Chapter 3
Abstract Argumentation and Values

Trevor Bench-Capon and Katie Atkinson

1 Introduction

Abstract argumentation frameworks, as described in Chapter 2, are directed towards determining whether a claim that some statement is true can be coherently maintained in the context of a set of conflicting arguments. For example, if we use preferred semantics, that an argument is a member of all preferred extensions establishes that its claim must be accepted as true, and membership of at least one preferred extension shows that the claim is at least tenable. In consequence, that admissible sets of arguments are conflict free is an important requirement under all the various semantics.

For many common cases of argument, however, this is not appropriate: two arguments can conflict, and yet both be accepted. For an example suppose that Trevor and Katie need to travel to Paris for a conference. Trevor offers the argument "we should travel by plane because it is quickest". Katie replies with the argument "we should travel by train because it is much pleasanter". Trevor and Katie may continue to disagree as to how to travel, but they cannot deny each other's arguments. The conclusion will be something like "we should travel by train because it is much pleasanter, even though travelling by plane is quicker". The point concerns what Searle [24] calls *direction of fit*. For matters of truth and falsity, we are trying to fit what we believe to the way the world actually is. In contrast, when we consider what we should do we are trying to fit the world to the way we would like it to be. Moreover, because people may have different preferences, values, interests and aspirations, people may rationally choose different options: if Katie prefers comfort

Trevor Bench-Capon
Department of Computer Science, University of Liverpool, UK, e-mail: tbc@liverpool.ac.uk

Katie Atkinson
Department of Computer Science, University of Liverpool, UK, e-mail: katie@liverpool.ac.uk

I. Rahwan, G. R. Simari (eds.), *Argumentation in Artificial Intelligence,*
DOI 10.1007/978-0-387-98197-0_3,

to speed she will rationally choose the train, but this does not mean that Trevor cannot rationally choose the plane if he prefers speed to comfort. We will return to this example throughout this chapter.

Within standard abstract argumentation frameworks one approach to recognising the importance of direction of fit [22] is to require sceptical acceptance for epistemic arguments but only credulous acceptance for practical arguments. This does successfully model the existence of a choice with respect to practical arguments, but it does not motivate the choice, nor does it allow us to predict choices on the basis of choices made in the past. Value based argumentation frameworks (VAFs) [6], described in this chapter, are an attempt to address issues about the rational justification of choices systematically.

Value based justification of choices is common in many important areas: in politics where specific policies are typically justified in terms of the values they promote, and where politicians' values are advanced as reasons to vote for them; in law, where differences in legal jurisdictions and decisions over time can be explained in terms of the values of the societies in which the judgements are made [11]; in matters of morality where individual and group ethical perspectives play a crucial role in reasoning and action [2]; as well as more everyday examples, such as given above.

In this chapter we will first give some philosophical background, in particular introducing the notion of *audience*, and some of the features that we require from practical reasoning. Section 3 will discuss the nature of values in more detail, in particular the distinction between values and goals. Section 4 will introduce the formal machinery of Value Based Argumentation Frameworks, and discuss some of their more important properties. Section 5 describes some applications of value based argumentation. Section 6 discusses some recent developments, and section 7 concludes the chapter with a summary.

2 Audiences

One of the first people to stress the importance of the audience in determining whether an argument is persuasive or not was Chaim Perelman [20], [19]:

> "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." [[19] p.150, italics ours].

A similar point was made by John Searle [24]:

> "Assume universally valid and accepted standards of rationality, assume perfectly rational agents operating with perfect information, and you will find that rational disagreement will still occur; because, for example, the rational agents are likely to have different and inconsistent values and interests, each of which may be rationally acceptable." [[24], xv]

Both Perelman and Searle recognise that there may be complete agreement on facts, logic, which arguments are valid, which arguments attack one another and the

rules of fair debate, and yet still disagreement as to the correct decision. This was true when Trevor and Katie were thinking about how to travel to Paris, and there are many other examples. Consider an example from politics.

One choice that any government must make is to decide on an appropriate rate of income tax. Typically there will be an argument in favour of increasing the rate of taxation, since this progressive form of taxation will reduce income inequalities. Against this, it can be argued that a decrease in taxation will promote more enterprise, increasing Gross National Product, and so raising the absolute incomes of everyone. It is possible to see both these arguments as valid, since both supply a reason to act: and yet a choice must be made, since the actions are incompatible. Which choice is made will depend on whether the chooser prefers equality or enterprise in the particular circumstances with which he is confronted. Two parties may be in agreement as to the consequences of a movement in the tax rate, and yet disagree as to the choice to be made because they differ in their fundamental aspirations. Different people will prize social values differently, and one may prefer equality to enterprise, while another prefers enterprise to equality. Thus while both arguments are agreed to be valid, one *audience* will ascribe more force to one of the arguments, while a different *audience* will make a different choice. In such cases these different audiences will rationally disagree, and agreement can only be reached by coming up with additional arguments which convince all audiences *in terms of their own preferences*, or by converting those who disagree to a different appraisal of social values. This will often require that different arguments be presented to different audiences. Thus when in the 1980s the UK Conservative Party under Margaret Thatcher were attempting to justify dramatic cuts in income tax for the highly paid, one argument was simply that fairness meant that people deserved to keep a larger proportion of their "earnings". This argument was quite acceptable at Party Conferences where the audience comprised predominately high earners, but was not persuasive to the country at large, since most people were not subject to higher rate taxation. To convince the nation at large a different argument was needed: namely that there would be a "trickle down" effect, benefitting everyone, whatever their level of income. This was clearly persuasive as Thatcher was twice re-elected.

