Agreeing to Differ: Modelling Persuasive Dialogue Between Parties With Different Values

T.J.M. Bench-Capon
Department of Computer Science
The University of Liverpool
Liverpool,
UK

Abstract. In some cases of disagreement, particularly in areas of practical reasoning such as ethics and law, it is impossible to provide any proof or other conclusive demonstration. The role of argument in such cases is to *persuade* rather than to prove, demonstrate or refute. Drawing on ideas of Perelman, we argue that persuasion in such cases relies on a recognition that the strength of an argument in these situations will vary according to its audience, and the strength accorded to an argument by an audience depends on the comparative weight that that audiences gives to the social values that the argument advances, when compared with the social values advanced by arguments competing with it. To model this we introduce the notion of Value Based Argumentation Frameworks (*VAFs*), an extension of Argumentation Frameworks as originally introduced by Dung. After defining *VAFs* we state some of their properties, and in particular we recall 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 agree to accept arguments, even when they differ as to which values are more important. We then describe a dialogue game based on *VAFs*, designed to model persuasive argumentation, which we illustrate with a widely discussed ethical problem.

Keywords: Argumentation, persuasion, dialogue games, practical reasoning

1. Introduction

Why do rational people disagree? There are many reasons. It may be through ignorance: if one of the parties is unaware of a crucial piece of information, they may refrain from drawing a conclusion until they discover it. It may be through weakness: it may be that one party, although in full possession of all the relevant information is incapable of drawing some required inference. It may be through deliberate fault: one party may simply refuse to accept a conclusion that has been demonstrated, although here rationality is called into question. Such disagreements can be resolved through education, explanation, or goodwill. Sometimes, however, the dispute may seem to be irreconcilable: both parties agree on the facts, are capable of making the required inferences, and be reasonable seekers after truth, and yet still they disagree. As Perelman, whose *New Rhetoric* (Perelman & Olbrechts-Tyteca 1969) has been highly influential in informal argument, puts it:

> "If men [sic] 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." (Perelman 1980 p. 150, italics mine, to indicate that we are concerned with practical reasoning here.)

It is to resolve this kind of disagreement that the need for argumentation, intended to secure assent through persuasion rather than intellectual coercion, arises. For example: many would argue that more money must be spent on UK Universities if standards are to be maintained. But this is resisted by the UK Government, as to do so would involve raising taxes. From the Government perspective, this is sufficient to defeat the academic argument. But it is not sufficient from the academic perspective: they recognise that the argument attacks their own, but deny that it is of sufficient force to defeat it. Neither party is irrational: it all depends on whether maintaining University standards is valued more than the social values promoted by leaving the tax rate unchanged. Similarly in law, disputes often come down to a clash of values. Perelman (1980) 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 rightly emphasises 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. Since arguments derive their force from the values they promote, this means that whether an argument is accepted is a function of the audience to which it is addressed as well as the argument itself. But although differences in the values of different audiences may mean that it is rational to differ, it is not necessarily the case that a difference in values will lead to disagreement. There can often be points of rational agreement, even if we allow the strength of an argument to be determined by the value it promotes. Indeed in some cases, we can show that an argument must be rationally accepted, however one ranks the values involved. In this paper we want to explore the notion of persuasion in the face of divergent values.

We will begin our exploration with the notion of an *Argumentation Framework* (*AF*) introduced in Dung (1995), which has proved to provide a fruitful way of looking at systems of conflicting argument. AFs 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 Bench-Capon (2002, under review), I have extended *AFs* to *Value Based Argumentation Frameworks* (*VAF*), which attempt to represent the kind of use of values to ground rational disagreement described above. I have shown there how *VAFs* can be used to resolve disputes which are undecideable in standard *AFs*.

Here, I will first recapitulate the standard notion of an *AF*, and next show how we can incorporate the notions of value and audience by showing how the key concepts are defined in a *VAF*. I will then draw attention to some of the important properties of *VAFs* established in previous work. I then describe a dialogue game based on *VAFs* which can be used to model persuasive dialogues when values differ. I will conclude with a small example.

2. Argumentation Frameworks

First let us recall Dung's original definition of Argumentation Frameworks. For Dung the notion of an argument is highly abstract: arguments are characterised only by the arguments they attack and are attacked by. This is especially suitable for modelling informal, natural language arguments, since the arguments are unconstrained in form, and there are no restrictions on what we can choose to count as an attack of one argument on another.

