Complexity in Value-Based Argument Systems

Paul E. Dunne and Trevor Bench-Capon

Dept. of Computer Science
University of Liverpool
Liverpool L69 7ZF, United Kingdom
{ped,tbc}@csc.liv.ac.uk

Abstract. We consider a number of decision problems formulated in *value-based argumentation frameworks* (VAFs), a development of Dung's argument systems in which arguments have associated abstract values which are considered relative to the orderings induced by the opinions of specific audiences. In the context of a single fixed audience, it is known that those decision questions which are typically computationally hard in the standard setting admit efficient solution methods in the value-based setting. In this paper we show that, in spite of this positive property, there still remain a number of natural questions that arise solely in value-based schemes for which there are unlikely to be efficient decision processes.

1 Introduction

Argument systems as a model of defeasible reasoning date from the seminal paper of Dung [11], and have subsequently proved useful both to theorists who can use them as an abstract framework for the study and comparison of non-monotonic logics, e.g. [5, 7–9], and for those who wish to explore more concrete contexts where defeasibility is central e.g., for the legal domain, [2], [18], and [15].

In many domains, especially those relating to practical reasoning, such as law, politics and ethics, however, it is not possible to demonstrate the acceptability of an argument absolutely. Which arguments are found persuasive depends on the opinions, values and, perhaps, even the prejudices of the audience to which they are addressed. The point is made by Perelman [17] thus:

> If men oppose each other concerning a decision to be taken, it is not because they commit some error of logic or calculation. They discuss *apropos* the applicable rule, the ends to be considered, the meaning to be given to values, the interpretation and characterisation of facts.

What this means is that because people may differ as to what they hold to be important or worth attempting to achieve, they may differ in their evaluations of the strengths of the arguments about the choices that should be made. For example: it is an argument in favour of raising income tax that it promotes equality, and for decreasing income tax that it promotes enterprise. Most people would acknowledge that both arguments are valid, but which they will choose to follow depends on the importance they ascribe to equality as against enterprise in the given situation. Thus which of the arguments is found persuasive by a given audience will depend on the ordering of these two values by that audience. So if I claim that income tax should be raised to promote equality, I can defend my claim against the attack that so doing would discourage enterprise not by attacking this counter argument but by declaring my preference for equality over enterprise. Whether this defence will be persuasive will depend on whether the audience shares my preference. A similar viewpoint has been proposed in philosophical studies of rational and persuasive argument, e.g from [19, p. xv]

> Assume universally valid and accepted standards of rationality. Assume perfectly rational agents operating with perfect information, and you will find that rational disagreement will still occur; because, for example, the rational agents are likely to have different and inconsistent values and interests, each of which may be rationally acceptable.

In order to reason about arguments dependent on values and to accommodate the notion of audiences with different values, Dung's original framework was extended in [3] and [4] to give what are termed there Value Based Argumentation Frameworks (VAFs). In those papers a number of properties of VAFs are demonstrated. In this paper we will consider some questions relating to the computational complexity of these frameworks.

Section 2 will provide the definitions of Argumentation Systems and Value Based Argumentation Frameworks, and of the decision problems we will address. Section 3 will present the proofs of our results, and section 4 will offer some discussion and concluding remarks.

2 Basic Definitions

The basic definition below of an Argument System is derived from that given in [11].

Definition 1 *An* argument system *is a pair $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$, in which $\mathcal{X}$ is a finite set of* arguments *and $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ is the* attack relationship *for $\mathcal{H}$. A pair $\langle x, y \rangle \in \mathcal{A}$ is referred to as 'y* is attacked by *x' or 'x* attacks *y'. For R, S subsets of arguments in the system $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$, we say that*

a. $s \in S$ *is* attacked *by R if there is some $r \in R$ such that $\langle r, s \rangle \in \mathcal{A}$.*
b. $x \in \mathcal{X}$ *is* acceptable with respect to *S if for every $y \in \mathcal{X}$ that attacks x there is some $z \in S$ that attacks y.*
c. S is conflict-free *if no argument in S is attacked by any other argument in S.*
d. A conflict-free set S is admissible *if every argument in S is acceptable with respect to S.*
e. S is a preferred extension *if it is a maximal (with respect to $\subseteq$) admissible set.*
f. S is a stable extension *if S is conflict free and every argument $y \notin S$ is attacked by S.*
g. $\mathcal{H}$ *is* coherent *if every preferred extension in $\mathcal{H}$ is also a stable extension.*

An argument x is credulously accepted *if there is* some *preferred extension containing it; x is* sceptically accepted *if it is a member of* every *preferred extension.*

Abstracting away concerns regarding the internal structure and representation of arguments affords a formalism which focuses on the relationship between individual arguments as a means of defining divers ideas of acceptance. In particular preferred extensions are of interest as these represent maximal coherent positions that can be defended against all attackers.