Thus whether an argument is persuasive depends not only on the intrinsic merits of the argument – of course, it needs to be based on plausible premises and must be sound – but also on the audience to which it is addressed. Moreover, for practical reasoning, what is important about the audience is what they want to see happen, and this seems to turn on how they rank the various values that accepting the arguments promote. In the next section we will consider values, and their relation to practical reasoning, in more detail.

3 Values

This far we have seen that whether a particular audience is persuaded by an argument depends on the attitude of that audience to the values on which the argument is

founded. Values are used in the sense of fundamental social or personal goods that are desirable in themselves, and should never be confused with any numeric measure of the strength, certainty or probability of an argument. Liberty, Equality and Fraternity, the values of the French Revolution, are paradigmatic examples of values. Values are widely recognised as the basis for persuasive argument. For example, the National Forensic League, which conducts debating competitions throughout the USA uses the "Lincoln-Douglas" (LD) debate format which is based on the notion of a clash of values. In an LD debate the resolution forces each side to take on competing values and argue about which one is supreme. For example, if the resolution is, "*Resolved*: An oppressive government is better than no government at all," the affirmative side might value "order" and the negative side might value "freedom". Such a debate would revolve around whether order is more valuable than freedom. In the original debate between Abraham Lincoln and Stephen Douglas on which LD debates are based, Douglas championed the rights of states to legislate for their particular circumstances, whereas Lincoln argued on the basis that there were certain inviolable human rights that all states had to respect, even though this constrained state autonomy.

But what is the role of values in practical reasoning? Historically the basis for treatments of practical reasoning has been the practical syllogism, first discussed by Aristotle. A standard modern statement is given in [16]:

K1 I'm to be in London at 4.15
If I catch the 2.30, I'll be in London at 4.15
So, I'll catch the 2.30.

The first premise is a statement of some desired state of affairs, the second an action which would bring about that state of affairs, and the conclusion is that the action should be performed. There are, however, problems with the formulation: it is abductive rather than deductive, does not consider alternative, possibly better, ways of achieving the desired state of affairs, or possibly undesirable side effects of the action. Walton [26] addresses these issues by regarding the practical syllogism as an argumentation scheme, which he calls the *sufficient condition scheme for practical reasoning*, which provides a presumptive reason to perform the action, but which can be critiqued on the basis of alternatives and undesirable consequences. He states the sufficient condition practical reasoning scheme as:

W2 G is a goal for agent *a*
Doing action A is sufficient for agent *a* to carry out goal G
Therefore agent *a* ought to do action A.

This, however, still does not explain why G is a goal for the agent, nor indicate how important bringing about G is to the agent. Neither K1 nor W1 make any mention of values: rather that the agent has certain values is implicit in calling the desired state of affairs a "goal" for that agent. Accordingly, to make this role of values explicit, Walton's scheme was developed by Atkinson and her colleagues [4] into the more elaborated scheme:

A1 In the circumstances R, we should perform action A
to achieve new circumstances S, which will realise some goal G
which will promote some value V.

What this scheme does in particular is to distinguish three aspects which are conflated into the notion of goal in K1 and W1. These aspects are: the state of affairs which will result from the action; the goal, which is those aspects of the new state of affairs for the sake of which the action is performed; and the *value*, which is the reason why the agent desires the goal. Making these distinctions opens up several distinct types of alternative to the recommended action. We may perform a different action to realise the same state of affairs; we may act so as to bring about a different state of affairs which realises the same goal; or we may realise a different goal which promotes the same value. Alternatively, since the state of affairs potentially realises several goals, we can justify the action in terms of promoting a different value. In coming to agreement this last possibility may be of particular importance: we may want to promote different values, and so agree to perform the action on the basis of different arguments. Our contention is that, in the spirit of the notion of audience developed in section 2, what is important, what is the appropriate comparison for choosing between alternatives, is the value.

In order to see the distinction between a goal and a value, consider again Trevor and Katie's journey to Paris. The *goal* is to be in Paris for the conference, and this is not in dispute: the dispute is how that goal should be realised and turns on the values promoted by the different methods of travel. What is important is not the state reached, but the way in which the transition is made.

The style of argumentation represented by A1 has been formalised in [1] in terms of a particular style of transition system, Alternating Action Based Transition systems (AATS) [27]. An AATS consists of a set of states and a set of agents and the transitions between the states are in terms of the *joint actions* of the agents, that is, actions composed from the actions available to the agents individually. In terms of A1 the circumstances R and S are represented by the states of the system, the goal G is realised if G holds in S (of course, G may hold in several of the states), and the action is the particular agent's component of a joint action which is a transition from R to S. The value labels the transition, indicating that it is the movement from R to S *using that particular transition* that promotes the value. A fragment of the AATS for Trevor and Katie's travel dilemma is shown in Figure 3.1, t/kt/p is the action of Trevor/Katie travelling by train/plane, C/St/k means that Comfort/Speed is promoted in respect of Trevor/Katie, 00 that both are in Liverpool and 11 that both are in Paris.