A formal definition of an Argumentation Framework, and the central notions concerning Argumentation Frameworks, is given as Definition 1.

Definition 1: An Argumentation Framework (*AF*) is a pair $AF = \langle X, A \rangle$, where X is a set of arguments and $A \subseteq X \times X$ is the attack relationship for *AF*. A comprises a set of ordered pairs of distinct arguments in X. A pair $\langle x,y \rangle$ is referred to as "*x attacks y*".

For R, S, subsets of X, we say that

(a) $s \in S$ is attacked by R if there is some $r \in R$ such that $\langle r,s \rangle \in A$.
(b) $x \in X$ is *acceptable* with respect to S if for every $y \in X$ that attacks x there is some $z \in S$ that attacks y (i.e. z, and hence S, defends x against y).
(c) S is *conflict free* if no argument in S is attacked by any other argument in S.
(d) A conflict free set is *admissible* if every argument in S is acceptable with respect to S.
(e) S is a *preferred extension* if it is a maximal (with respect to set inclusion) admissible subset of X.

A useful way to picture an *AF*, to which we will appeal on occasion, is as a directed graph with arguments as vertices and arcs representing the attacks relation.

The key notion here is the *preferred extension* which represents a position which is

- internally consistent
- can defend itself against all attacks
- cannot be further extended without becoming inconsistent or open to attack.

From Dung (1995) we know that every *AF* has a preferred extension (possibly the empty set if a cycle of odd length exists in *AF*), and that it is not generally true that an *AF* has a unique preferred extension. In fact any *AF* that contains a cycle of even length may have multiple preferred extensions (see Bench-Capon (2002) for a proof). 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. Where there are multiple preferred extensions, we can 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.

Note that in the standard argumentation framework, an attack will always succeed. While this seems well adapted for reasoning about matters of fact and formal systems such as mathematics, 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 ways 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 any practical force, it must be accepted that G is a good. Here we will always assume that the values advanced by arguments are *prima facie* acceptable, that they do have some force for all parties concerned. We will therefore focus on the attacks which depend on the relative weight of the values.

Once we allow that arguments may have different strengths, we open the possibility that an attack can fail, since the attacked argument may be stronger than its attacker. Thus, if an argument attacks an argument whose value is preferred it can be accepted, and yet not defeat the argument it attacks. To represent this possibility of unsuccessful attacks we must extend the standard argumentation framework so as to include the notion of value.

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 $a \in AR$, $val(a) \in V$.

The set *P* of audiences is introduced because, following Perelman, we want to be able to make use of the notion of an audience. We see audiences as individuated by their preferences between values, since if there is agreement on the ranking of values, there will be agreement on which attacks succeed. 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 3: 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) $Valpref_a \subset V \times V$, 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 $a \hat{I} P$ 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 4: *An argument* $A \hat{I} AF$ *defeats*$_a$ *an argument* $B \hat{I} AF$ *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 practical disputes can in fact 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.

We next define the other notions associated with an *AF* for a *VAF*,

Definition 5: An argument $A \in AR$ is *acceptable to audience a* (*acceptable*$_a$) with respect to set of arguments *S, (acceptable*$_a$*(A,S))* if:

$(\forall x)((x \hat{I} AR \& defeats_a(x,A)) \rightarrow (\exists y)((y \hat{I} S) \& defeats_a(y,x)))$.

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

$(\forall x)(\forall y)((x \hat{I} S \& y \hat{I} S) \rightarrow (\varnothing attacks(x,y) \acute{U} valpref(val(y),val(x)) \hat{I} valpref_a)))$.

Definition 7: A *conflict-free for audience a* set of arguments *S* is *admissible for an audience a* if

$(\forall x)(x \hat{I} S \rightarrow acceptable_a(x,S))$.

Definition 8: 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* subset 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 in particular 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. Because multiple preferred extensions can only arise from even cycles, and empty preferred extensions only from odd cycles, both af_a and vaf_a will have a unique, non-empty, preferred extension for such cases.