While this approach offers a powerful tool for the abstract analysis of defeasible reasoning, there are, however, several potential problems. While every argument system has *some* preferred extension, this may simply be the empty set of arguments and although the use of *stable* extensions avoids such difficulties these in turn have the drawback that there are systems which contain no stable extension. An additional concern is the computational complexity of a number of the associated decision problems that has been shown to range from NP–complete to Π_2^p–complete. A summary of these is given in Table 1 below. The classification of problems (3–5) follows from [10]; that of (6) and

	Problem	Decision Question	Complexity
1	$\text{ADM}(\mathcal{H}, S)$	Is S admissible?	P
2	$\text{STAB}(\mathcal{H}, S)$	Is S stable?	P
3	$\text{PREF-EXT}(\mathcal{H}, S)$	Is S *preferred*?	CO-NP-complete.
4	$\text{CA}(\mathcal{H}, x)$	Is x in a *preferred* S?	NP-complete
5	$\text{STAB-EXIST}(\mathcal{H})$	Has $\mathcal{H}$ a stable extension?	NP-complete
6	$\text{SA}(\mathcal{H}, x)$	Is x in *every* preferred S?	$\Pi_2^{(p)}$-complete
7	$\text{COHERENT}(\mathcal{H})$	Preferred$\equiv$stable?	$\Pi_2^{(p)}$-complete

Table 1. Decision Problems in Argument Systems

(7) has recently been demonstrated in [12]. Related problems arise with proof-theoretic mechanisms for establishing credulous acceptance, e.g. for the TPI-dispute mechanism proposed in [20], Dunne and Bench-Capon [13] show that this defines a weak propositional proof system under which proofs that arguments are not credulously accepted require exponentially many steps.

While the issues discussed above concern algorithmic and combinatorial properties of the standard argument system framework, there is also one interpretative issue of some importance. A typical argument system may contain many distinct preferred extensions and, in some cases, two different preferred extensions may define a partition of the argument set. Thus a single argument system can give rise to a number of disjoint internally consistent admissible argument sets. The abstract level at which Dung's formalism operates avoids any mechanism for distinguishing notions of the relative merit of such mutually incompatible outcomes. Thus the situation arises in which we appear to have several coherent positions that could be adopted, and no well motivated way of choosing between them.

Recognising the benefits of Dung's approach, a number of extensions have been mooted in order to ameliorate the various interpretative and computational difficulties outlined above. Among such are the *preference-based argumentation frameworks* (PAFs) of Amgoud and Cayrol [1] and, the formalism with which the present article is concerned, the *value-based argumentation frameworks* (VAFs) of Bench-Capon [3, 4]. It is important to note that while there are some superficial similarities, these two mechanisms are, in fact, quite distinct. We shall defer a more detailed comparison until the concluding section, since this distinction can be rather more readily discerned and appreciated in the light of our subsequent technical results.

As we have indicated, [3, 4] extend Dung's framework to provide a semantics for distinguishing and choosing between consistent but incompatible belief sets through the use of *argument values*. Thus arguments are seen as grounded on one of a finite number of abstract values and the interpretation of which of a set of arguments to "accept" is treated in terms of preference orderings of the underlying value set according to the views held by a particular *audience*. Thus while in the standard Argumentation system the choice between preferred extensions is arbitrary, in a VAF we are able to motivate such choices by reference to the values of the audience. The formal definition of such *value-based argumentation frameworks* is given below.

Definition 2 *A* value-based argumentation framework *(VAF), is defined by a triple* $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), \mathcal{V}, \eta \rangle$*, where* $\mathcal{H}(\mathcal{X}, \mathcal{A})$ *is an argument system,* $\mathcal{V} = \{v_1, v_2, \ldots, v_k\}$ *a set of* k values*, and* $\eta : \mathcal{X} \to \mathcal{V}$ *a mapping that associates a value* $\eta(x) \in \mathcal{V}$ *with each argument* $x \in \mathcal{X}$*. An* audience*,* α*, for a VAF* $\langle \mathcal{H}, \mathcal{V}, \eta \rangle$*, is a total ordering of the values* $\mathcal{V}$*. We say that* v_i is preferred to v_j in the audience α*, denoted* $v_i \succ_\alpha v_j$*, if* v_i *is ranked higher than* v_j *in the total ordering defined by* α*.*

Using VAFs, ideas analogous to those of admissible argument in standard argument systems are defined in the following way. Note that all these notions are now relative to some audience.

Definition 3 *Let* $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), \mathcal{V}, \eta \rangle$ *be a VAF and* α *an audience.*

a. *For arguments* x*,* y *in* $\mathcal{X}$*,* x *is a* successful attack *on* y *(or* x defeats y*)* with respect to the audience α *if:* $\langle x, y \rangle \in \mathcal{A}$ and *it is* not *the case that* $\eta(y) \succ_\alpha \eta(x)$*.*
b. *An argument* x *is* acceptable to the subset S *with respect to an audience* α *if: for every* $y \in \mathcal{X}$ *that* successfully attacks x *with respect to* α*, there is some* $z \in S$ *that successfully atttacks* y *with respect to* α*.*
c. *A subset* R *of* $\mathcal{X}$ *is* conflict-free with respect to the audience α *if: for each* $\langle x, y \rangle \in R \times R$*, either* $\langle x, y \rangle \notin \mathcal{A}$ *or* $\eta(y) \succ_\alpha \eta(x)$*.*
d. *A subset* R *of* $\mathcal{X}$ *is* admissible *with respect to the audience* α *if:* R *is conflict free with respect to* α *and every* $x \in R$ *is acceptable to* R *with respect to* α*.*
e. *A subset* R *is a* preferred extension *for the audience* α *if it is a maximal admissible set with respect to* α*.*
f. *A subset* R *is a* stable extension *for the audience* α *if* R *is admissible with respect to* α *and for all* $y \notin R$ *there is some* $x \in R$ *which successfully attacks* y*.*