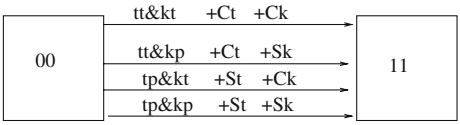

Fig. 3.1 AATS for travel to Paris example

Although there is only one destination state, each of the four potential ways of reaching it promotes different values, and hence give rise to different arguments in their favour. Which arguments will succeed will depend on the preferences between the values of Comfort and Speed of the two agents concerned.

Essentially then, in this problem there will be a number of possible audiences, depending on how the values are ordered. Suppose that Trevor values his own speed over his own comfort and Katie her comfort over her speed, and that neither consider values promoted in respect of the other. Then Trevor will choose to go by plane and Katie by train. Here the agents can choose independently, as their values are affected only by their own actions: in later sections we will introduce a third value which requires them to consider what the other intends to do also.

The basic idea underlying Value Based Argumentation Frameworks is that it is possible to associate practical arguments with values, and that in order to determine which arguments are acceptable we need to consider the audience to which they are addressed, characterised in terms of an ordering on the values involved. We need, however, to recognise that not all the arguments relevant to a practical decision will be practical arguments. For example, if there is a train strike (or it is a UK Bank Holiday when there are often no trains from Liverpool), the argument that the train should be used cannot be accepted no matter how great the audience preference is for Comfort over Speed. In order to recognise that such epistemic arguments constrain choice, such arguments are associated with the value *Truth*, and all audiences are obliged to rank Truth above all other values. In the next section we will give a formal presentation of Value Based Argumentation Frameworks.

4 Value Based Argumentation Frameworks

We present the Value Based Framework as an extension of Dung's original Argumentation Framework [14], defined in Chapter 2 of this book. We do this by extending the standard pair to a 5 tuple.

Definition 3.1. A *value-based argumentation framework* (*VAF*) is a 5-tuple:
VAF = $<\mathcal{A}, \mathcal{R}, V, val, P>$

where $\mathcal{A}$ is a finite set of arguments, $\mathcal{R}$ is an irreflexive binary relation on $\mathcal{A}$ (i.e. $<\mathcal{A}, \mathcal{R}>$ is a standard AF), V is a non-empty set of values, *val* is a function which maps from elements of $\mathcal{A}$ to elements of V and P is the set of possible audiences (i.e. total orders on V). 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 $a \in \mathcal{A}$, $val(a) \in V$.

When the VAF is considered by a particular audience, the ordering of values is fixed. We may therefore define an Audience Specific VAF (AVAF) as:

Definition 3.2. An *audience specific value-based argumentation framework* (*AVAF*) is a 5-tuple: $\text{VAF}_a = <\mathcal{A}, \mathcal{R}, V, val, Valpref_a>$

where $\mathcal{A}$, $\mathcal{R}$, 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 $\mathcal{A}$, $\mathcal{R}$, V and *val* are identical, and *Valpref* is the set of preferences derivable from the ordering $a \in P$ in the *VAF*.

Our purpose in introducing VAFs is to allow us to distinguish between one argument *attacking* another, and that attack *succeeding*, so that the *attacked* argument may or may not be defeated. Whether the attack succeeds depends on the value order of the audience considering the VAF. We therefore define the notion of *defeat for an audience*:

Definition 3.3. *An argument $A \in AF$ $defeats_a$ an argument $B \in AF$ for audience a if and only if both $\mathcal{R}(A,B)$ and not $(val(B),val(A)) \in Valpref_a$.*

We can now define the various notions relating to the status of arguments:

Definition 3.4. An argument $a \in \mathcal{A}$ is *acceptable-to-audience-a* ($acceptable_a$) with respect to set of arguments S, ($acceptable_a(A,S)$) if:
$(\forall x)((x \in \mathcal{A} \ \& \ defeats_a(x,A)) \rightarrow (\exists y)((y \in S) \ \& \ defeats_a(y,x)))$.

Definition 3.5. A set S of arguments is *conflict-free-for-audience-a* if:
$(\forall x)(\forall y)((x \in S \ \& \ y \in S) \rightarrow (\neg \mathcal{R}(x,y) \vee valpref(val(y),val(x)) \in Valpref_a)))$.

Definition 3.6. A *conflict-free-for-audience-a* set of arguments S is *admissible-for-an-audience-a* if: $(\forall x)(x \in S \rightarrow acceptable_a(x,S))$.

Definition 3.7. A set of arguments S in a value-based argumentation framework *VAF* 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 $\mathcal{A}$.

Now for a given choice of value preferences $valpref_a$ we are able to construct an *AF* equivalent to the *AVAF*, by removing from $\mathcal{R}$ those attacks which fail because they are faced with a superior value.

Thus for any *AVAF*, $vaf_a = <\mathcal{A}, \mathcal{R}, V, val, Valpref_a>$ there is a corresponding *AF*, $af_a = <\mathcal{A}, defeats>$, such that an element of $\mathcal{R}$, $\mathcal{R}(x,y)$ 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. A proof is given in [6]. Moreover, since the AF derived from an AVAF contains no cycles, the grounded extension coincides with the preferred extension for this audience, and so there is a straightforward polynomial time algorithm to compute it, also given in [6]. For the moment we will restrict consideration to VAFs which do not contain any cycles in a single value.