3. Properties of VAFS

In what follows we will restrict ourselves to *VAFs* which do not contain any cycles in which all the arguments relate to the same value. In practice we believe that this is not an undue restriction, and that such single value cycles are generally ill formed, except where we have a paradox or some inescapable dilemma, which will preclude rational resolution. As noted above, such *VAFs* will have a unique, non-empty preferred extension for each of its audiences. Consider, for example, the VAF comprising a four cycle in Figure 1.

Here there are two potential audiences representing the ordering v1 > v2 and the ordering v2 > v1. For the first audience the preferred extension will be {a,c} and for the second {b,d}. No agreement is possible here. But suppose the same audiences are considering the VAF in Figure 2. Now both audiences will have the preferred extension {a.c}. Thus even though they disagree as to values, they

will agree as to which arguments are accepted. Thus we can say that some arguments can be acceptable irrespective of how values are ranked: that is acceptable to all audiences.

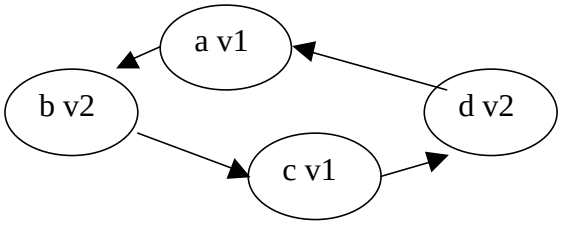

Figure 1: 4-cycle with alternating values

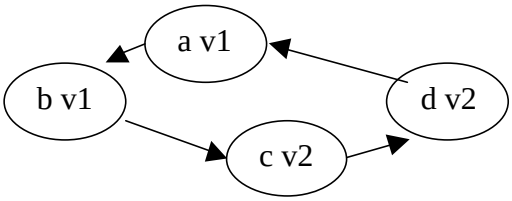

Figure 2: 4-cycle with connected values

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

Definition 9: *Objective Acceptance*. Given a *VAF, <AR, attacks,V,val,P>*, an argument $A \hat{I} AR$ is objectively acceptable if and only if for all $p \hat{I} P$, A is in every $preferred_p$.

Definition 10: *Subjective Acceptance*. Given a *VAF, <AR, attacks,V,val,P>*, an argument $A \hat{I} AR$ is subjectively acceptable if and only if for some $p \hat{I} P$, A is in some $preferred_p$.

An argument which is neither objectively nor subjectively acceptable is said to be *indefensible*.

We now introduce the notion of a *line of argument*.

Definition 11: A line of argument is a set L of n arguments $\{a_1 \ldots a_n\}$ such that:

i. a_1 *has no attacker in L;*

ii *For all* $a_i \hat{I} L$ *if* $i > 1$, *then* a_i *is attacked and the sole attacker of* a_i *is* a_{i-1}

In the special case where all the arguments in a line of argument have the same value, we call it an *argument chain*.

Definition 12: An *argument chain* in a *VAF, C* is a set of n arguments $\{a_1 \ldots a_n\}$ such that:

i. $(\forall a)(\forall b)(a \hat{I} C \& b \hat{I} C) \rightarrow val(a) = val(b))$;

ii. a_1 *has no attacker in C;*

iii. *For all* $a_i \hat{I} C$ *if* $i > 1$, *then* a_i *is attacked and the sole attacker of* a_i *is* a_{i-1} .

Clearly in an argument chain, the status of every argument depends on the status of the first argument: if that argument is accepted so are all the odd numbered arguments, whereas if it is rejected, all the even numbered arguments are accepted.

Using this notion we can come up with a characterisation of the status of arguments in a VAF considered as a set of argument chains. See Bench-Capon (2002) for a justification.

i *an argument is indefensible if it is an even numbered member of any chain preceded only by even chains; or if it is an even numbered member of a chain attacked by an odd chain, and is directly attacked by an odd chain;*

ii *an argument is objectively acceptable if it is only an odd numbered argument of a chain preceded only by even chains;*

iii *an argument is subjectively acceptable otherwise.*

An unattacked argument is considered to be preceded by a chain of length zero, hence an even chain.