A standard consistency requirement which we assume of the VAFs considered is that every directed cycle of arguments in these contains *at least two* differently valued arguments. We do not believe that this condition is overly restricting, since the existence of such cycles in VAFs can be seen as indicating a flaw in the formulation of the framework. While in standard argumentation frameworks cycles arise naturally, especially if we are dealing with uncertain or incomplete information, in VAFs odd length cycles in a single value represent paradoxes and even length cycles in a single value can be reduced to a self-defeating argument. Given the absence of cycles in a single value the following important property of VAFs and audiences was demonstrated in [3].

Fact 1 *For every audience, α, $\langle \mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle), \mathcal{V}, \eta \rangle$ has a unique non-empty preferred extension, $P(\mathcal{H}, \eta, \alpha)$ which can be constructed by an algorithm that takes $O(|\mathcal{X}| + |\mathcal{A}|)$ steps. Furthermore $P(\mathcal{H}, \eta, \alpha)$ is a stable extension with respect to α.*

From Fact 1 it follows that, when attention is focused on one specific audience, the decision questions analogous to those described in Table 1 become much easier. There are, however, a number of new issues that arise in the value-based framework from the fact that that the relative ordering of different values promoted by distinct audiences results in arguments falling into one of three categories.

C1. Arguments, x, that are in the preferred extension $P(\mathcal{H}, \eta, \alpha)$ for some audiences but not all. Such arguments being called *subjectively acceptable*.
C2. Arguments, x, that are in the preferred extension $P(\mathcal{H}, \eta, \alpha)$ for *every* audience. Such arguments being called *objectively acceptable*.
C3. Arguments, x, that do not belong to the preferred extension $P(\mathcal{H}, \eta, \alpha)$ for *any* choice of audience. Such arguments being called *indefensible*.

To show the advantages of taking values into account, consider the following ethical debate, discussed in, e.g. [6]. Hal, a diabetic, loses his insulin and can save his life only by breaking into the house of another diabetic, Carla, and using her insulin. We may consider the following arguments:

A. Hal should not take Carla's insulin as he may be endangering her life.
B. Hal can take the insulin as otherwise he will die, whereas there is only a potential threat to Carla.
C. Hal must not take Carla's insulin because it is Carla's property.
D. Hal must replace Carla's insulin once the emergency is over.

Now B attacks A, since the permission licensed by the actual threat overrides the obligation arising from the potential threat. A does not attack B, since the immediate threat represents an exception to the general rule which A instantiates. C attacks B, construing property rights as strict obligations whereas possible endangerment is a defeasible obligation. D attacks C. since it provides a way for the insulin to be taken whilst property rights are respected. Further, Christie argues in [6] that B attacks D, since even if Hal were unable to replace the insulin he would still be correct to act so as to save his life, and therefore he can be under no strict obligation to replace the insulin. The resulting argumentation system can be depicted as a directed graph as shown in Figure 1.

Considered as a standard Argumentation System there is no non-empty preferred extension, and it seems we have no coherent position, which is why it is seen and discussed as an ethical dilemma. If, however, we consider it as a VAF, we can see that arguments A and B rely on the importance of preserving life, whereas C and D depend on respect for property. We will now have two preferred extensions, depending on whether life or property is preferred. If we prefer life, we will accept $\{B, C\}$: whilst we respect Carla's property rights, we regard Hal's need as paramount. In contrast if we prefer property to life, the preferred extension is $\{B, D\}$: the property claim can

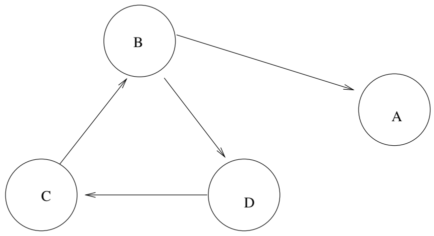

Fig. 1. VAF Example Argument System

be discharged if restitution is made. Thus B is objectively acceptable, C and D are subjectively acceptable and A is indefensible. This small example shows how we can use explicit value preferences to cut through what would otherwise be an irresolvable dilemma.

Our initial concern, in this article, is to consider the following questions specific to the value-based setting

Definition 4 *The decision problem* Subjective Acceptance *(*SBA*) takes as an instance a VAF $\langle \mathcal{H}, \mathcal{V}, \eta \rangle$ and an argument x in this. The instance is accepted if there is at least one audience, α, for which $x \in P(\mathcal{H}, \eta, \alpha)$. The decision problem* Objective Acceptance *(*OBA*) takes as an instance a VAF $\langle \mathcal{H}, \mathcal{V}, \eta \rangle$ and an argument x in this. The instance is accepted if $x \in P(\mathcal{H}, \eta, \alpha)$ for every audience α.*

The status of these problems was left unresolved in the original study presented in [3].