For such VAFs, the notions of sceptical and credulous acceptance do not apply, since any given audience will accept only a single preferred extension. These preferred extensions may, and typically will, however, differ from audience to audience. We may therefore introduce two useful notions, objective acceptance, arguments which are acceptable to all audiences irrespective of their particular value order, and subjective acceptance, arguments which can be accepted by audiences with the appropriate value order.

Definition 3.8. *Objective Acceptance*. Given a *VAF*, $<\mathcal{A}, \mathcal{R}, V, val, P>$ an argument $a \in \mathcal{A}$ is objectively acceptable if and only if for all $p \in P$, a is in every $preferred_p$.

Definition 3.9. *Subjective Acceptance*. Given a *VAF*, $<\mathcal{A}, \mathcal{R}, V, val, P>$ an argument $a \in \mathcal{A}$ is subjectively acceptable if and only if for some $p \in P$, a is in some $preferred_p$.

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

All arguments which are not attacked will, of course, be objectively acceptable. Otherwise objective acceptance typically arises from cycles in two or more values. For example, consider a three cycle in two values, say two arguments with V1 and one with V2. The argument with V2 will either resist the attack on it when it is preferred to V1, or, when V1 is preferred, fail to defeat the argument it attacks which will, in consequence, be available to defeat its attacker. Thus the argument in V2 will be objectively acceptable, and both the arguments with V1 will be subjectively acceptable. For a more elaborate example consider Figure 3.2.

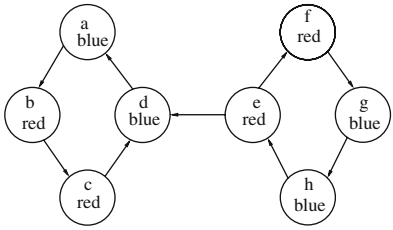

Fig. 3.2 VAF with values red and blue

There will be two preferred extensions, according to whether red $>$ blue, or blue $>$ red. If red $>$ blue, the preferred extension will be {e,g,a,b}, and if blue $>$ red, {e,g,d,b}. Now e and g and b are objectively acceptable, but d, which would have been objectively acceptable if e had not attacked d, is only subjectively acceptable (when blue $>$ red), and a, which is indefensible if d is not attacked, is also subjectively acceptable (when red $>$ blue). Arguments c, f and h are indefensible. Results characterising the structures which give rise to objective acceptance are given in [6].

4.1 VAF Example

We will illustrate VAFs using our running example of Trevor and Katie's conference travel arrangements. Recall that $VAF_a = <\mathcal{A}, \mathcal{R}, V, val, Valpref_a>$. We therefore need to instantiate the five elements of this tuple.

From Figure 3.1 above we get four arguments:

A1: Katie should travel by train (Kt) to promote her comfort (Ck).

A2: Katie should travel by plane (Kp) to promote her speed (Sk).

A3: Trevor should travel by train (Tt) to promote his comfort (Ct).

A4: Trevor should travel by plane to (Tp) to promote his speed (St).

But there are other considerations: it is far more boring to travel alone than in company. This gives two other arguments:

A5: Both Katie and Trevor should travel by train (Kt&Tt) to avoid boredom (B).

A6: Both Katie and Trevor should travel by plane (Kp&Tp) to avoid boredom (B).

Thus $\mathcal{A}_e$ = {A1,A2,A3,A4,A5,A6} and val = {A1→Ck, A2→Sk, A3→Ct, A4→St, A5→B, A6→B}.

We can now identify attacks between these arguments. Since neither Katie nor Trevor can travel by both train and plane, A1 attacks A2, and vice versa, and A3 attacks and is attacked by A4. Moreover A1and A3 attack and are attacked by A6, and A2 and A4 attack and are attacked by A5.

Thus $\mathcal{R}_e$ = {<A1,A2>, <A2,A1>, <A3,A4>, <A4,A3>, <A1,A6>, <A3,A6>, <A6,A1>, <A6,A3>, <A2,A5>, <A4,A5>,<A5,A2>,<A5,A4>, <A5,A6>, <A6,A5>}.

The values are given by the values used in the arguments, but for the present we will make no distinction at first between values promoted in respect of Trevor and values promoted in respect of Katie. Thus V_e = {B, C, S}. Finally the audiences P will be every possible ordering of the elements in V_e, so P = {B>C>S, B>S>C, S>B>C, S>C>B, C>B>S, C>S>B}

We can represent the VAF diagrammatically as a directed graph, as shown in Figure 3.3.

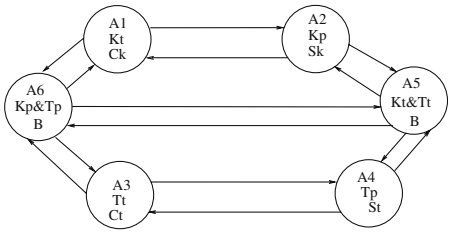

Fig. 3.3 VAF for travel example

Note that here we do have a cycle of two arguments with the same value, namely B. This means that some audiences will not have a unique preferred extension. This does not pose any serious problem in this small example.

Now consider specific audiences. Suppose that Katie, who very much dislikes flying, ranks C as her highest value, and S as her least important.

