VBF.

Definition 14. *Given a VBF and AT $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K})$, we define the subjective AT T_α, for an agent α, by $T_\alpha = (\mathscr{L}, \mathscr{R}^\alpha, n, ^-, \mathscr{K}_\alpha)$.*

Accordingly, for each subjective AT T_α, a subjective AF F_α can be constructed, by considering all arguments and attacks in T_α.

For reasoning, we make use of the grounded extension for each agent, i.e., $grd(F_\alpha)$ is the unique grounded extension for agent α. When clear from the context, we write G_α as the grounded extension of agent α.

4.2. Extended Example

Example 4. *Suppose a language $\mathscr{L}$ that contains $\{a, b, c, d, e, f, d_1, d_2, d_3, d_4\}$ and their complementary literals, two agents $\texttt{Agents} = \{\alpha, \beta\}$, and the following defeasible rules*

- $d_1 : a, b \Rightarrow c$,
- $d_2 : d \Rightarrow c$,
- $d_3 : c \Rightarrow e$, *and*
- $d_4 : f \Rightarrow a$.

We do not have strict rules in this instance, i.e., $\mathscr{R}_s = \emptyset$. Assume that the knowledge base is given by $\mathscr{K} = \mathscr{K}_n \cup \mathscr{K}_p$ in which $\mathscr{K}_p = \{b, d, f\}$ and $\mathscr{K}_n = \emptyset$.

Say, $\texttt{Prop} = \{a, b, c, d, e, f\}$ and assume that, after the filtering process, we get that $\texttt{PropBaseClean}$ of both agents has the following shapes:

- $\texttt{PropBaseClean}_\alpha = \{a, b, c, e, f\}$ *and*
- $\texttt{PropBaseClean}_\beta = \{a, b, c, d, e, f\}$.

In other words all atoms (from $\texttt{Prop}$ that are subject to values) but d are available for α, i.e., $\texttt{PropBaseClean}_\alpha = \texttt{Prop} \setminus \{d\}$, while all atoms are in $\texttt{PropBaseClean}_\beta$.

On the basis of the above, the subjective AT for agent α is AT $T_\alpha = (\mathscr{L}, \mathscr{R}_\alpha, n, ^-, \mathscr{K}_\alpha)$, where

- $\mathscr{K}_\alpha = \mathscr{K}_p^\alpha \cup \mathscr{K}_n$, *where $\mathscr{K}_p^\alpha = \{b, \neg d, f\}$ and $\mathscr{K}_n = \emptyset$, and*

- $\mathscr{R}_\alpha = \mathscr{R} \setminus \{d \Rightarrow c\}$.

Analogically, AT $T_\beta = (\mathscr{L}, \mathscr{R}_\beta, n, ^-, \mathscr{K}_\beta)$ is given by

- $\mathscr{K}_\beta = \mathscr{K}_p^\beta \cup \mathscr{K}_n$, *where* $\mathscr{K}_p^\beta = \{b, d, f\}$ *and* $\mathscr{K}_n = \emptyset$, *and*
- $\mathscr{R}_\beta = \mathscr{R}$.

On the basis of the above we can construct associated AFs (F_α and F_β) for both agents with the following arguments (see Table 1).

Table 1. Arguments of agents α and β

$\mathscr{A}_\alpha$	$\mathscr{A}_\beta$
$A_1 : b$	$A_1 : b$
$A_3 : f$	$A_2 : d$
$A_6 : A_3 \Rightarrow a$	$A_3 : f$
$A_7 : A_6, A_1 \Rightarrow c$	$A_4 : A_2 \Rightarrow c$
$A_8 : A_7 \Rightarrow e$	$A_5 : A_4 \Rightarrow e$
$A_9 : \neg d$	$A_6 : A_3 \Rightarrow a$
	$A_7 : A_6, A_1 \Rightarrow c$
	$A_8 : A_7 \Rightarrow e$

The set of attacks of both agents are empty $\mathscr{D}_\alpha = \mathscr{D}_\beta = \emptyset$ (the conflicts are between ATs from different agents). The complete extensions (grounded and preferred) of both frameworks contain all arguments in their sets $\{A_1, A_3, A_6, A_7, A_8, A_9\}$ for α and $\{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$ for β.

On the basis of that one can observe that even if no agent attacks their own arguments, some other agent might attack that agent's arguments, e.g., A_9 (of Agent α) attacks A_2 (of Agent β) (and vice-versa), A_4 and A_5. Conflict might arise due to an attack by another agent's argument. Intuitively, such a phenomenon can be interpreted as individualisation of conflicts, because agents may create different arguments based on their individualised knowledge bases. It is, to some extent similar to the model presented in [10], where defeat relations differ in various contexts, and discussed further below.