The remainder of this paper is structured as follows. In the next section we consider the decision problems SBA and OBA defined above and classify their complexity as, respectively, NP–complete and CO-NP–complete. We then consider decision problems motivated by approaches to determining subjective acceptance by attempting to identify which pair-wise orderings are "critical" in the sense that a given ordering will admit an audience for which an argument is subjectively accepted, whereas reversing this order will yield a context within which the argument of interest is never accepted. We show that the decision problem formulated with respect to this question is D^p-complete and hence within a class that is "harder" than either NP or CO-NP. Discussion and conclusions are the subject of the final section.

3 Complexity of Decision Problems for VAFs

Theorem 1 SBA *is* NP*–complete.*

Proof. For membership in NP simply use the algorithm which non-deterministically chooses an audience α from the $k!$ available and then tests if $x \in P(\mathcal{H}, \eta, \alpha)$, the latter test being accomplished by a polynomial-time algorithm. To prove that SBA is NP–hard we use a reduction from 3-SAT. Given an instance $\Phi(Z_n) = \wedge_{i=1}^{m} (y_{i,1} \vee y_{i,2} \vee y_{i,3})$ of this we construct a VAF $\langle \mathcal{H}_\Phi, \mathcal{V}_\Phi, \eta \rangle$ and argument x that is subjectively acceptable in $\langle \mathcal{H}_\Phi, \mathcal{V}_\Phi, \eta \rangle$ if and only if $\Phi(Z_n)$ is satisfiable. The framework uses $4n+m+1$ arguments which we denote $\{\Phi, C_1, \ldots, C_m\} \cup \cup_{i=1}^{n}\{p_i, q_i, r_i, s_i\}$. The relationship $\mathcal{A}$ contains attacks $\langle C_j, \Phi\rangle$ for each $1 \leq j \leq m$ and attacks $\{\langle p_i, q_i\rangle, \langle q_i, r_i\rangle, \langle r_i, s_i\rangle, \langle s_i, p_i\rangle\}$ for each $1 \leq i \leq n$. The remaining attacks in $\mathcal{A}$ are as follows. For each clause $y_{i,1} \vee y_{i,2} \vee y_{i,3}$ of $\Phi(Z_n)$ if $y_{i,j}$ is the literal z_k, the attack $\langle p_k, C_i\rangle$ is included in $\mathcal{A}$; if $y_{i,j}$ is the literal $\neg z_k$, then the attack $\langle q_k, C_i\rangle$ is added. The final part of the construction is to describe the value set $\mathcal{V}_\Phi$ and association of arguments with values prescribed by η. The set $\mathcal{V}_\Phi$ contains $2n+1$ values $\{con\} \cup \cup_{i=1}^{n}\{pos_i, neg_i\}$ and the mapping η assigns the value con to Φ and each argument in $\{C_1, \ldots, C_m\}$. Finally the arguments $\{p_i, r_i\}$ are mapped to the value pos_i and the arguments $\{q_i, s_i\}$ to the value neg_i. To complete the instance we set x to be Φ. We note that the constructed system satisfies the requirement that all cycles contain at least two distinct values.

We claim that Φ is subjectively accepted in the VAF $\langle \mathcal{H}_\Phi, \mathcal{V}_\Phi, \eta\rangle$ if and only if $\Phi(Z_n)$ is satisfiable. Suppose first that $\Phi(Z_n)$ is satisfied by an instantiation $\langle a_1, a_2, \ldots, a_n\rangle$ of Z_n. Consider any audience α for which $pos_i \succ_\alpha neg_i$ if $a_i = \top$, $neg_i \succ_\alpha pos_i$ if $a_i = \bot$, and $v \succ_\alpha con$ for all $v \in \mathcal{V}_\Phi/\{con\}$. Since $\Phi(Z_n)$ is satisfied, for each C_i there is some literal $y_{i,j}$ that is assigned $\top$ in the instantiation $\langle a_1, \ldots, a_n\rangle$. Consider the arguments $\{p_k, q_k, r_k, s_k\}$ for which $y_{i,j} \in \{z_k, \neg z_k\}$. If $y_{i,j} = z_k$ then p_k is acceptable in $\{p_k, r_k\}$ and, in addition, p_k successfully attacks C_i; if $y_{i,j} = \neg z_k$ then q_k is acceptable in $\{q_k, s_k\}$ and, again, successfully attacks C_i. Thus every argument C_i is successfully attacked by an argument p_k or q_k and thence Φ together with these form an admissible set. Thus we have an audience with respect to which Φ is subjectively accepted.

On the other hand, suppose α is an audience for which $\Phi \in P(\mathcal{H}_\Phi, \eta, \alpha)$. It cannot be the case that $C_i \in P(\mathcal{H}_\Phi, \eta, \alpha)$ since $\eta(\Phi) = \eta(C_i) = con$ and so the presence of any C_i would suffice to eliminate Φ. The audience α must therefore be such that every C_i is successfully attacked by one of its three possible attackers. Let $\langle t_1, t_2, \ldots, t_m\rangle$ be the choices which give successful attacks on $\langle C_1, \ldots, C_m\rangle$. First observe that we cannot have $t_i = p_k$ and $t_j = q_k$ for any $1 \leq k \leq n$ and distinct C_i and C_j: under α either $\eta(p_k) \succ_\alpha \eta(q_k)$ and so q_k would not succeed in its attack or $\eta(q_k) \succ_\alpha \eta(p_k)$ with the attack by p_k failing. It follows that the instantiation of Z_n by $z_i = \top$ if $p_i \in \langle t_1, t_2, \ldots, t_m\rangle$, $z_i = \bot$ if $q_i \in \langle t_1, t_2, \ldots, t_m\rangle$ is well-defined and yields a true literal in every clause, i.e. results in a satisfying instantiation of $\Phi(Z_n)$. This suffices to complete the proof.