Now $\text{AVAF}_{katie} = <\mathcal{A}_e, \mathcal{R}_e, \text{V}_e, \text{val}, \{<\text{C,B}>,<\text{C,S}>,<\text{B,S}>\}>$.

When we use Katie's preferences to eliminate unsuccessful attacks, this produces the corresponding $\text{AF}_{katie} = < \mathcal{A}_e, \{<\text{A1,A2}>,<\text{A3,A4}>, <\text{A1,A6}>, <\text{A3,A6}>, <\text{A5,A2}>,<\text{A5,A4}>, <\text{A5,A6}>\}>$. This AF has a unique preferred extension, $\text{PE}_{katie} = \{\text{A1,A3,A5}\}$, which means that she will be in favour of both Trevor and herself travelling by train.

Suppose, however, Trevor, who has no objection to flying, prefers speed to comfort, but dislikes travelling alone, so that he is a member of the audience $\{\text{B}>\text{S}>\text{C}\}$.

Now $\text{AVAF}_{trevor} = <\mathcal{A}_e, \mathcal{R}_e, \text{V}_e, \text{val}, \{<\text{B,S}>,<\text{B,C}>,<\text{S,C}>\}>$.

And $\text{AF}_{trevor} = <\mathcal{A}_e, \{<\text{A2,A1}>, <\text{A4,A3}>, <\text{A6,A1}>, <\text{A6,A3}>, <\text{A5,A4}>, <\text{A5,A6}>, <\text{A6,A5}>\}>$.

This contains a cycle for the two arguments in B, and so Trevor will have two preferred extensions: {A1,A3,A5}, and {A2,A4,A6}. Trevor could solve this dilemma by considering that A3 also promotes C and A4 also promotes S, and so choose {A2,A4,A6}. But what is required is a joint decision: neither Trevor nor Katie can act independently so as to ensure that A5 or A6 is followed. We therefore need to consider the joint audience, and to distinguish between values promoted in respect of Trevor and values promoted in respect of Katie.

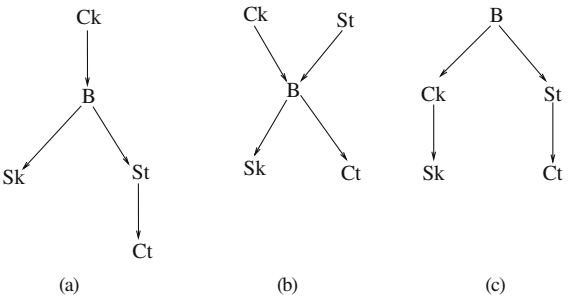

Fig. 3.4 Partial Orders representing combined audiences: (a) Katie C>B>S and Trevor B>S>C; (b) Katie C>B>S and Trevor S>B>C; (c) Katie B>C>S and Trevor B>S>C

Katie's order is Ck > B > Sk, while Trevor's is B > St > Ct. Since they have B in common – either both are bored or neither are bored – we can merge their orderings on B to get the partial order shown in Figure 3.4(a).

The AVAF for the combined audience is thus $<\mathcal{A}_e, \mathcal{R}_e, \text{V}_e, \text{val}, \{<\text{B,St}>, <\text{B,Ct}>, <\text{B,Sk}>, <\text{St,Ct}>, <\text{Ck,B}>, <\text{Ck,Sk}>, <\text{Ck,St}>, <\text{Ck,Ct}>\}>$. This gives rise to the AF shown in Figure 3.5.

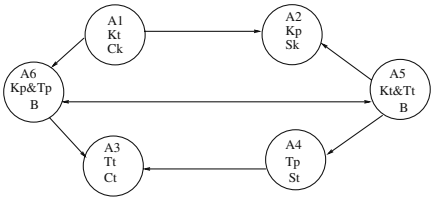

Fig. 3.5 AF for Combined Audience

We can use this VAF to illustrate the algorithm for finding the Preferred Extension given in [6]. First we include the arguments with no attacker: in this case A1. A1 attacks A2 and A6 and so they are excluded. Now A5 has no attacker and so it is included. A5 excludes A4, leaving A3 without an attacker, and so A3 is included to give the preferred extension of the combined audience as {A1,A3,A5}.

This case is straightforward, because the combined audience yields a single preferred extension. The same is true if Trevor preferred S to B, the combined order being shown in Figure 3.4(b). This would cause <A5,A4> to be replaced in $\mathcal{R}$ by <A4,A5>. Now both A1 and A4 are not attacked, and so they defeat the remaining arguments yielding the preferred extension {A1,A4}. This is possible: they simply agree to travel separately by their preferred means.

More complicated is the situation where Katie prefers B to C, so that the merged order is as shown in Figure 3.4(c), and <A6,A1> replaces <A1,A6> in $\mathcal{R}$. Now there is no longer any argument which has no attackers, and the algorithm must be applied twice; first including A5 and then including A6, so that are two preferred extensions, {A1,A3,A5}, and {A2,A4,A6}, both of which are acceptable to them both. Now, since Katie will lean towards the former and Trevor the latter they must find a way to decide between Ck and St. This might depend on who had the strongest opinions, or who is the more altruistic or conciliatory. Alternatively one person might change their preferences: if Katie moved back to her original ordering of C>B>S, Trevor would either have to decide to prefer S to B or to agree to travel by train. This possibility shows how preferences can emerge from the reasoning process: although initially Katie might express a preference for B to C, and Trevor for B to S, when the consequences are realised she may decide that C is actually more important than B, and he may decide S is more important than B.

