Persuasion in Practical Argument Using Value-based Argumentation Frameworks

TREVOR J. M. BENCH-CAPON, *Department of Computer Science, The University of Liverpool, Liverpool, UK.*
E-mail: tbc@csc.liv.ac.uk

Abstract

In many cases of disagreement, particularly in situations involving practical reasoning, it is impossible to demonstrate conclusively that either party is wrong. The role of argument in such cases is to persuade rather than to prove, demonstrate or refute. Following Perelman, we argue that persuasion in such cases relies on a recognition that the strength of an argument depends on the social values that it advances, and that whether the attack of one argument on another succeeds depends on the comparative strength of the values advanced by the arguments concerned. To model this we extend the standard notion of Argumentation Frameworks (*AFs*) to Value-based Argumentation Frameworks (*VAFs*). After defining *VAFs* we explore their properties, and show how they can provide a rational basis for the acceptance or rejection of arguments, even where this would appear to be a matter of choice in a standard *AF*. In particular we show that in a *VAF* certain arguments can be shown to be acceptable however the relative strengths of the values involved are assessed. This means that disputants can concur on the acceptance of arguments, even when they differ as to which values are more important, and hence that we can identify points for which persuasion should be possible. We illustrate the above using an example moral debate. We then show how factual considerations can be admitted to our framework and discuss the possibility of persuasion in the face of uncertainty and disagreement as to values.

Keywords: Argumentation, dialogue, persuasion, practical reasoning.

1 Introduction

Sometimes when there is disagreement, it is possible for one party to convince the other by means of a demonstration. In some fields, such as mathematics, this is even the typical case. But in most areas of dispute involving practical reasoning, such as law and ethics, the case is rather different. As Perelman, whose *New Rhetoric* [9] has been highly influential in informal argument, puts it:

> 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. [10, p. 150, italics mine].

It is to resolve this kind of disagreement that the need for argumentation, intended to secure assent through persuasion rather than intellectual coercion, arises. Such disagreement is common in law. When a case is brought to court, it is because the two parties disagree about what should be done in the light of some set of particular circumstances. Often no conclusive demonstration of the rightness of one side is possible: both sides will plead their case, presenting arguments for their view as to what is correct. Their arguments may all be sound. But their arguments will not have equal value for the judge charged with deciding the case: the case will be decided by the judge preferring one argument over the other. And when the judge decides the case, the verdict must be supplemented by an argument, intended to convince the parties to the case, fellow judges and the public at large, that the favoured argument is the one that *should* be favoured. This means that the judge's preference for one argument over the other should be rational, or at least capable of rationalization.

One way of giving rationality to the preference is to relate the arguments to the purposes of the law under consideration, or the values that are promoted by deciding for one side against the other. Perelman [10] says that each party to a legal dispute 'refers in its argumentation to different values' and that the 'judge will allow himself to be guided, in his reasoning, by the spirit of the system, i.e., by the values which the legislative authority seeks to protect and advance' (p. 152). A key element in persuasion is identifying the value conflict at the root of the disagreement so that preference between values can explicitly inform the acceptance or rejection of the competing arguments. Becoming convinced is importantly bound up with identifying how the decision argued for advances the values one holds. Perelman makes much of the fact that an argument is addressed to an *audience*: in many cases this will be a particular audience with a particular set of values, and a particular ranking of them. As an example consider the regular debate as to whether taxes should be raised or lowered. Typically some will say they should be raised to promote social equality while others will says they should be lowered to promote enterprise. Both parties accept that the effects argued by their opponents will be achieved, and both regard greater equality and greater enterprise as good things. Which side they support, however, will depend on whether they value equality or enterprise more. In this case the dispute can only be resolved through *choosing* a preference. As we shall see, however, there are circumstances where particular arguments should be accepted by all audiences, however they choose to rank their values.

Since they were introduced in [6], Argumentation Frameworks (*AF*) have been a fruitful way of looking at systems of conflicting argument. They do not, however, always provide a rational basis for preferring one argument over another: they can identify which points of view are defensible, but are often silent as to which should be preferred. In this paper we extend these Argumentation Frameworks to Value-based Argumentation Frameworks (*VAF*), to attempt to represent the kind of use of values to ground rational disagreement described above. We also show that *VAFs* have some nice properties which can be used to render tractable problems which are intractable in standard *AFs*, and to resolve certain disagreements which cannot be resolved in standard *AFs*. The introduction to *The New Rhetoric* concludes:

> Logic underwent a brilliant development during the last century when, abandoning the old formulas, it set out to analyze the methods of proof used effectively by mathematicians. . . . One result of this development is to limit its domain, since everything ignored by mathematicians is foreign to it. Logicians owe it to themselves to complete the theory of demonstration obtained in this way by a theory of argumentation [9, p. 10]

Our intention in extending *AFs* to *VAFs* is to begin to provide this kind of completion.

We will first recapitulate the standard notion of an *AF*, and consider how persuasion is possible with respect to an *AF*. We then introduce the notion of a *VAF*, and discuss the properties of *VAFs*. We then use an example of a well-known moral debate to illustrate these properties. We then see how practical and factual arguments can be combined in a *VAF*, Finally we provide a summary of our argument.

2 Standard argumentation frameworks

Dung [6] defines an argumentation framework as follows.

DEFINITION 2.1
An *argumentation framework* is a pair

$$AF = \langle AR,\ \textit{attacks} \rangle$$

where *AR* is a set of arguments and *attacks* is a binary relation on AR, i.e.

$$\textit{attacks}\ \subseteq AR \times AR.$$

For two arguments A and B, the meaning of *attacks(A,B)* is that A represents an attack on B. We also say that a set of arguments S attacks an argument B if B is attacked by an argument in S. An *AF* is conveniently represented as a directed graph in which the arguments are vertices and edges represent attacks between arguments. This picture of an *AF* underlies much of our discussion.

The key question to ask about such a framework is whether a given argument $A, A \in AR$, should be accepted. One reasonable view is that an argument should be accepted only if every attack on it is rebutted by an accepted argument This notion produces the following definitions.

DEFINITION 2.2
An argument $A \in AR$ is *acceptable* with respect to set of arguments $S(\textit{acceptable}(A, S))$, if:

$$(\forall x)((x \in AR)\&(\textit{attacks}(x, A)) \rightarrow (\exists y)(y \in S)\&\ \textit{attacks}(y, x).$$

Here we can say that y defends A, and that S defends A, since an element of S defends A.

DEFINITION 2.3
A set S of arguments is *conflict-free* if $\neg(\exists x)(\exists y)((x \in S)\&(y \in S)\&\ \textit{attacks}(x, y))$.