The structure introduced in the proof of Theorem 1 provides the basis for developing a number of our subsequent results. Thus,

Theorem 2 OBA *is* CO-NP*–complete.*

Proof. Membership is CO-NP follows by the algorithm which tests that all $k!$ audiences accept x. For CO-NP–hardness, we employ a reduction from 3-UNSAT, the problem of deciding if a 3-CNF formula $\Phi(Z_n) = \wedge_{i=1}^{m}(y_{i,1} \vee y_{i,2} \vee y_{i,3})$ is unsatisfiable. The reduction constructs an identical VAF to that of the previous theorem, but with one additional argument, $\{test\}$, having $\eta(test) = con$ and whose sole attacker is the argument Φ. We claim that $test$ is objectively acceptable if and only if Φ is unsatisfiable. From the proof of Theorem 1, $test$ will fail to be acceptable with respect to any audience α for which Φ is admissible. Such an audience exists if and only if $\Phi(Z_n)$ is satisfiable. We therefore deduce that $test$ is objectively accepted if and only if $\Phi(Z_n)$ is unsatisfiable.

In applying the value-based framework to promote a particular argument an important consideration is the relationship between the value of the argument defended to that of the other values employed in the system. Thus the existence of an audience that provides subjective acceptance may depend on the specific ordering of a subset of the values. Thus, an audience granting subjective acceptance of some x exists if $v_i \succ v_j$, but should an audience prescribe $v_j \succ v_i$ then x cannot be defended. For example in law we find that values are ranked differently in different jurisdictions. Can we determine which difference affect the status of a given argument? Such scenarios motivate the concept of a *critical pair*.

Definition 5 *Let $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), \mathcal{V}, \eta \rangle$ be a VAF, and $\langle v, v' \rangle$ be an ordered pair of distinct values from $\mathcal{V}$. The pair $\langle v, v' \rangle$ is* critical *with respect to an argument $x \in \mathcal{X}$ if there is an audience α for which $v \succ_\alpha v'$ and $x \in P(\mathcal{H}, \eta, \alpha)$, but for every audience β for which $v' \succ_\beta v$ it holds that $x \notin P(\mathcal{H}, \eta, \beta)$.*

We can, of course, generalise the notion of critical pair from Definition 5 to encompass relationships involving three or more values.

Definition 6 *Let $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), \mathcal{V}, \eta \rangle$ be a VAF, $\mathcal{W} \subseteq \mathcal{V}$ and $\mathcal{C} = \{c_1, c_2, \ldots, c_r\}$ a finite set of* constraints *on $\mathcal{W}$ each of which is of the form $w \succ w'$ for $\{w, w'\} \subset \mathcal{W}$. The structure $\langle \mathcal{W}, \mathcal{C} \rangle$ is a* critical set *for an argument $x \in \mathcal{X}$ if there is an audience α under which $w \succ_\alpha w'$ for each $c_i \in \mathcal{C}$ and with $x \in P(\mathcal{H}, \eta, \alpha)$, but for any audience β in which at least one constraint of $\mathcal{C}$ is broken it holds that $x \notin P(\mathcal{H}, \eta, \beta)$.*

From these we define the following decision problems.

Definition 7 *The decision problem* Critical Pair *(*CP*) takes as an instance a triple of the form $\langle \langle \mathcal{H}, \mathcal{V}, \eta \rangle, \langle v, v' \rangle, x \rangle$ comprising a VAF, ordered pair of values $\langle v, v' \rangle$ within this, and argument x. An instance is accepted if $\langle v, v' \rangle$ is critical with respect to x. The decision problem* Critical Set *(*CS*) takes as its instance a quadruple $\langle \langle \mathcal{H}, \mathcal{V}, \eta \rangle, \mathcal{W}, \mathcal{C}, x \rangle$ of VAF, subset of values, set of constraints, and argument. The instance is accepted if $\langle \mathcal{W}, \mathcal{C} \rangle$ defines a critical set for x.*

Noting that the problem CP is a restricted case of CS, we can establish a lower bound on the complexity of the latter by considering the complexity of CP only. Under the standard complexity-theoretic assumptions this problem turns out to be "more difficult" than Subjective Acceptance and Objective Acceptance. Formally we show that it is D^p–complete, the class D^p comprising those languages L formed by the intersection of some NP language L_1 with some CO-NP language L_2, i.e. $L \in \mathrm{D}^p$ if $L = L_1 \cap L_2$, $L_1 \in$ NP, and $L_2 \in$ CO-NP. The class D^p was introduced in [16] where the problem SAT-UNSAT – instances of pairs of formulae $\langle \Phi_1, \Phi_2 \rangle$ for which Φ_1 is satisfiable and Φ_2 unsatisfiable – was shown to be D^p–complete.

