Persuasion in Practical Argument Using Value Based Argumentation Frameworks

T.J.M. Bench-Capon
Department of Computer Science
The University of Liverpool
Liverpool,
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. This forms the basis for an account of persuasive argument in dialogues pertaining to 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], p150, 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 that the judge's preference for one argument over the other should be rational, or at least capable of rationalisation.

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" (p152). 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. Perelman, however, also wishes to allow for a more objective status for arguments. This is achieved through the notion of the universal audience. Those who address the universal audience "think that all who understand their reasons will have to accept their conclusions. *The agreement of a universal audience is thus a matter, not of fact, but* *of right*". [9], p31, *italics theirs*). Part of what we wish to do in this paper is to show that there can be such universally acceptable arguments, even if we allow the strength of an argument to be determined by the value it promotes.

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 problems which are intractable in standard *AFs* tractable, 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], p10).

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 a running example of a well known moral debate to illustrate these properties and to consider persuasion dialogues with respect to a *VAF*. We then see how practical and factual arguments can be combined in a *VAF*, and how this affects persuasion dialogues. Finally we provide a summary of our argument.

2. STANDARD ARGUMENTATION FRAMEWORKS

Dung [6] defines an argumentation framework as follows.

Definition 1: An *argumentation framework* is a pair

$AF = \langle AR, attacks \rangle$

Where *AR* is a set of arguments and *attacks* is a binary relation on AR, i.e. $attacks \subseteq AR \times AR$.

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*, $A \subseteq AR$, should be accepted. One reasonable view is that an argument should be accepted only if for every argument that attacks it, there is an argument which attacks that other argument. This notion produces the following definitions:

Definition 2: An argument $A \in AR$ is *acceptable* with respect to set of arguments *S*, *(acceptable(A,S))*, if:

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

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

Definition 3: A set *S* of arguments is *conflict-free* if

$\neg(\exists x)\ (\exists y)((x \in S)\ \&\ (y \in S)\ \&\ attacks(x,y))$.

Definition 4: A conflict-free set of arguments *S* is *admissible* if

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

Definition 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 is defensible 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 from [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 possible.

The plurality of preferred extensions derives from the presence of cycles in the graph. For multiple preferred extensions to exist, there must be a cycle of *even* length.

Theorem 6: *If AF = <AR,attacks> has two (or more) preferred extensions, then the directed graph of AF contains a directed cycle of even length.*

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

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

Both sets are non-empty 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 $attacks(p_i,p_j)$ 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

$\{p \rightarrow q_k\} \rightarrow \{p_{k-1} \rightarrow q_{k-1}\} \rightarrow \ldots \rightarrow \{q_2 \rightarrow p_1\} \rightarrow q_1$; or
$q \rightarrow \{p_k \rightarrow q_k\} \rightarrow \{p_{k-1} \rightarrow q_{k-1}\} \rightarrow \ldots \rightarrow \{q_1 \rightarrow p_1\}$
are found for which
$p \in \{p_{k-1},p_{k-2}, \ldots, p_1\}$ or $q \in \{q_k,q_{k-1},\ldots,q_1\}$

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*. $\square$

Moreover, it can be shown that the unique preferred extension of an *AF* which contains no even length 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. For a more formal presentation of the algorithm, see [2].

Taken together these results mean that **f** an *AF* contains no even cycles, the dispute is resoluble, and that its resolution can be achieved in time linear to the number of arguments. 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 polynomial time algorithm has been found, although neither has the problem been shown to be NP-complete. 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.