DEFINITION 2.4
A conflict-free set of arguments S is *admissible* if

$$(\forall x)((x \in S) \rightarrow\ \textit{acceptable}(x, S).$$

DEFINITION 2.5
A set of arguments S in an argumentation framework *AF* is a *preferred extension* if it is a maximal (with respect to set inclusion) admissible set of *AR*.

The notion of a preferred extension is interesting because it represents a consistent position within *AF*, which can defend itself against all attacks and which cannot be further extended without introducing a conflict. We can now view a *credulous* reasoner as one who accepts an argument if it is in *at least one* preferred extension, and a *sceptical* reasoner as one who accepts an argument only if it is in *all* preferred extensions.

From [6] we know that every *AF* has a preferred extension (possibly the empty set), and that it is not generally true that an *AF* has a unique preferred extension. In the special case where there is a unique preferred extension we say the dispute is *resoluble*, since there is only one set of arguments capable of rational acceptance.

It is known [8] that establishing whether an argument is credulously accepted is NP-complete, and that deciding whether an *AF* has a unique preferred extension is CO-NP complete. Thus, determining whether a dispute is resoluble is not in general a tractable problem.

In normal circumstances, the plurality of preferred extensions derives from the presence of cycles in the graph. For multiple preferred extensions to exist in a finite argument framework without self-attacks, there must be a simple cycle of *even* length.

THEOREM 2.6
If $AF = \langle AR, \textit{attacks}\rangle$ where AR is finite and attacks contains no self-attacks, has two (or more) preferred extensions, then the directed graph of AF contains a simple directed cycle of even length.

PROOF. Suppose that P and Q are different preferred extensions of *AF*. Let

$$P/Q = \{p_1, p_2, \dots p_r\}; Q/P = \{q_1, q_2, \dots q_s\}.$$

Both sets are nonempty since otherwise $P \subseteq Q$ or $Q \subseteq P$, which would violate the condition that preferred extensions are maximal admissible sets. For each $p_i \in P/Q$ there must be some $q_j \in Q/P$ such that *attacks*$(p_i q_j)$ or *attacks*(q_j, p_i). Without loss of generality assume that *attacks*(p_1, q_1). Since Q is an admissible set, there is some $q \in Q/P$ for which *attacks*(q, p_1). If $q = q_1$ then the pair $\{p_1, q_1$ forms an even length cycle. Otherwise, by continuing to identify successive defences in P/Q (resp Q/P) to the attack on the most recent defence, the point is reached whereby paths

$$\{p \to q_k\} \to \{p_{-1} \to q_{k-1}\} \to \dots \to \{q_2 \to p_1\} \to q_1; \text{ or}$$
$$q \to \{p_k \to q_k\} \to \{p_{k-1} \to q_{k-1}\} \to \dots \to \{q_1 \to p_1\}$$

are found for which

$$p \in \{p_{k-1}, p_{k-2}, \dots p_1\}, \text{ or } q \in \{q_k, q_{k-1}, \dots, q_1\},$$

both yielding an even length directed cycle with t less than or equal to r distinct arguments from each of P/Q and Q/P. ∎

Moreover, it can be shown that the unique preferred extension of an *AF* which contains no cycles can be constructed in a number of steps linear to the number of attacks in *AF*. The method is to select all unattacked arguments and include them in the preferred extension. Next remove all arguments attacked by those included so far. Either no arguments remain, or there are some new unattacked arguments. Include these and repeat until no arguments remain. This method will always succeed in an acyclic graph.

EXTEND$(AF, \textit{attacks})$

1. $s := \{s \in AR : (\forall y)\textit{not defeats}(y, s))\}$
2. $R := \{r \in AR : \exists s \in S \textit{ for which defeats } (s, r)\}$
3. *If* $S = \varnothing$ *then return S and Halt*
4. $AR' := AR/(S \cup R)$
5. *attacks*$' :=$ *attacks*$/((S \times R) \cup (R \times AF) \cup (AF \times R))$
6. *Return* $S \cup$ *EXTEND*$(AR', attacks')$

To see that this method is correct first note that the condition that there are no cycles holds throughout: removing arguments from *AF* cannot create a cycle. Since at least one argument is removed on each pass, the algorithm will eventually halt. It remains to show that the set returned is a preferred extension. The arguments in S must be included in the preferred extension because they are not defeated. Either they were initially not defeated, or their attackers were removed in an earlier pass before they were included in S. Similarly no argument from R can be in the preferred extension, because their inclusion in R means that they are defeated by an argument in S. The new system $\langle AR', \mathit{attacks}' \rangle$ now contains a subset S' of arguments with no attackers in AR'. These are those arguments which were originally attacked by arguments in R, and we know that a defence to these attacks is provided by S. These arguments may therefore be included in the preferred extension.

Taken together, these results mean that if an *AF* contains no even cycles, the dispute is resoluble, and that its resolution can be achieved in time linear to the number of attacks. Unfortunately, this is not as promising for a standard *AF* as might appear, since the complexity status of the problem of checking whether a directed graph in fact contains an even cycle is open: no general polynomial time algorithm has been found, although neither has the problem been shown to be NP-complete, and some special cases, such as directed planar graphs do have polynomial time algorithms. When, however, we are dealing with a Value-based Argumentation Framework, these tractability problems can be, under conditions that typically hold, ignored.

3 Persuasion in a standard argumentation framework

Using a standard argumentation framework we can develop a notion of persuasion. I will illustrate this using the argumentation framework shown in Figure 1. The nodes representing arguments are labelled with the conclusions of the arguments.

This represents a situation where we do not know whether three propositions, P, Q and R, are true or false. One type of argument is the simple assertion of a proposition; clearly this attacks, and is attacked by the assertion of its negation. We know, however, that P, Q and R are related, and we have arguments that conclude $P \rightarrow Q, -P \rightarrow Q$ and $Q \rightarrow R$. The first of these is attacked by $-P$, the second by P and the third by $-Q$. The first and second of these attack $-Q$ and the third attacks $-R$. From Figure 1 we can see that there are two preferred extensions: $\{P, Q, R, P \rightarrow Q, -P \rightarrow Q, Q \rightarrow R\}$ and $\{-P, Q, R, P \rightarrow Q, -P \rightarrow Q, Q \rightarrow R\}$. We can therefore see that Q and R are sceptically acceptable and P and $-P$ are credulously acceptable. This means that we should be able to persuade someone to accept Q. Suppose we assert Q: our interlocutor may challenge this with $-Q$. We attack this with $P \rightarrow Q$. He in turn attacks this with $-P$. I concede not P, and attack $-Q$ with $-P \rightarrow Q$. Now my opponent cannot attack this with P, since this is attacked by the already asserted $-P$. Therefore my opponent should be persuaded of the truth of Q.