Theorem 3 CP *is D^p–complete.*

Proof. For membership in D^p, define the language L_1 to be

$$\{\langle \langle \mathcal{H}, \mathcal{V}, \eta \rangle, \langle v, v' \rangle, x \rangle \ : \ \exists \alpha \text{ with } v \succ_\alpha v' \text{ and } x \in P(\mathcal{H}, \eta, \alpha)\}$$

Similarly, define L_2 as

$$\{\langle\langle\mathcal{H},\mathcal{V},\eta\rangle,\langle v,v'\rangle,x\rangle \;:\; \forall\alpha \text{ with } v' \succ_\alpha v,\, x \notin P(\mathcal{H},\eta,\alpha)\}$$

Then $\langle\langle\mathcal{H},\mathcal{V},\eta\rangle,\langle v,v'\rangle,x\rangle$ is accepted as an instance of CP if and only if it belongs to the set $L_1 \cap L_2$. Since it is immediate that $L_1 \in$ NP and $L_2 \in$ CO-NP this suffices to give CP $\in$ D^p. To prove that CP is D^p–hard, we first show that the problem *Critical Variable* (CV) is D^p–hard: instances of this comprise a CNF formula $\Phi(Z_n)$ and a variable $z \in Z_n$ with instances accepted if there is a satisfying instantiation in which $z = \top$ but no satisfying instantiation in which $z = \bot$. To see that CV is D^p–hard we use a reduction from the D^p–complete problem SAT-UNSAT. Given an instance $\langle\Phi_1(Z_n),\Phi_2(Z_n)\rangle$ of this, the instance $\langle\Psi,z\rangle$ of CV is simply $\langle(\neg z\vee\Phi_1)\wedge(z\vee\Phi_2),z\rangle$ where z is a new variable. We note that for Φ in CNF, $z\vee\Phi$ translates to the CNF formula in which every clause C of Φ is replaced by the clause $z\vee C$. It is easy to see that $\langle(\neg z\vee\Phi_1)\wedge(z\vee\Phi_2),z\rangle$ is a positive instance of CV if and only if $\langle\Phi_1(Z_n),\Phi_2(Z_n)\rangle$ is a positive instance of SAT-UNSAT: if Φ_1 is satisfiable then $(\neg z \vee \Phi_1) \wedge (z \vee \Phi_2)$ has a satisfying instantiation with $z = \top$ since it reduces to Φ_1; if Φ_2 is unsatisfiable then there is no satisfying instantiation with $z = \bot$ since the formula now reduces to Φ_2, hence if $\langle\Phi_1,\Phi_2\rangle$ accepted as an instance of SAT-UNSAT then $\langle(\neg z\vee\Phi_1)\wedge(z\vee\Phi_2),z\rangle$ is accepted as an instance of CV. Similarly, if $\langle(\neg z \vee \Phi_1) \wedge (z \vee \Phi_2),z\rangle$ is a positive instance of CV then $(\neg z \vee \Phi_1) \wedge (z \vee \Phi_2)$ is satisfiable when $z = \top$, i.e. Φ_1 is satisfiable, and $(\neg z \vee \Phi_1) \wedge (z \vee \Phi_2)$ is unsatisfiable when $z = \bot$, i.e. Φ_2 is unsatisfiable.

The proof that CP is D^p–hard now follows easily, using the reduction of Theorem 1: given an instance $\langle\Phi(Z_n),z\rangle$ of CV form the VAF $\langle\mathcal{H}_\Phi,\mathcal{V}_\Phi,\eta\rangle$ described in the proof of Theorem 1 (where we note that this trivially extends to arbitrary CNF formulae). Set the value pair in the instance of CP to be $\langle pos_z, neg_z\rangle$ and the argument x to be Φ. Consider the resulting instance $\langle\langle\mathcal{H}_\Phi,\mathcal{V}_\Phi,\eta\rangle,\langle pos_z,neg_z\rangle,\Phi\rangle$. If it is a positive instance of CP then there is an audience α within which $\Phi \in P(\mathcal{H}_\Phi,\eta,\alpha)$ and $pos_z \succ_\alpha neg_z$: it has already been seen that this indicates $\Phi(Z_n)$ has a satisfying instantiation with $z = \top$. Similarly, if it is a positive instance of CP, then $\Phi \notin P(\mathcal{H}_\Phi,\eta,\alpha)$ for any audience within which $neg_z \succ_\alpha pos_z$ so that, from our earlier analysis, $\Phi(Z_n)$ has no satisfying instantiation with $z = \bot$. On the other hand should $\langle\Phi(Z_n),z\rangle$ be a positive instance of CV then the argument of Theorem 1 yields an audience α with $pos_z \succ_\alpha neg_z$ for which $\Phi \in P(\mathcal{H}_\Phi,\eta,\alpha)$ from a satisfying instantiation of $\Phi(Z_n)$ with $z = \top$ while the unsatisfiability of $\Phi(Z_n)$ when $z = \bot$ indicates that no audience α having $neg_z \succ_\alpha pos_z$ will result in $\Phi \in P(\mathcal{H}_\Phi,\eta,\alpha)$. We deduce that $\langle\Phi(Z_n),z\rangle$ is a positive instance of CV if and only if $\langle\langle\mathcal{H}_\Phi,\mathcal{V}_\Phi,\eta\rangle,\langle pos_z,neg_z\rangle,\Phi\rangle$ is a positive instance of CP, thereby establishing that CP is D^p–complete.