5 Example Applications

As noted in Section 1, reasoning with values is common to many application domains. In previous work [2, 3, 5, 9] we have shown how the application of abstract argumentation with values can be applied to problems in law, medicine, ethics and

e-democracy, and we will briefly discuss these applications here. We begin by considering legal reasoning with values.

5.1 Law

Reasoning with legal cases has often been viewed as a decision being deduced about a particular case through the application of a set of rules, given the facts of the case, e.g. [25]. However, the facts of cases are not set in stone as they can be open to interpretation from different lawyers. Additionally, the rules used to reach decisions are defeasible by their nature and many are derived from precedent cases, so they too may be open to interpretation. Thus, within the AI and Law literature it has been recognised that when considering arguments in legal cases, the purposes of the law – the values intended to be promoted or upheld through the application of the law – must be represented and accounted for, e.g. [8] [21]. In the literature on legal case-based reasoning the issue was first brought to attention in Berman and Hafner's seminal paper on the topic [8] arguing that legal case-based reasoning needs to recognise teleological as well as factual aspects. This is so since the law is not composed arbitrarily, rather it is constructed to serve social ends, so when conflicts in the application of rules occur in legal cases they can be resolved more effectively by considering the purposes of these rules and their relative applicability to the particular case in question. This enables preferences amongst purposes to be revealed, and then the argument can be presented appropriately to the audience through an appeal to the social values that the argument promotes or defends.

In order to demonstrate how the values of the law can be represented and reasoned about within a case, we have previously presented a reconstruction [3] of a famous case in property law by simulating the opinion and dissent in that case. The case is that of *Pierson vs Post*[1] which concerned a dispute about ownership of a hunted fox. The said fox was being pursued by Post who was hunting with hounds on unoccupied waste-land. Whilst Post was in pursuit of the fox another man, Pierson, came along and intercepted the chase, killing and carrying off the fox. Central to the arguments considered in the case was whether ownership of a wild animal can be attributed through mere pursuit. However, there are numerous other arguments that need to be considered which draw out the emphasis placed on the values considered within the case.

Firstly, the value 'public benefit' was considered as it was argued that fox hunting is of benefit to the public because it assists farmers, so it should be encouraged by giving the sportsman such as Post protection of the law. There are of course counter arguments to this based on the humane treatment of animals. Furthermore, there are arguments concerning consideration of public benefit based on the desire to punish malicious behaviour as allegedly shown by Pierson in intercepting the fox that he could see Post was chasing.

[1] 3 Cai R 1752 Am Dec 264 (Supreme Court of New York, 1805)

Secondly, there were arguments set forth about the need for the law to be clear: in attributing ownership without bodily possession this would encourage a climate of litigation based on similar claims related to pursuit alone.

Thirdly, the value of 'economic benefit' was considered in relation to the protection of property rights where the claimant is engaged in a profitable enterprise.

Given the facts of the case and the values stated above that have been recognised as pertinent to the reasoning in the case, the argument scheme for practical reasoning can be applied to generate the competing arguments about who to decide for in the case. Once generated these arguments can be organised into a VAF and evaluated in the usual manner. In the actual case the court found for Pierson, thus holding that clarity was more important than the values promoted by finding for Post. Preference orderings of values that led to this decision are reflected in our full representation of the case, which can be found in [3]. Explicitly representing the values promoted by the arguments put forward in the case helps to clarify the justifications for the arguments advanced and ground those justifications within the purposes that law is intended to capture and uphold.

5.2 Medicine

A second example scenario that has been considered in terms of value-based argumentation is one concerning a system for reasoning about the medical treatment of a patient [5]. Decision making in this domain often requires consideration of a wide range of options, some of which may conflict, and may also be uncertain. Thus, value-based argumentation can play a role in supporting the decision making process in this domain.

The scenario modelled in [5] illustrates a running example of a patient whose health is threatened by blood clotting. In deciding which particular treatment to administer to the patient there are a number of policies and concerns that affect the decision, and each must be given its due weight. In the computational model of the scenario a number of different perspectives are represented that are given as values of individual agents. The arguments and subsequent conclusions drawn by the individual agents are then adjudicated by a central agent which comes to a decision based on an evaluation of the competing arguments. Concerning the individual agents' values, these represent perspectives such as: the *treatment* of the patient based on general medical policy; the *safety* of the patient concerning knowledge of contraindications of the various drugs; the *efficacy* of the treatments in reference to specific medical knowledge; and, the *cost* of the different treatments available.

Given the above agent perspectives (and others that we do not detail here), the practical reasoning argument scheme can be used to generate arguments about which drug should be used to treat a particular patient. These arguments can be critiqued by agents other than those that generated the recommendation, based on their individual knowledge, through the posing of the appropriate critical questions. This may lead to different agents recommending different treatments, one of which

must be chosen. In order to decide between the competing choices, the arguments justifying each are organised into a VAF and evaluated according to the preference given over the values represented by the individual agents. For example, it may be the case that the treatment agent recommends a particular drug that is known to be highly effective (since no critique from the efficacy agent indicates otherwise) and has no contraindications (according to the safety agent), yet the cost agent has an argument that the drug cannot be used on monetary grounds. The question then is whether treatment is to be preferred to cost (which may be the case if there are no suitable alternatives are identified). Resolution of this issue will be determined by the central adjudicating agent who provides the value ordering to decide upon the winning argument and subsequent treatment recommended, in accordance with the policy of the relevant health authority at the time.