What of a credulously acceptable argument, such as P? Here I cannot persuade my opponent because he can counter with $-P$, and I have no independent way of arguing against $-P$. So here I cannot persuade my opponent that P should be accepted, but neither can I be persuaded that it should be abandoned. There is no rational way of choosing between P and $-P$; it is an empirical fact which must be determined by observation.

In this situation, the act of persuasion is akin to a demonstration of the truth of the proposition; it is rather like giving a *proof*. In the sense that the conclusion cannot be rationally

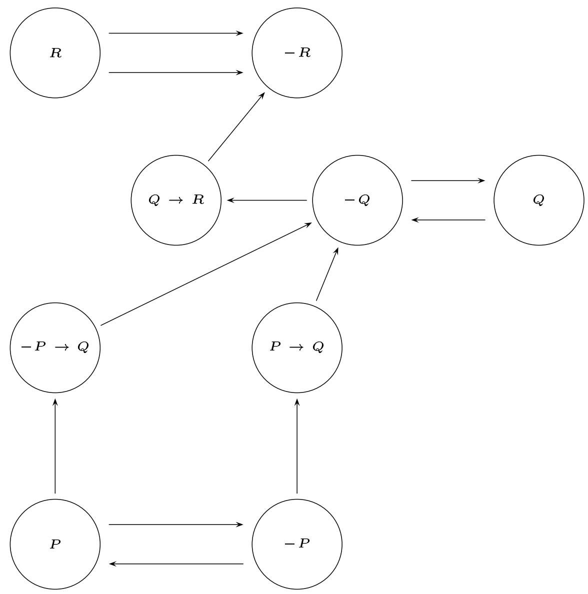

FIGURE 1. Example argumentation framework

rejected this is less persuasion than coercion. While this is appropriate in some domains, it seems rather more problematic in areas of practical reasoning, such as law or ethics. For there, disagreement is less a matter of lack of awareness of some facts or some chain of reasoning than a fundamental disagreement as to what is more important in the given situation, and so which arguments actually succeed in defeating the arguments they attack.

We will look at practical reasoning in the next section.

4 Practical reasoning

While the standard argumentation framework seems well adapted for reasoning about matters of fact, it is less so for practical reasoning. In practical reasoning an argument often has the following form:

1. Action A should be performed in circumstances C, because the performance of A in C would promote some good G.

This kind of argument may be attacked in a number of ways. It may be that circumstances C do not obtain; or it may be that performing A in C would not promote good G. These are similar to the way in which a factual argument can be attacked in virtue of the falsity of a premiss, or because the conclusion does not follow from the premisses. Alternatively it can be attacked because performing some action B, which would exclude A, would also promote G in C. This is like an attack using an argument with a contradictory conclusion. However, a practical argument such as (1) can be attacked in two additional ways: it may be that G is not accepted as a good worthy of promotion, or that performing action B, which would exclude performing A, would promote a good H in C, and good H is considered more desirable than

G. The first of these new attacks concerns the ends to be considered, and the second the relative weight to be given to the ends. For (1) to have deontic force, it must be accepted that G is a good. Here we will always assume that the values advanced by arguments are acceptable, that they do have deontic force for all parties concerned. We will therefore focus on the attacks which depend on the relative weight of the values.

Attacks which make no reference to value will always succeed, provided the attacking argument is accepted. This is what Dung's framework models. However, if an argument attacks an argument whose value is preferred it can be accepted, and yet not defeat the argument it attacks. Thus we can, for arguments which derive their force from the promotion of a value, distinguish between attack and defeat (a successful attack). In order to represent this we must extend the standard argumentation framework so as to include the notion of value. This extension is presented in the next section.

5 Value-based argumentation framework

To record the values associated with arguments we need to add to the standard argumentation framework a set of values, and a function to map arguments on to these values.

DEFINITION 5.1
A *value-based argumentation framework* (*VAF*) is a 5-tuple:

$$VAF = \langle AR, \textit{attacks}, V, val, P \rangle$$

where *AR* is a finite set of arguments, *attacks* is an irreflexive binary relation on *AR*, V is a nonempty set of values, *val* is a function which maps from elements of *AR* to elements of V and P is the set of possible audiences. We say that an argument A relates to value v if accepting A promotes or defends v: The value in question is given by *val(A)*. For every $A \in AF, val(A) \in V$.

The set P of audiences is introduced because, following Perelman, we want to be able to make use of the notion of an audience. Audiences are individuated by their preferences between values. We therefore have potentially as many audiences as there are orderings on V. We can therefore see the elements of P as being names for the possible orderings on V. Any given argumentation will be assessed by an audience in accordance with its preferred values. We therefore next define an audience-specific value-based argumentation framework, *AVAF*.

DEFINITION 5.2
An *audience-specific value-based argumentation framework* (*AVAF*) is a 5-tuple:

$$VAF_a = \langle AR, attacks, V, val, Valpref_a \rangle$$

where *AR, attacks, V and val* are as for a *VAF, a* is an audience, and *Valpref* $_a$ is a preference relation (transitive, irreflexive and asymmetric) *Valpref* $_a \subseteq V \times V$, reflecting the value preferences of audience a. We write v_1 is preferred to v_2 as valpref(v_1,v_2). The *AVAF* relates to the *VAF* in that *AR*, *attacks*, V and *val* are identical, *a* $\in$ *P* and *Valpref* is the set of preferences derivable from the ordering a in the *VAF*.

Our purpose in extending the *AF* was to allow us to distinguish between one argument attacking another, and that attack succeeding, so that the attacked argument is defeated. We therefore define the notion of *defeat for an audience:*

DEFINITION 5.3
An argument $A \in AF$ defeats$_a$ an argument $B \in AF$ for audience a if and only if both *attacks*(A, B) and not $valpref(val(B), val(A))$.

Note that an attack succeeds if both arguments relate to the same value, or if no preference between the values has been defined. If V contains a single value, or no preferences are expressed, the *AVAF* becomes a standard *AF*. If each argument can map to a different value, we have a preference-based argument framework [1]. In practice we expect the number of values to be small relative to the number of arguments. Many disputes can be naturally modelled using only two values. Note that defeat is only applicable to an *AVAF*: defeat is always relative to a particular audience. We write *defeats*$_a(A, B)$ to represent that A defeats B for audience a, that is A defeats B in *VAF*$_a$.

[6] introduces the important notions, described in Section 2, of *acceptability, conflict free set, admissible set*, and *preferred extension* for *AFs.* We next need to define these notions for an *AVAF*.