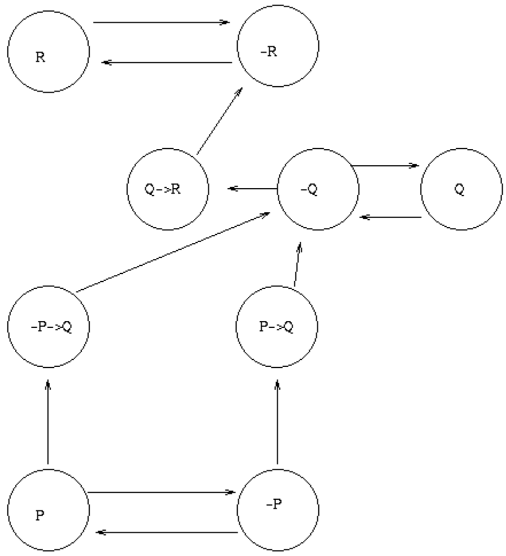

Figure 1: Example Argumentation Framework

This represents a situation of incomplete information concerning three propositions, P, Q and R. 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 → Q, -P → Q and Q → 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 → Q, -P → Q, Q → R} and {-P,Q,R, P → Q, -P → Q, Q → 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 → Q. He in turn attacks this with –P. I concede not P, and attack –Q with -P → 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 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 that 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 premise, or because the conclusion does not follow from the premises. 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 2: A *value-based argumentation framework* (*VAF*) is a 5-tuple:

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

Where *AR*, and *attacks* are as for a standard argumentation framework, *V* is a non-empty 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 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, $a \in P$, and $Valpref_a$ is a preference relation (transitive, irreflexive and asymmetric) $Valpref_a \subseteq V \times V$, reflecting the value preferences of audience *a*. The *AVAF* relates to the *VAF* in that *AR*, *attacks*, *V* and *val* are identical, 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 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 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:

("x)((x Î AR & defeats$_a$*(x,A)) ® ($y)((y Î S) & defeats*$_a$*(y,x)))*.

Definition 5: A set *S* of arguments is *conflict-free for audience a* if

("x) ("y)((x Î S & y Î S) ® (Øattacks(x,y) Ú valpref(val(y),val(x)) Î valpref$_a$*)))*.

Definition 6: A *conflict-free for audience a* set of arguments *S* is *admissible for an audience a* if

("x)(x Î S ® acceptable$_a$*(x,S))*.

Definition 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 = <*AR, attacks, V, val,* $Valpref_a$> there is a corresponding *AF*, af_a = <*AR, defeats*>, 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, non-empty, 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.

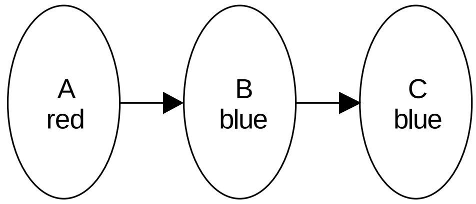

Figure 2: VAF with two values

If this were a standard *AF*, *A* and *C* would 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 of 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 8: *Objective Acceptance*. Given a *VAF*, <*AR, attacks,V,val,P*>, an argument *A Î AR* is objectively acceptable if and only if for all *p Î P*, *A* is in every $preferred_p$.

Definition 9: *Subjective Acceptance*. Given a *VAF*, <*AR, attacks,V,val,P*>, an argument *A Î AR* is subjectively acceptable if and only if for some *p Î P*, *A* is in some $preferred_p$.

An argument which is neither objectively nor subjectively acceptable 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, (*monochromatic)* cycles.

Is the avoidance of monochromatic 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 present of a monochromatic cycle in a *VAF* is a sure indication that the reasoning which gives rise to it is flawed.

Three important properties of *VAFs* with no monochromatic cycles are given below. Here I give an informal justification: readers who wish to see formal proofs should consult [2] and [8].

- *VAFs* with no monochromatic cycles have a unique, non-empty preferred extension, given an ordering on values. This simply follows from the fact that the *AF* corresponding to an *AVAF* with no monochromatic cycles is cycle free.
- The status of an argument can be determined by considering only the chain of which the argument is part (where *chain* is an unbroken sequence of arguments pertaining to the same value) and the chains which directly attack elements of that chain. Either the value of the chain is stronger than its attacker, in which case the first and every odd numbered argument in the chain will be in the preferred extension, or the value of the chain is weaker than that of its attacker. In this case if the last argument of the attacking chain is undefeated, it will defeat the first argument of the next chain, and so every even numbered argument of that chain will be in. Otherwise, the first and every odd numbered argument of the chain will be undefeated.
- An efficient algorithm exists to compute the preferred extension for an audience of a *VAF* with no monochromatic cycles. The algorithm is the same as that described for cycle free *AFs* in section 2: first we construct the corresponding AF_a for the desired audience, and then we apply the algorithm.