4 Discussion

Preferences and Values

We referred earlier to the preference-based formalism (PAFs) of Amgoud and Cayrol [1] as another approach to developing Dung's abstract framework in order to enrich the idea of one argument "defeating" another. Formally, [1] defines a PAF as a pair

$\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), Pref \rangle$ wherein $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is a standard argument system and *Pref* a binary preorder (i.e. reflexive and transitive) relation on the argument set $\mathcal{X}$. The property that "*x successfully attacks y*"[1], then holds if $\langle x, y \rangle \in \mathcal{A}$ *and* $\neg(\langle y, x \rangle \in Pref)$, that is: a (potential) attack exists ($\langle x, y \rangle \in \mathcal{A}$) and it is not the case that the *argument y* is preferred to the *argument x*, i.e. $\neg(\langle y, x \rangle \in Pref)$. Compare this with the VAF notion of "successful attack": $\langle x, y \rangle \in \mathcal{A}$ and $\neg(\eta(y) \succ_\alpha \eta(x))$, i.e. a (potential) attack exists (as before) and *with respect to the audience* α it is not the case that *the value promoted by y* is preferred to the *value promoted by x*. Now, despite the superficial "syntactic" similarity that is present, these two approaches have significant differences. Certainly it is the case that one could describe a particular VAF instantiation by a PAF: the consequent *Pref* relationship, however, would relate only to a fixed audience α. Similarly one could model a given PAF by a VAF simply by assuming that the number of distinct values is $|\mathcal{X}|$ and considering an audience defined in a manner consistent with the preference relationship, i.e. so that $\eta(x) \succ_\alpha \eta(y)$ if $\langle x, y \rangle \in Pref$.

Of course, the fact that one may mutually relate the formal models of both systems, is very far from providing a credible case for the redundancy of either. If we consider the basic schemata of standard argument systems, PAFs and VAFs, in very informal terms one might view Dung's systems as VAFs in which a single value is present and PAFs as one in which $|\mathcal{X}|$ values are related (within *Pref*) to the views of a single audience. It has been observed in [3, 4] that typically the number of values is "small" by comparison with the number of arguments and thus the interpretation of PAFs within a value context as we have outlined does not reflect this. We also observe that the problems of subjective and objective acceptance in VAFs, arising as they do from questions concerning properties of *audiences* with respect to a set of values, do not have "sensible" counterparts in the PAF context. For example, were one to consider "preference-based subjective acceptance" in PAFs as "does there exists a preference relation on $\mathcal{X}$ under which a given argument x is accepted?", then it is bordering on the facile to observe that this question is of minimal interest: every instance returns the answer true by virtue of any preference relation under which x is a maximal element. We have seen that this is far from being the case as regards "value-based subjective acceptance", cf. Theorem 1. This is because, in VAFS, the strengths of arguments are not independent of one another. Thus raising the priority of one argument will raise the priority of all other arguments associated with the same value. In particular, if an argument is attacked by an argument associated with the same value, that attack will succeed, if the attacker, is not itself defeated, whatever the relative rank of the associated value.

In total these two formalisms although deriving from similar motivating factors – extending the concept of "acceptability" to take into account subjective relationships – take quite distinct approaches: PAFs by "embedding" a *single* preferrence relation within an argument framework; VAFs by an abstract association of values with arguments with the ordering of these being a feature "external" to the framework itself.

[1] [1] employs the term "defeats" rather than "attacks"

Summary and Further Work

The above results show that the identification of an argument as subjectively or objectively acceptable is just as hard as the corresponding problems of credulous and sceptical acceptance in standard *coherent* Argumentation Systems, cf. [12, p. 202]. Moreover Theorem 3 demonstrates that the effort required to identify the points of disagreement on which the acceptance or rejection of an argument turns is likely to be not well spent.

This does not, however, vitiate the intent underlying the use of VAFs. The situations in which VAFs are intended to be deployed are specific rather than entirely general: a particular legal case, a particular political decision. In such concrete situations, even where the values are many the audiences are few, and thus the relevant status of an argument can be determined by reference to the particular audiences engaged in the debate. Where the ordering of preferences of the audiences involved is not given in advance - as in the dialogue situations envisaged in [3] - a major thrust of the dialogue is to clarify the order of preferences of the participants. Since the motivation of VAFs is to resolve disagreements among particular parties with different value preferences, we can, for the purpose of any given dispute, rely on there being only a small set of audiences that need to be considered.

It could be claimed, with some justification, that the emphases in developing practical exploitation of the VAF formalism should be directed towards such studies. Thus one can consider algorithmic approaches by which a defender of a particular argument can determine in the light of information discovered concerning individual preferences whether there is a possible defence under which the argument in question is acceptable to all of the participating audiences. In principle such methods ought not to require an exhaustive enumeration of *all* possible value orderings since almost all of these will be irrelevant. An important related issue concerns *processes* for uncovering value preferences. Given that the question of whether a particular argument is accepted by a given audience can be decided by efficient methods, it may be possible to determine the exact value preferences of a participant from the answers given to questions concerning whether particular arguments are accepted or not, e.g. if x and y are different arguments with $\eta(x) \neq \eta(y)$ and $\langle x, y\rangle \in \mathcal{A}$, then should some participant answer that *both* x and y are acceptable, it must be the case that the ordering $\eta(y) \succ \eta(x)$ pertains. This prompts the algorithmic question, given a VAF and an *unknown* value ordering, of identifying a suitable set of queries regarding acceptance the answers to which allow the specific audience to be fully determined. It is interesting to note that the preliminary study of these issues reported in [14] indicates that the problem of constructing an audience with respect which a specific set of arguments is a *preferred extension* admits an efficient algoirthmic solution.