DEFINITION 5.4
An argument $A \in AR$ is acceptable to audience a (acceptable$_a$) with respect to set of arguments S, (*acceptable*$_a(A, S)$) if:

$$(\forall x)((x \in AR \& \textit{defeats}_a(x, A)) \rightarrow (\exists y)((y \in S) \& \textit{defeats}_a(y, x))).$$

DEFINITION 5.5
A set S of arguments is *conflict-free for audience a* if

$$(\forall x)(\forall y)((x \in S \& y \in S) \rightarrow (\neg \textit{attacks}(x, y) \vee \textit{valpref}(\textit{val}(y), \textit{val}(x)) \in \textit{valpref}_a))).$$

DEFINITION 5.6
A *conflict-free for audience a* set of arguments S is *admissible for an audience a* if

$$(\forall x)(x \in S \rightarrow \textit{acceptable}_a(x, S)).$$

DEFINITION 5.7
A set of arguments S in a value-based argumentation framework *AF* is a *preferred extension for audience a* (*preferred*$_a$) if it is a maximal (with respect to set inclusion) *admissible for audience a* set of *AR*.

Now for a given choice of value preferences *valpref*$_a$ we are able to construct an *AF* equivalent to the *AVAF*, by removing from *attacks* those attacks which fail because faced with a superior value.

Thus for any *AVAF*, *vaf*$_a$ $= \langle$*AR, attacks, V, val, Valpref*$_a\rangle$ there is a corresponding *AF*, *af*$_a$ $= \langle$*AR, defeats*$\rangle$, such that an element of attacks, *attacks(x,y)* is an element of *defeats* if and only if *defeats*$_a(x, y)$. The preferred extension of *af*$_a$ will contain the same arguments as *vaf*$_a$, the preferred extension for audience a of the *VAF*. Note that if *vaf*$_a$ does not contain any cycles in which all arguments pertain to the same value, *af*$_a$ will contain no cycles, since the cycle will be broken at the point at which the attack is from an inferior value to a superior one. Hence both *af*$_a$ and *vaf*$_a$ will have a unique, nonempty, preferred extension for such cases.

6 Acceptance in value-based argument frameworks

We can now look at notions of acceptance in value-based argumentation frameworks. Consider the framework with two values (called *red* and *blue*) in Figure 2.

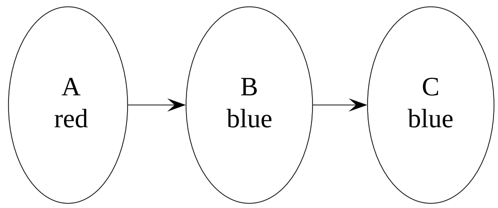

FIGURE 2. VAF with two values

If this were a standard *AF*, A and Cwould be sceptically acceptable. If, however, we consider the values for the two possible audiences, *red* and *blue*, we get the following two preferred extensions. For *red*, which prefers red to blue, we get *preferred*$_{red}$ is $\{A,C\}$. For *blue*, which prefers blue to red, however, *preferred*$_{blue}$ is $\{A,B\}$. There are two points to note here: first that a sceptically acceptable argument in a value free framework may be rejected by a consideration of values, and second that some arguments, like A in Figure 2, may be acceptable irrespective of the choice of values. We will term arguments which are acceptable irrespective of choice of value preferences, that is acceptable to every audience, *objectively acceptable*. Arguments which are acceptable to some audiences, B and C in Figure 2, will be termed *subjectively acceptable*. Note also that sceptical acceptance in the framework considered as an *AF* is not only not sufficient for objective acceptance, but is also not necessary. Suppose we add an attack from C to A in Figure 2: now the preferred extension of the *AF* is empty, since we have a three-cycle: A remains, however, objectively acceptable, since either it is not defeated by C, or else C is defeated by B, which A fails to defeat. Note that objective acceptance of an attacked argument requires that the number of values be smaller than the number of arguments: otherwise it is always possible to prefer the value of the attacker, and that value to that of any of its attackers. A *VAF* is most useful when the number of values is small, since a single choice of preference between values is then able to determine whether a number of attacks succeed or fail.

We may define the notions of objective and subjective acceptance as follows.

DEFINITION 6.1
Objective Acceptance. Given a *VAF*, $\langle AR, attacks, V, val, P\rangle$, an argument $A \in AR$ is objectively acceptable if and only if for all $p \in P, A$ is in every *preferred*$_p$.

DEFINITION 6.2
Subjective Acceptance. Given a *VAF*, $\langle$*AR, attacks,V,val,P* $\rangle$, an argument $A \in AR$ is subjectively acceptable if and only if for some $p \in P, A$ is in some *preferred*$_p$.

An argument which is neither objectively nor subjectively acceptable (such as one attacked by an objectively acceptable argument with the same value) is said to be *indefensible*.

[2] discusses the properties of value-based argumentation frameworks, particularly for cases with two values and no cycles containing arguments relating to a single value. Is the avoidance of single valued cycles a severe limitation? We do not think so. While there is a natural requirement for even cycles in a standard *AF* (Figure 1 shows that a two-cycle is the obvious way to deal with uncertain and incomplete information), and Dung argues strongly in [6] that an interpretation of an *AF* with an odd cycle is plausible, we believe that they

should be avoided in *VAFS*. An odd cycle means that nothing can be believed: it is akin to a paradox, and paradoxes are best avoided. Even cycles represent dilemmas, and require that a choice between alternatives be made. While such dilemmas have their place in cases of uncertainty, we believe that they should be resolved before practical arguments giving rise to them are advanced. The presence of a single-valued cycle in a *VAF* is a sure indication that the reasoning which gives rise to it is flawed.

Some important properties of *VAFs* with no single-valued cycles are discussed below. First, however, we define the useful notion of an argument chain.

DEFINITION 6.3
An *argument chain* in a *VAF*, C is a set of n arguments $\{a_1 \ldots a_N\}$ such that:

(i) $(\forall a)(\forall b)(a \in C \& b \in C) \rightarrow \mathit{val}(a) = \mathit{val}(b))$;

(ii) a_1 has no attacker in C;

(iii) for all $a_i \in C$ if $i > 1$, then a_i is attacked and the sole attacker of a_i is a_{i-1}.

In an argument chain C it is obvious that, since all attacks will succeed because all arguments have the same value, if a_1 is accepted, then *every odd* argument of C is also accepted and *every even* argument of C is defeated. Similarly if a_1 is defeated, every odd argument of C is defeated and every even argument of C is accepted.

THEOREM 6.4
Every *AVAF* with no single-valued cycles has a unique, nonempty preferred extension.