For the purposes of this paper the most important observation follows from the second of these points. Suppose we wish to consider the status of an argument. We shall need to consider every path leading into that argument. We do not, however, have to follow the path all the way back to an unattacked argument: once the value of an attacking argument changes from the original value, no further arguments on that path pertaining to that value need be considered. That this is so can be seen as follows. Suppose an argument with value *red* is attacked by an argument with value *blue*. If red is greater than blue then that attack will fail, and the path is terminated. Thus for the path to have significance, blue must be greater than red. But if this is so, no subsequent attack of a red argument on a blue argument can succeed, and so we need consider no subsequent attacks of red on blue arguments. This is so, even if there are more than two values. Suppose the blue argument is attacked by a *green* argument. For this to be considered we must suppose green to be greater than blue, and hence greater than red also. Therefore no subsequent attack of blue (or red) on a green argument can succeed.

This observation has significance in two ways:

- First it obviously limits the number of arguments that need to be considered when debating the status of a particular argument.
- Second, and perhaps more importantly, it shows how an argument can be used to attack an opponents argument without undermining one's own position. The point here is that an argument which a person wished to defend with value red may be attacked by an argument with value blue, and defended by attacking this blue argument with another blue argument. Now to accept the red argument, the defender must also accept the blue argument. But if this is not wanted, the defender can attack the second blue argument with another red argument, relying on the preference *red > blue* to defend the original claim.

Both points will be illustrated by an example in the next section. I have previously discussed this example in [2].

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 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 that it is wrong to infringe the property rights of another. If, however, Hal compensates Carla, we have a property based argument that Carla's rights have not been infringed. This position is illustrated in Figure 3.

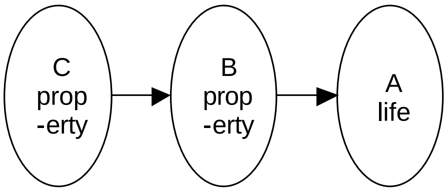

Figure 3: Coleman's version of Hal and Carla

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 4.

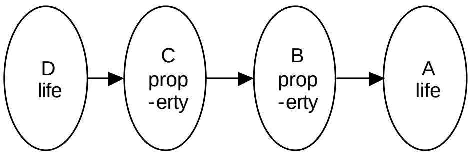

Figure 4: Christie’s version of Hal and Carla

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). 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 5. But for this attack to succeed, property must be valued above life, and now (C) is not defeated by (D). Interestingly, in this scenario, the life based (A) is reduced to subjective acceptance, and requires that its own value be rated as the lesser of the two.

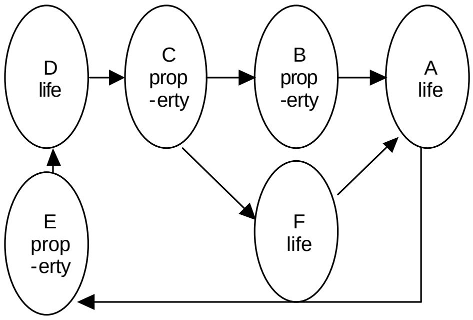

Figure 5: Final Hal and Carla scenario

It is, of course, possible, to extend the argument framework further so that we can have both (A) and value life more than property. I shall return to the example later in the paper.

8. Persuasive Dialogues

We are now in a position to look at the notion of *persuasive dialogue*. It might perhaps be felt that if two disputants differ as to their ranking of values, persuasion would be difficult, if not impossible, and we have all experienced instances where argument has broken down through mutual lack of sympathy with the other’s worldview. None the less the existence of objectively acceptable arguments in a *VAF* indicates that persuasion should on occasion be possible. Since the value order does not affect the acceptability of such arguments, persuasion should be possible, even against a background of different value rankings. What is true, however, is that a persuasive dialogue must be directed towards the value judgements of the *audience* not the *speaker*. It may well be, therefore, that the speaker may have to offer a line of reasoning which he does not himself find persuasive in order to convince his audience. This need not, however, compromise sincerity, since he will independently believe his claim by his own lights.