Having defined the reasoning of an individual agent, i.e., G_α, the next step is to define a collective outcome, according to a given (non-subjective) AT and VBF. To look at collectively acceptable arguments, under the grounded semantics, we consider the intersection of all G_α, of each agent α. When clear from the context, we denote this intersection by G^*. That is, given a VBF and AT

$$G^* = \bigcap_{\alpha \in \mathtt{Agents}} G_\alpha.$$

Example 5. *Let $T = (\mathscr{L}, \mathscr{R}, n, ^-, \mathscr{K}, \leq)$ be an AT with $\mathscr{L} = \{a, \neg a, b, \neg b, c, d, \neg c, \neg d\}$ and three strict rules $a \to \neg c$, $b \to \neg c$, and $c \to \neg d$. The set $\{a, b, c, d\}$ are the ordinary premises of this AT.*

Say we have two agents α and β. The first agent α has as their subjective knowledge base all atoms, except b, and, symmetrically, β does not include a. We implicitly assume a VBF matching these criteria.

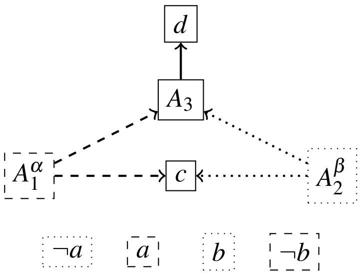

Figure 1. AT from Example 5, with arguments and attacks for agent α (dashed) and for agent β (dotted). The grounded extension for T_α is $\{a, \neg b, A_1^\alpha, d\}$ and for T_β is $\{b, \neg a, A_2^\beta, d\}$. The intersection of both grounded extensions is $\{d\}$. Note that both agents can construct A_3 based on their subjective knowledge base, i.e., $\{d\}$ is not admissible in neither T_α, T_β, nor T.

Except for the arguments corresponding to ordinary premises, we can construct the three arguments A_1^α consisting of ordinary premise a and claim ¬c, A_2^β based on b and claiming also ¬c, and A_3 based on c claiming ¬d. Figure 1 shows the associated AF.

We find that $\{a, \neg b, A_1^\alpha, d\}$ is the grounded extension for agent α, while $\{b, \neg a, A_2^\beta, d\}$ is the grounded extension of agent β. The intersection contains only d. This indicates that both agents accept d (under the cautious grounded semantics), while their reasons significantly differ. Indeed, they do not agree on the values which "justify" accepting d: α uses A_1^α, while β utilizes A_2^β, based on conflicting values.

We remark that the set of arguments $\{d\}$ (containing only the intersection of both grounded extensions G_α and G_β) is not admissible. In fact, $\{d\}$ is not admissible in the subjective ATs T_α and T_β, nor is this set admissible in the "original" AT.

The concept of collectively acceptable arguments can be related to human argumentation, when two (or more) agents agree on some arguments, but with different justifications, where they aim to satisfy different values; the agents disagree on other arguments.

4.3. Propositions

A direct result from the definition of subjective ATs (and subjective knowledge bases) is that arguments constructible in subjective ATs only contain atoms and rules that pass the value filter (or are not subject to filtering).

Proposition 1. *Let T be an AT, V a VBF, and α an agent in V. It holds that if A is an argument in T_α, then $\texttt{Prem}_\texttt{d}(A)$ and $\texttt{DefRules}(A)$ do not contain an element of $\mathit{Prop} \setminus \mathit{PropBaseClean}_\alpha$.*

This proposition follows from the definition of subjective ATs: any element of $\texttt{Prop} \setminus \texttt{PropBaseClean}_\alpha$ is removed from the subjective knowledge base and from the defeasible rules.

In a subjective AT T_α if $\neg p \in \texttt{CompsProps}_\alpha$, then $\neg p$ is in the grounded extension G_α. This follows by our construction of a T_α: p is effectively removed and $\neg p$ added as an ordinary premise (that is unattacked, by construction).

Proposition 2. *Let T be an AT, V a VBF, and α an agent in V. If $\neg p \in \texttt{CompsProps}_\alpha$, then $\neg p \in G_\alpha$.*

As shown in Example 5, the collective grounded extension is not admissible. However G^* is conflict-free (in the original AT and all subjective ATs).