PROOF. Let *avaf* be an *AVAF*, and let *af* be the standard argumentation framework resulting from removing all failing attacks. If *avaf* is cycle free, then *af* is cycle free and that it has a unique nonempty preferred extension follows immediately from Theorem 6.4. Suppose *avaf* has a cycle. We know that this contains at least two values. Let v be the least preferred value in the cycle, and *arg* be the final argument in a chain relating to this value. The attack from *arg* to the next argument in the cycle will fail. Therefore this attack will not appear in *af* and the cycle will be broken at this point. This applies to all cycles in *avaf*. Therefore *af* is cycle free and hence by Theorem 6.4 has a unique, nonempty preferred extension. ■

From this it follows that an efficient algorithm exists to compute the preferred extension for an audience of a A*VAF* with no single valued cycles. The algorithm is the same as that described for cycle free *AFs* in Section 2: first we construct the corresponding *af* for the desired audience, and then we apply the algorithm.

If, however, we do not have a particular audience, but wish to know whether the argument is indefensible, or subjectively or objectively acceptable, we need to take into account all possible audiences. Thus for a *VAF* with no single-valued cycles the complexity of determining the status of an argument is, in the worst case, the factorial of the number of values in V. A straightforward algorithm is: choose an ordering and compute the preferred extension. If the argument is in this preferred extension, the argument is acceptable, else it is either subjectively acceptable or indefensible. We must now choose other orderings and compute the preferred extension until a different result is obtained. If we do obtain a different result the argument is subjectively acceptable, but if not we may have to compute the preferred extension for all factorial(V) orderings, in order to confirm that it is objectively acceptable or indefensible.

Fortunately, however, we typically do not have to consider the whole *VAF* in order to determine the status of a given argument. Suppose we wish to determine the status of a particular argument *arg*, and that *val(arg)* is v. The status of *arg* can be determined by considering the paths which lead to it. Each path will end with a chain of arguments with value v, and then be followed by zero or more chains relating to other values. The first important observation is that, working back from the argument whose status is in question, once a new chain has been entered, no further arguments relating to v need be considered. This is so because either v will be greater than the value of the next chain, call it v_1, or it will be less. If v is greater, the status of the arguments relating to v_1 can have no effect on *arg*, since the attack will fail. If it is less than v_1, then no subsequent argument relating to v can have an effect on the status of arguments relating to v_1. This argument applies equally to the arguments in the chain relating to v_1 with regard to the next chain should it relate to a different value. To have an effect each new chain must relate to a value ranked more highly than its predecessor. Of course, this will be the ordering for some audience, and so it must be considered: but this condition cannot be satisfied if a value repeats. Thus we may terminate a path as soon as we encounter a repeated value.

We may now define a *line of argument*.

DEFINITION 6.5
A line of argument for an argument *arg*, l_{arg}, comprises a set of n argument chains $\{C_1 \ldots C_n\}$, each relating to distinct values, such that *arg* is the last argument of C_1 and the last argument of C_n attacks the first argument of C_{n-1}, and the first argument of C_n has no attacker with a value not already present in l_{arg}.

Note first that the number of chains in a line of argument cannot exceed the number of values in the *VAF*, since otherwise some value would be repeated. Note second that the line of argument will always terminate. The set of arguments in *VAF* is finite, so there can be no paths containing an infinite number of distinct arguments. An argument can repeat only through a cycle. But we are considering only *VAFS* with no cycles relating to a single value. Therefore if a cycle is travelled around so that an argument repeats, the value must have changed, and so the repetition of the argument will be a repetition of the value, and the line of argument halts since it requires distinct values for its chains.

Next consider the case of a *VAF* in which every argument has at most one attacker. Obviously in such a *VAF* there will be a single line of argument relevant to any argument. Now suppose that *arg* relates to v_1 and that v_1 is the greatest value. If this is so the status of *arg* for this audience will be determined solely by its position in C_1. If *arg* is odd numbered it will be acceptable for this audience, and if it is even it will be unacceptable for this audience.

Next consider C_2. If the attack of the last argument of this chain defeats the first argument of C_1, then it will alter the status of the arguments in C_1, and so all the arguments in C_1, including *arg*, will be subjectively acceptable with respect to that line of argument. Suppose that C_2 is an even chain. Now its last argument will be available to attack the first argument of C_1 only if its first argument is defeated. But suppose all the chains in the line of argument are even chains. Since the first argument of C_n is not defeated, its last argument is defeated and so the first argument of C_{n-1} is undefeated. Thus no first argument in any chain will be defeated, and hence the status of the arguments in C_1 with regard to that line of argument will be determined solely by their position in C_1. Therefore if all the chains $C_2 \ldots C_n$ are even chains, then *arg* is either objectively acceptable or indefensible, according to whether it is odd or even numbered in C_1.

But now suppose C_2 is an odd chain, and that its value is the most highly ranked. Now its attack on the first argument of C_1 will succeed, and all arguments in C_1, including *arg*, will have a different status from the case where v_1 is preferred and so will be subjectively acceptable. This will be so, wherever the odd chain appears in the line of argument: if C_n is the first odd chain, then all chains $C_2 \ldots C_{n-1}$ will be even chains. In the case where the ranking of the values of these chains increases with n, in each case the first argument will be defeated, and so the last argument will not be defeated, and so the last argument of C_2 will defeat the first argument of C_1. Note, however, that it is never necessary to consider the line argument beyond its first odd chain, since that will have the most highly ranked value for some audience.

We may summarize this as

THEOREM 6.6
For a *VAF* containing no cycles relating to a single value, in which every argument has at most one attacker, the status of any argument is determined by the parity of the chains in its line of argument.

(a) An argument is objectively acceptable if and only if it is odd numbered in C_1 and the line of argument contains no subsequent odd chain.
(b) An argument is indefensible if and only if it is even numbered in C_1 and the line of argument does not contain a subsequent odd chain.
(c) An argument is subjectively acceptable if and only if the line of argument contains an odd chain, C_n, where $n > 1$.

This theorem will allow us to determine a number of simple cases. Consider for example a cycle of three arguments relating to two values. One argument will relate to, say, v_1 and the other two to v_2. The cycle will thus comprise an odd chain of v_1 and an even chain with value v_2. Thus the arguments with v_2 will be attacked by an odd chain, and hence subjectively acceptable, while the argument relating to v_1 will be an odd numbered argument of a chain attacked by no odds chains, and hence will be objectively acceptable. In any odd cycle with only two values one value must appear in an odd chain and the other must appear in an even chain, irrespective of what other chains may occur. Note that to determine the value of an argument in a *VAF* with only two values we need consider only its own chain and its predecessor, since the next chain will repeat the original value. Again the odd numbered arguments of the chain preceded by the even chain will objectively acceptable. Thus any odd cycle with exactly two values will contain at least one objectively acceptable argument.