Whilst a key motivation for the example application described above was the representation of the different perspectives within the situation, there are other advantages worthy of note. Firstly, the reasoning involved in medical scenarios is often highly context dependant and relative to specific individuals so there is a high degree of uncertainty. Thus any ordering of preferences must take the specific context into account and the argumentation based approach enables this. Secondly, the argumentation element is effected inside a single agent and the information that it uses is distributed across different information sources, which need not themselves consider every eventuality, and play no part in the evaluation. This simplifies their construction and facilitates their reuse in other applications. Finally, the critiques that are posed against putative solutions are made only as and when they can affect the evaluation status of arguments already advanced. This means that all reasoning undertaken is potentially relevant to the solution.

5.3 Moral Reasoning

The running example that we have presented in this paper concerning travel to a conference is represented in terms of an AATS. We now turn to briefly discussing another example scenario, concerning moral reasoning, that has been modelled in these terms.

The scenario is a particular ethical dilemma discussed by Coleman [12] and Christie [11], amongst others, and it involves two agents, called Hal and Carla, both of whom are diabetic. The situation is that Hal, through no fault of his own, has lost his supply of insulin and urgently needs to take some to stay alive. Hal is aware that Carla has some insulin kept in her house, but Hal does not have permission to enter Carla's house. The question is whether Hal is justified in breaking into Carla's house and taking her insulin in order to save his life. By taking Carla's insulin, Hal may be putting her life in jeopardy, since she will come to need that insulin herself. One possible response is that if Hal has money, he can compensate Carla so that her insulin can be replaced before she needs it. Alternatively if Hal has no money but Carla does, she can replace her insulin herself, since her need is not immediately

life threatening. There is, however, a serious problem if neither have money, since in that case Carla's life is really under threat. Coleman argued that Hal may take the insulin to save his life, but should compensate Carla. Christie's argument against this was that even if Hal had no money and was unable to compensate Carla he would still be justified in taking the insulin by his immediate necessity, since no one should die because of poverty.

In [2] we have represented this scenario in terms of an AATS and considered the arguments that can be generated concerning how the agents could justifiably act. Following our methodology, we take the arguments generated and organise them into a VAF to see the attack relations between them and evaluate them in accordance with the particular value preference orderings. An interesting point that can be taken from this particular example concerns the nuances between different 'levels' of morality that can be drawn out by distinguishing the individual agents *within* the value orderings. For example, *prudential reasoning* takes account of the different agents, with the reasoning agent preferring values relating to itself, whereas *strict moral reasoning* ignores the individual agents and treats the values equally. For example, in the insulin scenario two values are recognised: *life*, which is demoted when Hal or Carla ceases to be alive, and *freedom*, which is demoted when Hal or Carla ceases to have money. Thus, an agent may rank life over freedom, but within this value ordering it may discriminate between agents; for example, the agent may place equal value on its own and another's life, or it may be that it prefers its own life to another's (or vice versa). This leads to distinctions such as *selfish* agents who prefer their own interests above all those of other agents, and *noble* agents whose values are ordered, but within a value the agent prefers another's interests.

In addition to the AATS representation set out in [2], simulations have also been run, which are reported in [10], that confirm the reasoning as set out.

5.4 e-Democracy

The final application area that we discuss is an e-Democracy setting whose focus is more on the support given by value based argumentation within a system to facilitate the collection and analysis of human arguments within political debates.

The application is presented as a discussion forum named Parmenides whose underlying structure is based upon the practical reasoning argument scheme and the latest version of the system is described in [9]. The system is intended as a forum by which the government is able to present policy proposals to the public so users can submit their opinions on the justification presented for the particular policy. The justification for action is structured in the form of the practical reasoning argument scheme, though this imposed structure is hidden from the user. Within a particular topic of debate, a justification upholding a proposed government action is presented to users of the system in the form of the argument scheme. Users are then led in a structured fashion through a series of web pages that pose the appropriate critical questions to determine which parts of the justification the users agree or disagree

with (the circumstances, the action, the consequences or the value). Users are not aware (and have no need to be aware) of the underlying structure for argument representation but it is, nevertheless, imposed on the information they submit. This enables the collection of information which is structured in a clear and unambiguous fashion from a system which does not require users to gain specialist knowledge before being able to use it.

In addition to collecting arguments, Parmenides also has analysis facilities that make use of AFs. All the information that the users submit through the system is stored in a back-end database. This information is then organised into an argumentation framework to show the attacking arguments between the positions expressed. Associated with the arguments in the AF is statistical information concerning a breakdown of support for the arguments, i.e. the number of users agreeing/disagreeing with a particular element of the justification. Thus, arguments can be assessed by considering which ones are the most controversial to the users.