Proposition 3. *Let T be an AT, V a VBF, and α an agent in V. It holds that*

- *G^* is in general not admissible in a T_α or T, and*
- *G^* is conflict-free in T_β and T, for any agent β.*

Proof of second item. Let G^* be the collective grounded extension. Suppose, for the sake of reaching a contradiction, that $A_1, A_2 \in G^*$ and that A_1 attacks A_2, in either T or T_β. First note that A_1 attacks A_2 in T iff A_1 attacks A_2 in T_β (for constructed arguments, attacks coincide on ATs and subjective ATs). If A_1 attacks A_2 in T_β, then G_β is not conflict-free, a contradiction. Thus, A_1 and A_2 are conflict-free in T. Thus, by the reasoning above, A_1 and A_2 are conflict-free in any T_β, for any agent β, a contradiction to the assumption that G^* is not conflict-free, in either T or T_β. □

5. Related Work

The role of values in argumentation has been discussed in many papers. One of the most influential approaches are Value-based Argumentation Frameworks (VAFs) [11], where the author extended Dung's abstract argumentation frameworks [9]. Values, in this work, are used to establish preferences between arguments. Moreover, [11] also introduces the concept of audience which is an ordering between values. This concept represents disagreement about preferences amongst values, which impact on argument attacks; this is similar, to some extent, to the concept of subjective ATs in VBF. The differences are, however, significant. Firstly, in VBF, values are used in structured, not abstract, models as in VAFs. Moreover, instead of preferences between values used to define audiences in [11], values in VBF are used as filters which remove atoms which do not fit to the agent's value profile. In VBF, values are private and inaccessible for others, while in [11] they are used explicitly to justify preferences. Approaches similar to VAFs appear in the argumentation schemes of Walton and papers on practical reasoning (e.g. [12,13]).

Another relevant approach to model value based argumentation in ASPIC+ has been presented in [10]. The authors deploy the concept of context, which is used to represents different parties of discourse. The context contains value-based orderings between norms. These orderings allow for differentiating parties (agents). [10] introduces also the concept of consensus which is similar to our concept of collectively accepted arguments. Apart from general similarities, there are significant differences between VBF and [10]. The most important is that our model does not use preferences. It is also important that, contrasting to [10], our model separates values from constructing the AT. Although values influence the construction of an AT, we cannot infer the values from the atoms without knowing `ValLimit` and `ValProp` functions; this would appear to correspond to human argumentation, where we do not necessarily know which values influenced an agent's arguments. In [10] values are explicitly involved in argumentation mechanism. The problem of the joint semantics of multiple agents has been discussed in the context of Multi-agent argumentation systems from various perspectives. Due to length limitations, we cannot introduce here a full discussion of this topic (see [10,14] for examples), which we discuss in a future work.

More broadly, the ArgMAS workshop series (2004-2013) considers works bearing a range of other topics than incorporating values in instantiated argumentation to yield agent-relative ATs, e.g., argumentation schemes, dialogue, abstract argumentation, belief revision, and multi-agent systems. In future, we aim to identify relevant literature.

This paper is rooted in the VBF framework [1,2] which introduces the model of reasoning with values. In contrast to other models, VBF does not represent values (or values' strength) as orderings, but introduces two functions `ValLimit` (`AgentValueToWeight` in older versions) and `ValProP` (previously `AgentValuePropWeight`); these two functions (separately for every agent and value) represent the minimal acceptable weight of a given value and the weight of a given value assigned to particular proposition, respectively. From these for each agent, we create the set `PropBaseClean`, which is the set of propositions which pass an agent's value-based filter. On the basis of such filtered sets we construct knowledge bases of agents, which is the basis of constructing ASPIC+ based ATs. What is important and different, in comparison to other models, is that values in VBF are used to shape the knowledge base and thence the construction of arguments, not to solve conflicts between arguments as in other approaches for abstract (e.g. [11]) and structured (e.g. [10]) argumentation.

6. Conclusions

This paper introduces a mechanism to incorporate VBF [1,2] in the ASPIC$^+$ argumentation framework. The framework allows for incorporating values into argumentation frameworks, which is distinct from [10] and VAFs [11]. The key point of VBF is in the individualisation of agents' knowledge bases, on the basis of which they create their own arguments. This results in individualisation of attack relations (attack relations can differ across the agents). Our model creates separate ATs for each agent. On the basis of that we construct the set of collectively accepted arguments G^* which is the intersection of all grounded extensions of all agents. We also briefly discuss some properties of collectively accepted arguments.