Turning to lines of argument, we can discover a very useful restriction on the extent to which we need to follow the line of argument. If we are considering the an argument in a line of reasoning, we need to consider its attacker to discover its status. If the attacker has a different value, no argument relating to the original value can affect the status of the original argument. Suppose we have a line of reasoning with two values x and y. which runs x? y? x. If x is preferred to y, then the first argument will not be

defeated. But if y is preferred to x, the first argument will be defeated, and cannot be rescued by the third argument since its attack will fail. Since we need never consider a line of argument back beyond the point at which a value is reintroduced, we can considerably shorten the task of establishing the status of an argument.

Finally we should note that we have an efficient algorithm (given in Bench-Capon (2002)) to establish the preferred extension given a value ordering. Thus determining objective acceptance is always tractable, for a small number of values.

4. 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 with respect to them, 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, (Dunne & Bench-Capon, 2001) 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. (Dunne & Bench-Capon, 2001) 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 describing the game in (Dunne & Bench-Capon, 2001), since it presents features that I wish to incorporate in the persuasion games described below. Let us call this game *CA* (for credulous acceptance). CA allows only three moves: COUNTER, BACKUP and RETRACT. The game has two players, Defender (Def) and Challenger (Chal). Def begins play by stating an argument which he wishes to defend. Chal 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 Chal 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 Def. RETRACT is made only by Def; it involves returning to the original claim, and means that the subset of arguments played by Def so far cannot be recreated. CA is won by Def if a preferred extension including the argument in dispute is created, and by Chal if this proves impossible.

An example dispute using this game given in (Dunne & Bench-Capon, 2001) is based on the *AF* shown in Figure 3. 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 Def at k, C_k are the arguments available to Chal at k, P_k are the arguments proposed by Def as a (subset) of some admissible set, and Q_k are the set of subsets that Chal has shown not to be subsets of an admissible set at k.

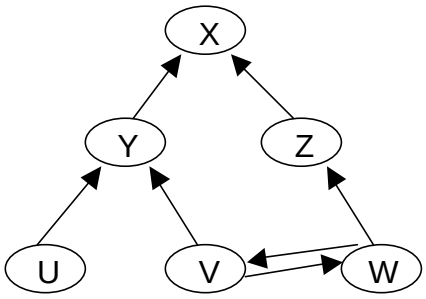

Figure 3: AF for Dispute Example

A possible play of the game relating to the *AF* in Figure 3 would run as follows. Def claims X, which is attacked by Chal with Y. Def attacks Y with V. Chal now chooses to back up and attack X with Z. Def cannot now play W, because this is attacked by the already played V. Def must therefore retract. Chal again attacks X with Y, but this time Def defends by attacking Y with the unassailable U. Chal has no choice but to back up and try the attack with Z. This time W is available to Def to attack Z, and Chal cannot attack with V, since it is already attacked by W. Therefore Chal 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 3

k	player	$move_k$	v_k	D_k	C_k	P_k	Q_k
0	Def	-	X	{U,V,W}	{Y,Z,U,V,W}	{X}	{}
1	Chal	C(Y)	Y	{U,V,W}	{Y,Z,U,V,W}	{X}	{}
2	Def	C(V)	V	{U}	{Z,U}	{X.V}	{}
3	Chal	B(0,Z)	Z	{U}	{U}	{X.V}	{}
4	Def	R	X	{U,V,W}	{Y,Z,U,V,W}	{X}	{X,V}
5	Chal	C(Y)	Y	{U,V,W}	{Z,U,V,W}	{X}	{X,V}
6	Def	C(U)	U	{V,W}	{Z,V,W}	{X,U}	{X,V}
7	Chal	B(4,Z)	Z	{V,W}	{V,W}	{X,U}	{X,V}
8	Def	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 Def is successful he retains the right to accept his claim, but Chal need not accept it, since there 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 above that there is no place for monochromatic cycles in a *VAF* – the preferred extension is unique for any given value ordering. In order that Chal may be persuaded, Chal 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 Chal 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;
- Chal 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 Chal 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 following scenario. The scenario we will consider is taken from an example discussed by Coleman (1992) and further discussed by Christie in (2000). 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?

The VAF is shown in Figure 4. As presented by Coleman, 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 (A). But should Hal compensate Carla? His justification can be attacked by an argument that it is wrong to infringe the property rights of another (B). If, however, Hal compensates Carla, we have a property based argument that Carla's rights have not been infringed (C). Christie, however, does not want to insist on compensation. He therefore introduces a fourth argument 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 (D). 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, with {A,B,D} as the accepted arguments..