Another possibility is that the value order of the audience may not be known to the speaker in advance. Therefore we must allow the possibility of value orderings emerging from the dialogue.

A good framework for modelling dispute as to the acceptability of an argument is to use the notion of a *dialogue game*. For example, [7] gives a game for establishing credulous acceptance. This game, and the others we will introduce here are examples of Two-Party Immediate (TPI) Response Disputes, in which we restrict ourselves to two parties, and in which responses can only be directed towards the last move of one's opponent. [7] offers a formal presentation of their game: in this paper I will provide only informal sketches of this and other games, so that we can focus on the intentions of the games, rather than the details.

I shall begin by recalling the game in [7], since it presents features that I wish to incorporate in the persuasion games described below. Let us call this game *CA*. CA allows only three moves: COUNTER, BACKUP and RETRACT. The game has two players, Defender (D) and Challenger (C). D begins play by stating an argument which he wishes to defend. C wishes to render this argument indefensible.

COUNTER may be played by either player. Given an argument, the player offers an argument which attacks it. BACKUP is played by C when no attack is available. It involves moving back through the sequence of arguments played and offering an alternative attack on one of the arguments put forward by D. RETRACT is made only by D; it involves returning to the original claim, and means that the subset of arguments played by D so far cannot be recreated. CA is won by D if a preferred extension including the argument in dispute is created, and by C if this proves impossible.

An example dispute using this game given in [7] is based on the *AF* shown in Figure 6. The state of the dispute is given by the tuple $<T_k,v_k,D_k,C_k,P_k,Q_k>$, where T_k is the dispute tree after k moves, v_k is the current argument vertex of T_k, D_k are the arguments available to D at k, C_k are the arguments available to C at k, P_k are the arguments proposed by D as a (subset) of some admissible set, and Q_k are the set of subsets that C has shown not to be subsets of an admissible set.

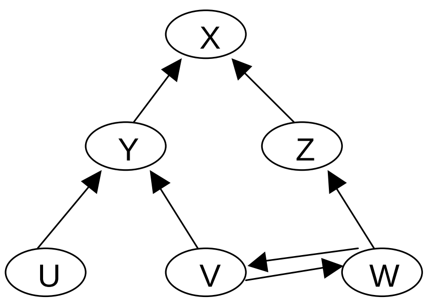

Figure 6: AF for Dispute Example

A possible game relating to the *AF* in Figure 6 would run as follows. D claims X, which is attacked by C with Y. D attacks Y with V. C now chooses to back up and attack X with Z. D cannot now play W, because this is attacked by the already played V. D must therefore retract. C again attacks X with Y, but this time D defends by attacking Y with the unassailable U. C has no choice but to back up and try the attack with Z. This time W is available to D to attack Z, and C cannot attack with V, since it is already attacked by W. Therefore C has successfully defended X. A summary is given in Table 1. This is, of course not "best play", but it does illustrate the various features of the game.

Table 1: CA played on AF shown in Figure 6

k	$move_k$	v_k	D_k	C_k	P_k	Q_k
0	-	X	{U, V, W}	{Y,Z, U,V, W}	{X}	{}
1	C(Y)	Y	{U, V, W}	{Y,Z, U,V, W}	{X}	{}
2	C(V)	V	{U}	{Z,U}	{X.V}	{}
3	B(0,Z)	Z	{U}	{U}	{X.V}	{}
4	R	X	{U, V, W}	{Y,Z, U,V, W}	{X}	{X,V}
5	C(Y)	Y	{U, V, W}	{Z,U, V,W}	{X}	{X,V}
6	C(U)	U	{V, W}	{Z,V, W}	{X,U}	{X,V}
7	B(4,Z)	Z	{V, W}	{V,W}	{X,U}	{X,V}
8	C(W)	W	{}	[]	{X,U, W}	{X,V}

Features to note in this game are:

1) we need a move to enable a player to attack an argument presented in the last move by the opponent;
2) only certain arguments are available to attack the opponent's argument; essentially these must attack the argument in the underlying *AF*, and must not themselves be attacked by an argument already presented;
3) Both challenger and defender need to be able to retrace their steps if they have plunged into a bad line of argument. The moves for challenger and defender are not, however, symmetrical, and so two different moves, one for each role, are required.
4) CA is *not* a persuasion game: if D is successful he retains the right to accept his claim, but C need not accept it, since they may be a preferred extension not containing the claim.