If we consider a four-cycle with two values, there are three possibilities. Call the arguments in the cycle A, B, C, D. First the cycle may be unbalanced, with three arguments relating to one value and one to the other. Here we have two odd chains, and all arguments are subjectively acceptable. Second the cycle may alternate values, with A and C relating to one and B and D to the other. Here we have four odd chains and every argument is subjectively acceptable. Finally, the values may be consecutive with A and B relating to one and C and D to the other. Now we have two even chains and so we will have two objectively acceptable and two indefensible arguments. We can summarize all this to precisely characterise the preferred extension of a cycle with two values.

COROLLARY 6.7
The preferred extension of a cycle with only two values comprises:

(i) the odd numbered arguments of all chains preceded by an even chain;
(ii) the odd numbered arguments of chains with the preferred value;
(iii) the even numbered arguments of all other chains.

Note that those included under (i) are objectively acceptable and those included under (ii) and (iii) are subjectively acceptable. The even numbered arguments of a chain preceded by an even chain are indefensible.

Most *VAFs*, however, contain arguments with more than one attacker. What can we say about these? First we may observe that an argument will either be not objectively acceptable or not indefensible on the basis only of the arguments with its own value connected to it with no intervening values. This follows simply by considering the audience for which its value is ranked as the highest value. To determine this form the subgraph representing the *AF* af_{v_1}, which contains only those arguments for which $val(a) = v_1$ and apply the algorithm EXTEND.

Now if the argument is not acceptable with respect to its own value, we must consider the unattacked arguments in af_{v_1}. At least one of these will, if accepted, cause *arg* to be rejected (has an odd numbers of attacks to reach *arg*). Therefore *arg* will be subjectively acceptable only if all of these can be defeated for a given audience. But this audience may also reject an argument required to defend *arg* (has an even number of steps to *arg*). Therefore we may need to consider all unattacked arguments in af_{v_1}. To see this consider Figure 3. Suppose all the arguments have the same value. Now c is indefensible, because it is defeated by g, and cannot be rescued by f which is itself defeated by e. Suppose we add an argument h relating to a different value and attacking e. If this is value is the stronger, e will be defeated, and so f will be able to rescue c to subjective acceptability. But suppose h also attacks a. Now a will be defeated, and b will defeat c, which thus will remain indefensible. Similarly, if *arg* is acceptable with respect to its own value, such as d in Figure 3, then some unattacked arguments in af_{v_1} will cause it to be accepted. For it to be objectively acceptable, all of these must be considered to see if all of them can be defeated for some audience, since this would provide a means of defeating *arg*. Should they be defeated, we must then consider whether the other unattacked arguments in af_{v_1} are also defeated for this audience, since this would rescue *arg*. This can be seen from a consideration of argument d in Figure 3, and applying the same reasoning by which we showed argument c to be indefensible.

$$
\begin{tikzcd}
a \arrow[r] & b \arrow[r] & c \arrow[r] & d \\
e \arrow[r] & f \arrow[r] & g \arrow[u] &
\end{tikzcd}
$$

FIGURE 3. Argumentation framework

The requirement to consider all lines of argument is what leads to the potential need to consider factorially (in the number of values) many lines of argument., since in pathological cases, we may have lines of argument with the values ordered in this many ways.

While therefore the general problem of determining the status of an argument may be intractable, this can typically be ignored in practice since:

- Typically the number of values will be small. If we have only two values, for example, we need apply EXTEND only twice. Two-valued *VAFs* are of considerable importance:

many disputes can be seen as a clash between two principles, or positions (consider for example political debates in a two-party system).

- That we need only pursue the lines of argument until a value is repeated obviously limits the number of arguments that need to be considered when debating the status of a particular argument. This ability to curtail a line of argument is particularly useful where we are modelling a dispute as a two-person dialogue as in [7].
- Dialogical considerations may also enable us to restrict the number of value orders that need to be considered, since the participants may reveal their value orderings in the dialogue. If we know the value orderings of the participants, then we can determine, for the purposes of a dispute between these parties, the status of an argument by applying EXTEND for each audience.
- Importantly, the fact that a repeated value does not change the status of the argument at the head of a line of argument shows how an argument can be used to attack an opponent's argument without undermining one's own position. The point here is that an argument which a person may wish to defend with a particular value may be attacked by an argument with some other value, and defended by attacking that argument with another argument with that value. Now to accept the original argument, the defender must also accept the other arguments. But if this is not wanted, the defender can attack the third argument with another argument relating to the value of the original argument, relying on the value preference to defend the original claim.

Both points will be illustrated by an example in the next section.

7 An example moral debate

The scenario we will consider is taken from an example discussed by Coleman in [5] and further discussed by Christie in [4]. Hal, a diabetic, loses his insulin in an accident through no fault of his own. Before collapsing into a coma he rushes to the house of Carla, another diabetic. She is not at home, but Hal enters her house and uses some of her insulin. Was Hal justified, and does Carla have a right to compensation?

As presented by Coleman in [5], the first argument is (A) that Hal is justified, since a person has a privilege to use the property of others to save their life — the case of necessity. But should Hal compensate Carla? His justification can be attacked by an argument (B) that it is wrong to infringe the property rights of another. If, however, Hal compensates Carla, we have a property-based argument (C) that Carla's rights have not been infringed. This position is illustrated in Figure 4.

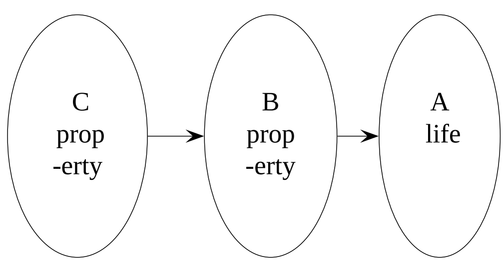

FIGURE 4. Coleman's version of Hal and Carla

The first argument (A) is based on the value that life is important (*life*), the second (B) and third (C) on the value that property owners should be able to enjoy their property (*property*). As it stands (A) and (C) are objectively acceptable: Hal can take the insulin, but must compensate Carla. This appears to be Coleman's view. Christie, however, in [4] does not want to insist on compensation. He therefore introduces a fourth argument (D), which says that if Hal were too poor to compensate Carla, he should none the less be allowed to take the insulin, as no one should die because they are poor. Moreover, he says that since Hal would not pay compensation if too poor, neither should he be obliged to do so, even if he can. We thus have a life-based argument that defeats (C), assuming that life is valued more than property. This situation is shown in Figure 5.