As one final issue there is the development of dialogue and reasoning processes specifically intended for VAFs. Although, [3] presents a reasoning semantics for subjective acceptance with respect to a particular audience, akin to the formal basis for TPI-disputes in standard frameworks described in [13], there are several directions in which this may need further refinement. Thus, sound and complete dialogue mechanisms for deciding subjective and objective acceptance in general would be of interest. In view of Theorem 2 it is likely to be the case that “reasonable” sound and complete schemes determining objective acceptance engender exponential length reasoning processes in some cases, cf. [13], however, when the notion of "objective acceptance" is qualified to refer to relevant audiences only (so that the context of Theorem 2 does not apply) it may be the case that such reasoning processes are feasible within appropriate VAF settings.

References

1. L. Amgoud and C. Cayrol. 'A reasoning model based on the production of acceptable arguments', *Annals of Math. and Artificial Intelligence*, **34**, 197–215, (2002)
2. T. J. M. Bench-Capon. 'Representation of Case Law as an Argumentation Framework' in *Legal Knowledge and Information Systems*, eds., T. Bench-Capon, A. Daskalopoulu and R. Winkels, IOS Press: Amsterdam. 103-112 (2002)
3. T. J. M. Bench-Capon. 'Agreeing to differ: modelling persuasive dialogue between parties with different values', *Informal Logic*, **22** 3 (2003).
4. T. J. M. Bench-Capon. 'Persuasion in Practical Argument Using Value Based Argumentation Frameworks', *Journal of Logic and Computation*, **13** 3 429-48 (2003).
5. A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. 'An abstract, argumentation-theoretic approach to default reasoning', *Artificial Intelligence*, **93**(1–2), 63–101, (1997).
6. G. Christie. *The Notion of an Ideal Audience in Legal Argument*, Kluwer Academic, Dordrecht, (2000).
7. Y. Dimopoulos, B. Nebel, and F. Toni. 'Preferred arguments are harder to compute than stable extensions', in *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-99-Vol1)*, ed., T. Dean, pp. 36–43, San Francisco, (1999). Morgan Kaufmann Publishers.
8. Y. Dimopoulos, B. Nebel, and F. Toni. 'Finding admissible and preferred arguments can be very hard', in *KR2000: Principles of Knowledge Representation and Reasoning*, eds., A. G. Cohn, F. Giunchiglia, and B. Selman, pp. 53–61, San Francisco, (2000). Morgan Kaufmann.
9. Y. Dimopoulos, B. Nebel, and F. Toni. 'On the compuational complexity of assumption-based argumentation by default reasoning', *Artificial Intelligence*, **141**, 57–78, (2002).
10. Y. Dimopoulos and A. Torres. 'Graph theoretical structures in logic programs and default theories', *Theoretical Computer Science*, **170**, 209–244, (1996).
11. P. M. Dung. 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *N*-person games', *Artificial Intelligence*, **77**, 321–357, (1995).
12. P.E. Dunne and T.J.M. Bench-Capon. 'Coherence in finite argument systems', *Artificial Intelligence*, **141**, 187–203, (2002).
13. P.E. Dunne and T.J.M. Bench-Capon. 'Two party immediate response disputes: properties and efficiency', *Artificial Intelligence*, **149**, 221–250, (2003).
14. P.E. Dunne and T.J.M. Bench-Capon. 'Identifying Audience Preferences in Legal and Social Domains', Technical Report, ULCS-04-010, Dept. of Computer Science, Univ. of Liverpool, 2004 (submitted)
15. H. Jakobovits and D. Vermeir. 'Dialectic semantics for argumentation frameworks', in *Proceedings of the Seventh International Conference on Artificial Intelligence and Law (ICAIL-99)*, ACM SIGART, pp. 53–62, N.Y., (June 1999). ACM Press.
16. C. H. Papadimitriou and M. Yannakakis. 'The complexity of facets (and some facets of complexity)', in *Proceedings of the Fourteenth ACM Symposium on the Theory of Computing (STOC-82)*, pp. 255–260, San Francisco, CA, (1982).
17. C. Perelman. *Justice, Law and Argument*, Reidel: Dordrecht, 1980.

18. H. Prakken. *Logical Tools for Modelling Legal Argument*, Kluwer Academic Publishers, 1997.
19. J.R. Searle. *Rationality in Action.* MIT Press, Cambridge Mass., 2001
20. G. Vreeswijk and H. Prakken. 'Credulous and sceptical argument games for preferred semantics.', in *Proceedings of JELIA'2000, The 7th European Workshop on Logic for Artificial Intelligence.*, pp. 224–238, Berlin, (2000). Springer LNAI 1919, Springer Verlag.