To play a game using values we must begin with a *VAF*, instead of an *AF*. Now, provided that there are no monochromatic cycles – and we have argued that there is no place for monochromatic cycles in a *VAF* – the preferred extension is unique for any given value ordering. In order that C may be persuaded, C must be allowed to determine the value ordering as he chooses: it is the value preferences of the *audience* that determines whether an argument is accepted. But because C has been allowed to determine the value order, if he fails to mount a successful challenge to the claim, he must accept the claim, for there is no alternative preferred extension for this value order to which to appeal. Since then in this case sceptical and credulous acceptance are the same, we may take CA as a starting point.

CA will, however, need some adaptation. First we must place an extra constraint on which arguments are available. Recall that once there has been a value change in line of argument, the value can never be usefully repeated. Therefore if there is a value change at move k, all arguments with the value of the argument played at move k-1 become unavailable, since no argument with this can affect the status of the claim. This has the desirable effect of shortening lines of argument.

Next we need to allow value preferences to be declared. A player will wish to declare a value preference when he would have otherwise lose the dispute. The move effectively severs a link in the chain of reasoning by declaring that one of the attacks fails. We call this move VALUE, and it may only be played by the challenger. Only the challenger may play this move because it is the task of the defender to persuade the challenger. Therefore it is only the challenger who can be allowed to determine what value order is to be used.

VALUE may be played when

- two arguments, a and b in P_k relate to different values, val_a and val_b ;
- attacks(a,b) $\in$ attacks;
- C has not previously played a move expressing or implying that $val_a > val_b$.

The move has a number of effects:

- the challenger is now committed to the preference $val_b > val_a$ and any preferences implied by it. For example if C had previously expressed a preference for val_c over val_b, he is now also committed to $val_c > val_a$.
- Neither player can any longer use any attack of an argument with val_a on an argument with val_b. Such attacks can no longer persuade.
- Moreover neither player can now use any attack which will fail in the face of an implied value preference.
- The dispute returns to argument b.

To provide an example, let us consider the dispute concerning Hal and Carla, as shown in Figure 5. Let us first consider it as a value free dispute using the original game CA.

D puts forward (A) to start the dispute. C could challenge this with either (B) or (F). In either case D counters this with (C) and D counters in turn with (D). Now D has won, since (E) is not available to C, because it is attacked by (A).

Now consider the dispute using values. Again C may counter (A) by using either (B) or (F), and D counters either of these with (C). But now (D) is not available to C, since it would repeat the value *life*. Therefore C will lose the dispute unless he chooses to play VALUE(life, property). Note that this will help only if (A) was countered with (F): otherwise the effect is to break the chain of reasoning by removing the attack of (B) on (A). Had C played (B) initially the correct response would now be BACKUP(0,F). At this point D has no way to persuade C, since (C) is no longer available to attack (F), because of the declared value preference.

Note that although D has failed to persuade C, D is not forced to abandon acceptance of the argument in dispute. What D accepts depends on D's ordering of values, not C's.

In this game, persuasion is possible only if the claim is objectively acceptable: otherwise C may choose whatever value preferences are required to defeat the claim. Suppose, however, we extend the game so that we do not have a single argument at issue, but rather a set of arguments that both participants are prepared to defend. In this scenario it is possible that the need to defend some arguments may require a participant to commit to value preferences that take away the ability to successfully challenge some of his opponent's claims. For example, in Figure 5, C may need to choose to commit to *property > life* in order to defend (C). Once this is done, he can no longer express the different value preference to attack (A). This seems a plausible situation: disputes are rarely about isolated arguments, and the tactic of establishing what values the audience prefers by first considering an uncontroversial issue, and then showing that this requires acceptance of a more debatable position is quite common. We do not elaborate further on this extended game here, although its definition will be a topic for future exploration.