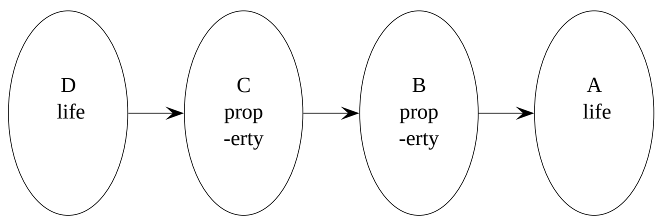

FIGURE 5. Christie's version of Hal and Carla

If we accept (D), then (C) becomes subjectively acceptable, only allowed if we value property more than life. Note, however, that if we value life more than property, (B) is now accepted, its attacker (C) being defeated by (D). (A), however, remains objectively acceptable since its value is strong enough to resist the attack from (B) in this case and (B) is defeated in the other. Note that (D) can only be introduced without threatening (A) because the line of reasoning relevant to (A) terminates when the value pertaining to (A) is re-introduced. In a value-free *AF*, introducing (D) would render (A) indefensible.

Suppose we want to resist Christie's conclusion, that $\{A,B,D\}$ are the acceptable arguments, and do want to insist on compensation. A natural way would be to attack (D) by an argument (E) to the effect that poverty is no defence for theft, that we prosecute the starving when they steal food. (E) is based on property. But this would not achieve our ends, since it would repeat the property value. (Note also that (E) is attacked by (A)). If life is valued over property, (D) is not defeated, and while it is defeated if property is valued over life, it is unnecessary for the defence of (C) which resists (D) unaided. Resistance to Christie can only come from another life-based argument. For example, suppose we attack (A) on the grounds that Hal is endangering Carla's life (F). Now (F) will defeat (A), which Christie wants to defend. He can now attack (F) with (C): if Carla is properly compensated her life is not endangered. This scenario is shown in Figure 6. But for this attack to succeed, property must be valued above life, and now (C) is not defeated by (D). Interestingly, in the scenario of Figure 6, the life-based (A) is reduced to subjective acceptance, and requires that its own value be rated as the lesser of the two.

It is possible to extend the argument framework further so that we can have both (A) and value life more than property, but this requires that we make some assumptions about the facts of the situation. In the next section we will consider how we can combine facts with value-based reasoning.

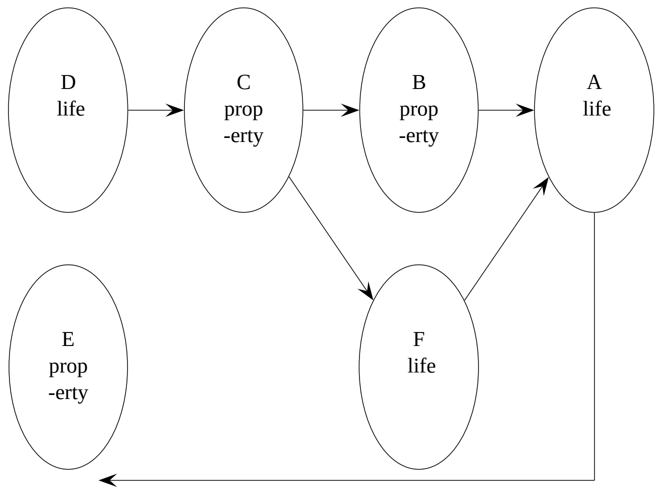

FIGURE 6. Final Hal and Carla scenario

8 Facts in moral debate

In the discussion thus far we have assumed that all arguments relate to some value. But sometimes we need to consider matters of fact as well as opinion grounded in values. In the Hal and Carla example it is usual to include as part of the description that Hal checks that Carla has abundant insulin before using it, in order to exclude from the discussion the line of attack involving danger to Carla. That Carla has abundant insulin (G) clearly attacks (F). This scenario is shown in Figure 7.

Now in the dispute of the last section, the challenger was able to resist persuasion that (A) by preferring life to property. Can C continue to resist by preferring life to fact as well? I answer no: if Carla has abundant insulin, then (F) must fall since the circumstances are such that the desired good is not promoted. My solution is to treat *fact* as if it were a value, but *fact is always the value with the highest preference for all parties*. Whether we prefer life to property is a matter of choice, but to deny facts is to depart from rational argument by resorting to wishful thinking. This is accommodated by including in the initial state of the dispute preferences of the form $fact > val_i$, for every value in V in the *VAF*, and for every audience. Thus in Figure 7, someone can be persuaded of (A) since it is acceptable under any value order for which $fact > life$. We will continue to refer to arguments in preferred extensions for all reasonable value orders (those which rate fact as the highest value) as objectively acceptable.

Introducing facts can bring with it uncertainty. For example, we may not know whether Carla has sufficient insulin. Thus argument (G) in Figure 7 may be attacked by another factual argument (H) to the effect that Carla does not have ample insulin. Note (H) is itself attacked by (G). The situation is shown in Figure 8. This introduces a cycle which is single valued in that both arguments relate to fact.

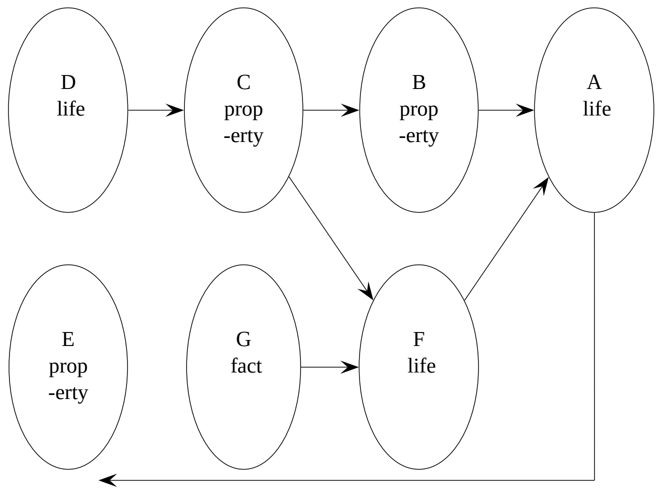

FIGURE 7. Hal and Carla with a Factual Argument

This means that we may get multiple preferred extensions, even if we have an ordering on values. In Figure 8, for *fact* $>$ *life* $>$ *property* we can have either $\{H,F,D,B\}$, or $\{G,A,D,B\}$, and for *fact* $>$ *property* $>$ *life*, either $\{H,C,A,E\}$ or $\{G,A,C,E\}$.

Now we can see that there are four possibilities for the status of an argument. Arguments may be objectively acceptable *sceptically*, if they appear in every preferred extension. They may be objectively acceptable *credulously*, if they appear in every preferred extension corresponding to some choice of facts: thus in the above example (A) is objectively acceptable on the assumption that (G). They may be subjectively acceptable *sceptically* if they appear in every preferred extension relating to some value order; in the above example (D) and (B) are subjectively acceptable however the conflict between (G) and (H) is resolved if life is preferred to property, and all of (A), (C) and (E) are acceptable whenever property is preferred to life. Finally they may be subjectively acceptable *credulously* if they appear in some preferred extension. All the arguments in Figure 8 fulfil this condition.