The most important difference between our model and other approaches to model value-based reasoning in argumentation is in the role of values. Instead of using orderings to justify preferences as in [11], we use the VBF mechanism to filter out the propositions which do not fit a given agent's value profile. Our model focuses on a different stage of the argumentation process; rather than deriving "winning" arguments *after all arguments have been constructed for all agents*, we focus on the construction of arguments for each agent and thence how such arguments are or are not accepted by each agent. Since our approach does not make use of preferential reasoning, there is a potential computational benefit as well, e.g., recent works found that incorporation of preferences to structured argumentation can lead to increased computational complexity [15].

Apart from formal issues discussed in previous sections, there are some intuitions relating to human argumentation which are addressed by our model:

- Values are an internal element of an agent's reasoning process, and they are not necessarily visible to others. Often we do no know what are the *real* intentions of adversaries, though we might try to infer them.
- Agents underlyingly use values to shape arguments and to introduce conflicts.

- An agent can proscribe the formation of a particular argument because some parts (i.e., literals amongst premises or conclusion) are *immoral* according to the agent.[5]
- Since some conflicts may be rooted in values, agents can differ not only in their knowledge bases but also in their understanding of which atoms and arguments are in conflict (see [1,2] for a discussion); negation is not the root of conflict.
- Agents can collectively accept some arguments, but from different underlying justifications, as agents can differ in how they weigh values and assess propositions.

For the future work, we aim to provide a more comprehensive analysis of the collective semantics of agents. In particular, we will discuss properties of collective extensions, for example, the admissibility of grounded and other extensions. We will consider how an agent's arguments can influence (persuade) another agent's relations to values and `PropBaseClean`. Finally, the domain of values has internal structure, which ought to be reflected in the formal analysis.

References

[1] Zurek T, Wyner A. Towards a Formal Framework for Motivated Argumentation and the Roots of Conflict. In: Grasso F, Green NL, Schneider J, Wells S, editors. Proceedings of the CMNA '22. vol. 3205 of CEUR Workshop Proceedings; 2022. p. 39-50.

[2] Wyner A, Zurek T. On Legal Teleological Reasoning. In: Proceedings of Jurix 2023; 2023. p. 83 -88.

[3] Modgil S, Prakken H. Abstract Rule-Based Argumentation. In: Baroni P, Gabbay D, Giacomin M, van der Torre L, editors. Handbook of Formal Argumentation. College Publications; 2018. p. 287-364.

[4] Baroni P, Gabbay D, Giacomin M, van der Torre L, editors. Handbook of Formal Argumentation. College Publications; 2018.

[5] Atkinson K, Bench-Capon TJM. Value-based Argumentation. FLAP. 2021;8(6):1543-88.

[6] Perelman C. Justice, law, and argument: Essays on moral and legal reasoning. vol. 142. Springer Science & Business Media; 2012.

[7] Schwartz S. An Overview of the Schwartz Theory of Basic Values. Online Readings in Psychology and Culture. 2012;2(1).

[8] Modgil S, Prakken H. A general account of argumentation with preferences. Artif Intell. 2013;195:361-97.

[9] Dung PM. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. Artif Intell. 1995;77(2):321-58.

[10] Yu Z, Ju S, Chen W. Context-based argumentation frameworks and multi-agent consensus building. Journal of Logic and Computation. 2023 10;34(2):199-228.

[11] Bench-Capon TJM. Persuasion in Practical Argument Using Value-based Argumentation Frameworks. J Log Comput. 2003;13(3):429-48.

[12] Atkinson K, Bench-Capon T. States, Goals and Values: Revisiting Practical Reasoning. Argument and Computation. 2016;7(2-3):135-54.

[13] Weide TL, Dignum F, Meyer JJC, Prakken H, Vreeswijk GAW. Practical Reasoning Using Values. In: McBurney P, Rahwan I, Parsons S, Maudet N, editors. Argumentation in Multi-Agent Systems. vol. 6057 of Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2010. p. 79-93.

[14] Kido H. Practical Argumentation Semantics for Pareto Optimality and Its Relationships with Values. In: McBurney P, Parsons S, Rahwan I, editors. Argumentation in Multi-Agent Systems. vol. 7543 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 28-45.

[15] Lehtonen T, Wallner JP, Järvisalo M. Computing Stable Conclusions under the Weakest-Link Principle in the ASPIC+ Argumentation Formalism. In: Proceedings of KR 2022; 2022. p. 215-25.

[5] Suppose an argument "In order to avoid overpopulation on the Earth, we should kill half of the human race". We can acknowledge the truth of both propositions and the rule, yet reject the legitimacy of the argument.