9. 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 6.

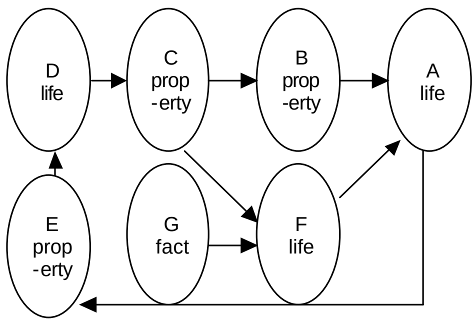

Figure 6: Hal and Carla with a Factual Argument

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 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. If we now play the dispute through again, after C has defended (F) against (C) by use of VALUE, D can attack (F) with (G), leaving (C) with no further challenge. Thus in Figure 6, C 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 6 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 7. This introduces a cycle which is monochromatic in that both arguments relate to fact.

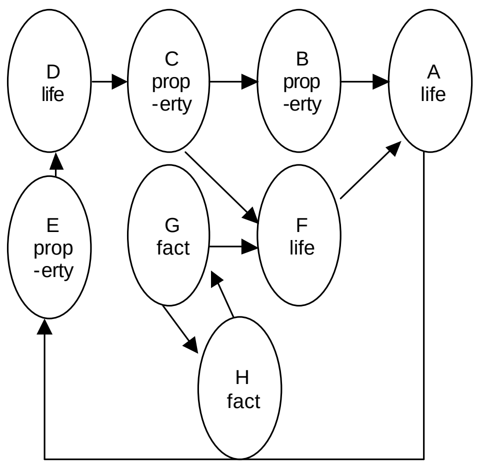

Figure 7: Hal and Carla with Uncertainty

This means that we may get multiple preferred extensions, even if we have an ordering on values. In Figure 7, 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 if preferred to life. Finally they may be subjectively acceptable credulously if they appear in some preferred extension. All the arguments in Figure 7 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. This requires a further modification to CA. Consider CA played on the *VAF* shown in Figure 7. When the defender of A plays (G) to attack (F), (H) is made unavailable because attacked. But (H) is the argument the challenger needs to obtain his preferred extension *{H,F,D,B}*. Therefore we must allow C to play an argument, even if it is attacked, provided that *it attacks all its attackers itself*. Note, however, that this relaxation of COUNTER applies only to C: 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).

Although the treatment of facts suggested here allows monochromatic two-cycles to appear, and thus means that we may not have a unique preferred extension even given the value ordering, the tractability implications of this do not present a problem for the persuasion game. This is because we do not have to entertain all the different possibilities: instead the challenger is allowed to resolve any dilemma as seems most favourable to him.

10. 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 of 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 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.

I have also shown how we can use these ideas in a dialectical setting by sketching extensions to a dialogue game originally proposed for determining credulous acceptance in a standard AF to allow the challenger to choose both value orderings and facts to be assumed. Note that it is giving these choices to the audience rather than the speaker that enables the case for an argument to be persuasive, rather than simply defensible.

All of this has been illustrated by a running example relating to a well known moral dilemma.

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. The dialogue games proposed allow these disputes to be modelled in a natural manner.

Acknowledgement

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] Amgoud, L., and Cayrol, C., (1998). *On the Acceptability of Arguments in Preference-Based Argumentation*, in Cooper, G., and Moral, S., (eds), Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence.

[2] Bench-Capon, T.J.M., *Value Based Argumentation Frameworks*. In Proceedings of Non Monotonic Reasoning 2002, pp444-453.

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

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

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

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

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

[8] Dunne, P.E., and Bench-Capon, T.J.M., (2001). *Coherence in Finite Argument Systems*. Research Report ULCS-01-006, Department of Computer Science, The University of Liverpool.

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

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