For persuasion against this background of uncertainty, only arguments whose objective acceptance is sceptically acceptable can be made persuasive for a determined challenger. Otherwise some choice of facts and value preferences will allow him to resist the defence. While the challenger may resist persuasion by a choice of which of two uncertain alternatives to believe, the defender cannot make such a choice and hope to be persuasive. It is for this reason that the choice must be made *in the problem description* when setting up the Hal and Carla scenario, so excluding (H) from consideration, by those who wish to make a persuasive case for accepting (A). Alternatively it is necessary to resolve the factual disputes before attempting to persuade someone to accept the value-based arguments.

Although the treatment of facts suggested here allows single-valued two-cycles to appear, and thus means that we may not have a unique preferred extension even given a value ordering, the tractability implications of this do not present a problem for persuasion. This is because we do not have to entertain all the different possibilities: instead the person to be

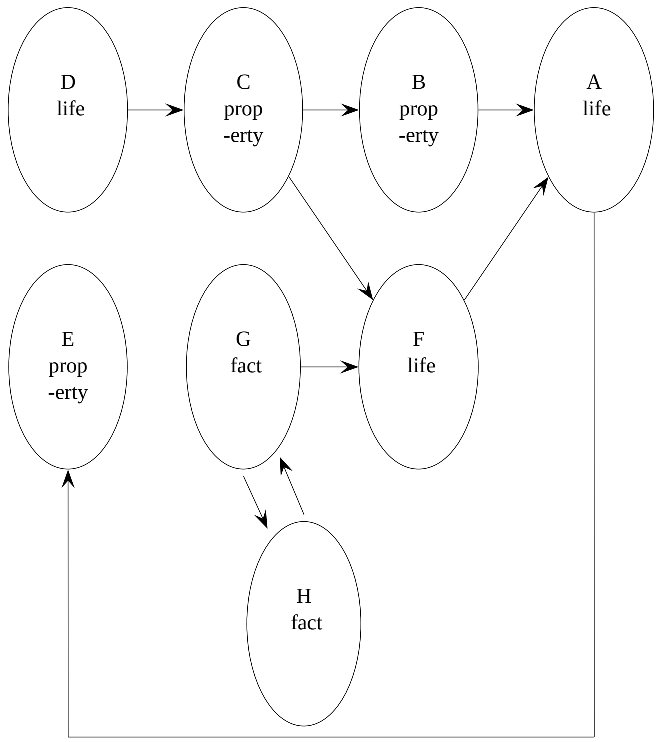

FIGURE 8. Hal and Carla with uncertainty

persuaded is allowed to resolve any dilemma as seems most favourable to him.

9 Summary

In this paper I have been concerned with persuasion. How is it possible that two people may disagree, and yet one convince the other by argument rather than by pointing to some new information or logical connection? To explore this phenomenon I have made use of the idea, originally put forward by Dung, of argumentation frameworks. Note that this abstracts away from the details of individual arguments: the assumption throughout is that the disputants agree as to what arguments should be considered, and as to which arguments attack other arguments. Persuasion is thus a matter of showing the critic that the argument under dispute must be accepted in any coherent position relating to this argument framework. A *coherent position* is given precision through the notion of a preferred extension, a maximal set of arguments able to defend itself against all attacks on any of its members.

Disagreement about some arguments is possible because there is not in general a unique preferred extension, and so any of several coherent positions can be taken. The task of the persuader here is to show that the argument that he wishes to advance is in every preferred extension, that it is what is often termed sceptically acceptable. This is particularly apt against a background of uncertain or incomplete information, since persuasion requires that however the debatable facts are resolved the argument in question will be in the preferred extension.

In practical reasoning, however, there is an additional way of disagreeing. The disputants may agree on which arguments attack which other arguments, but differ as to which of these attacks succeed. They can differ because the success of the attack depends on the relative strengths of the arguments *for an audience*, which in turn relates to the values to which the arguments pertain. To handle this notion of value, I have introduced the notion of *Value-based Argumentation Frameworks (VAF)*. For persuasion to be possible here, the argument must be in the preferred extension with respect to every *AF* derivable from the *VAF* by choosing an ordering on the values involved, a state of affairs which I termed *objectively acceptable*. For *VAFs* which contain no cycles in which all the arguments relate to the same value, the preferred extension given a value ordering is unique. This greatly simplifies the situation, since there is no difference between sceptical and credulous acceptance. The problem of determining whether an argument is objectively acceptable or not in such a framework can therefore be determined in time of order $n * m$where n is the number of attacks and m the number of value orderings. Remember that m is intended to be relatively small, since we envisage a limited number of values in any *VAF* — in many cases two will suffice.

In some cases we will wish to deal with practical reasoning in cases where there is also uncertainty. This I handled by making fact a special value, special because it is always given the highest preference. Note, however, that uncertain facts will form cycles of two arguments relating to the same value. Now each value ordering will no longer have a unique preferred extension. Therefore if we are dealing with both value and uncertainty for an argument to be persuasive it must be in every preferred extension for every value ordering: its objective acceptability must be sceptically acceptable.

All of this has been illustrated by an example relating to a well-known moral problem.

The extension to Value-based Argumentation Frameworks allows the representation of rational discussion pertaining to matters of value as well as fact and logic, and the accommodation of the phenomenon that different audiences will find different reasons persuasive. This is essential if we are to effectively model dispute about practical questions, ethics and law.

Acknowledgements

I am grateful to my colleague Paul Dunne, for many fruitful discussions on argumentation, and for a suggestion from another colleague, Marc Pauly. Several ideas in this paper were presented in a session on Argument, Dialogue and Decision at NMR02, [2], and have benefited from discussion with the participants of that conference, especially John Horty and David Makinson.

References

[1] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, G. Cooper and S. and Moral, eds, 1998.

[2] T.J.M. Bench-Capon. Value-based argumentation frameworks. In *Proceedings of Non Monotonic Reasoning*, pp. 444–453, 2002.

[3] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, **93**, 63–101, 1997.

[4] G.C. Christie. *The Notion of an Ideal Audience in Legal Argument*. Kluwer Academic Publishers, 2000.

[5] J. Coleman. *Risks and Wrongs*. Cambridge University Press, 1992.

[6] P.H. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, **77**, 321–57, 1995.

[7] P.E. Dunne and T.J.M. Bench-Capon. Two Party Immediate Response Disputes. Research Report ULCS-01-005, Department of Computer Science, The University of Liverpool, 2001.

[8] P.E. Dunne and T.J.M. Bench-Capon. Coherence in Finite Argument Systems. *Artificial Intelligence*, **141**, 1–2, 187–203, 2002.

[9] C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame, 1969.

[10] Ch. Perelman. *Justice, Law and Argument*. Reidel, Dordrecht, 1980.

Received 15 June 2002