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.

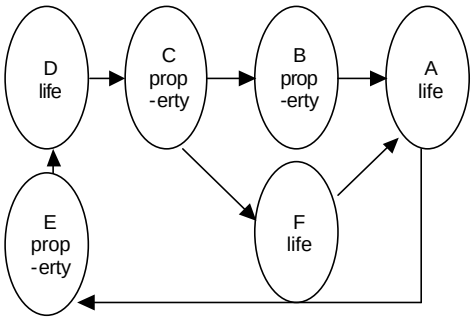

Figure 4: Hal and Carla scenario

Let us first consider it as a value free dispute using the original game CA.

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

Now consider the dispute using values. Again Chal may counter (A) by using either (B) or (F), and Def counters either of these with (C). But now (D) is not available to Chal, since it would repeat the value *life*. Therefore Chal 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). At this point Def has no way to persuade Chal, since (C) is no longer available to attack (F), because of the declared value preference. Had Chal played (B) initially the correct response would have been BACKUP(0,F).

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

In this game, persuasion is possible only if the claim is objectively acceptable: otherwise Chal 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 each participant is 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 challenge successfully some of his opponent's claims. For example, in Figure 4, suppose that Def wishes to say Hal has an absolute right to take the insulin, and Chal wishes to argue that Hal can do so only if he pays compensation. Now Chal must choose to commit to *property > life* in order to defend (C) against the attack of (D). Once this is done, Chal can no longer express the different value preference to attack (A): thus (C) will defeat (F). Here Chal is obliged to accept Def's argument in order to save his own, and Def is obliged to accept Chal's argument or surrender his own. The preferred extensions of Figure 4 are {B,D,E,F} if life is preferred to property, and {A,C,E} if property if preferred to life. (Note in passing that it is the one argument enshrined in law that appears in every preferred extension.) Thus anyone who wishes to defend (A), must also allow (C), and *vice versa*. We thus have a situation of mutual persuasion. 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.

Another issue, which we shall discus here, is that many disputes mix values and facts. In Bench-Capon, (under review.) we show how factual and value based arguments can be mixed, and also show how to allow for uncertainty as to the facts.

6. Summary

Our aim in this paper has been to explore issues of agreement and disagreement in situations where the acceptability of arguments depends on the audiences which receive them, and on the way they rank the values promoted or defended by the arguments. Where the two parties to a dispute represent different audiences, it is possible that disagreement is rational, since the acceptability of the arguments may

depend on the way in which values are ranked. Equally, however, some arguments can be shown to be acceptable to all audiences, opening up the possibility of persuasion.

In order to explore these questions we have first presented a formal framework which allows us to represent the notions of values promoted by arguments and audiences with preferences as to the ranking of values. With this framework we can establish which arguments have a status dependent on the audience, and which are acceptable to all audiences, and tractable algorithms exist to enable us to do so. Clearly, persuasion is possible for arguments which are acceptable to every audience. The properties of this framework have been discussed in detail elsewhere Bench-Capon (2002, under review), Dunne & Bench-Capon (2002).

We then introduced a dialogue game to model persuasive argument against a background of values. The game highlighted a second situation where persuasion is possible. It may be that it is possible to force a value ranking on someone in order for them to maintain a desired position, and then to use this ranking to demonstrate that some other argument must also be accepted.

Much remains to be considered if we are to get a full account of persuasion against a background of divergent values. I believe, however, that the framework put forward here will prove a fruitful tool in this exploration.

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.. (2002). *Value Based Argumentation Frameworks*. In Proceedings of Non Monotonic Reasoning 2002, pp444-453.

[3] Bench-Capon, T.J.M. (under review). Persuasion in Practical Argument Using Value Based Argumentation Frameworks. Technical Report ULCS-02-017, Department of Computer Science, The University of Liverpool. Submitted to *Journal of Logic and Computation*.

[4] 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.

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

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

[7] 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.

[8] 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. Submitted to *Artificial Intelligence*.

[9] Dunne, P.E., and Bench-Capon, T.J.M., (2002). *Coherence in Finite Argument Systems*. *Artificial Intelligence*, September 2002.

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

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