The Parmenides system is intended to overcome some of the problems faced by existing discussion forum formats, such as unstructured blogs and e-petitions. In such systems where there is no structuring of the information, it is undoubtedly very difficult for the policy maker to adequately address each person's concerns since he or she is not aware of users' specific reasons for disagreeing. Furthermore, it may be difficult to recognise agreement and disagreement *between* multiple user replies. In contrast, the structure imposed by Parmenides allows the administrator of the system to see exactly which particular part of the argument is disagreed with by the majority of users, e.g. arguments based on a description of the circumstances, or arguments based on a disagreement about the importance of promoting a particular value. Identifying these different sources of disagreement allows the policy maker to see why his policy is disliked, so he may be able to better respond to the criticisms made, or indeed change the policy. In particular, it can indicate whether the values motivating the policy are shared by the respondents.

Parmenides has been tested on a number of different political debates, including: the UK debate about banning fox hunting[2]; the justification for the 2003 war in Iraq; and, a debate about the proposal to increase the number of speed cameras on UK roads. Work on the Parmenides system is ongoing to further extend its representation facilities, through the use of schemes additional to the practical reasoning scheme, and to further extend the facilities for analysing the arguments through the use of argumentation frameworks.

6 Developments of Value Based Argumentation

In this section we will mention some developments of Value Based Argumentation.

In [13] there is an interesting exploration of the relation between neural networks, in particular neural-symbolic learning systems, and value based argumentation sys-

[2] For this particular debate on the system see:
`http://cgi.csc.liv.ac.uk/~parmenides/foxhunting/`

tems, including an extensive discussion of the insulin example described in the last section. In [15] there is a formal generalisation of VAFs to allow for arguments that promote multiple values, and in which preferences among values can be specified in various ways. In [7] a method is given to determine which audiences can accept a particular set of arguments. Here, however, we will look in detail only at the application of Modgil's extended argumentation frameworks (EAF) [17] to VAFs. For a preliminary exploration of the relation between EAFs and VAFs see [18].

The core idea of EAFs is, like VAFs, to enable a distinction between an argument attacking an argument, and an argument defeating another argument. Whereas VAFs, however, rely on a comparison of properties of the arguments concerned, EAFs achieve this in an entirely abstract manner by allowing arguments to attack not only other arguments, but also attacks. EAFs thus enable arguments to resist an attack for a number of reasons. In VAFs arguments resist attacks solely in virtue of a preference between the values concerned. This enables VAFs to be rewritten as standards AFs, by introducing some auxiliary arguments to articulate the notion of an attack on an attack. These auxiliary arguments represent the status of arguments, value preferences, and arguments representing particular audiences. Suppose we have a VAF with two arguments, A and B which attack one another. A is associated with value V1 and B with Value V2. A will be defeated if B defeats it, and B will be defeated if A defeats it. Defeat is only possible if the attacking argument is not defeated, and if the value of A is not preferred to that of B. Thus the attack on the attack of A on B in an EAF becomes an attack on the argument that A defeats B.

This enables us to represent a VAF as a standard AF, with preferred extensions depending on the choices made regarding value preferences. We can extend the AF to include audiences as well. Suppose Audience X prefers V1 to V2 and Audience Y prefers V2 to V1. This can be shown as in Figure 3.6.

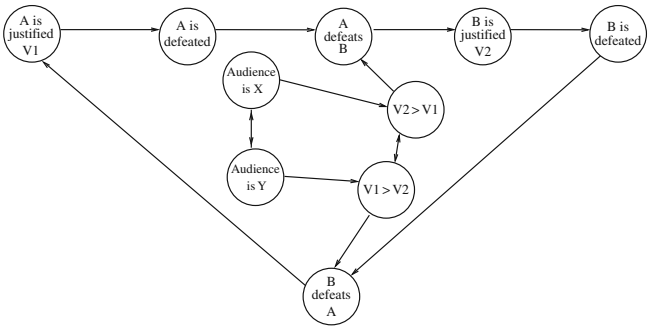

Fig. 3.6 AF representing VAF with audiences

The rewriting of VAFs in this way is shown to be sound and complete with respect to EAFs in [18]. When we rewrite VAFs in this way, subjective acceptance in the VAF is equivalent to credulous acceptance in the rewritten AF, and objective acceptance in the VAF is equivalent to sceptical acceptance in the rewritten AF. This

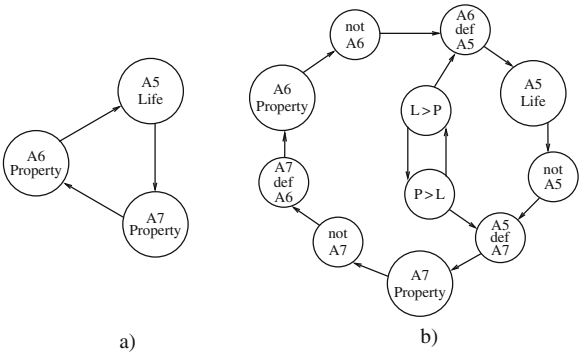

Fig. 3.7 3 cycle and re-write

can be seen by considering the three cycle in the two value case shown in Figure 3.7.

There will be two preferred extensions depending on which preference is chosen: {L> P, A5defA7,not A7,A6,A5} and {P>L,A7,A7defA6,notA6,A5}. Thus A5 is correctly sceptically acceptable in 7b, and objectively acceptable in 7a, and the remaining arguments, other than notA5, are credulously acceptable in 7b, and A6 and A7 are subjectively acceptable in 7a.

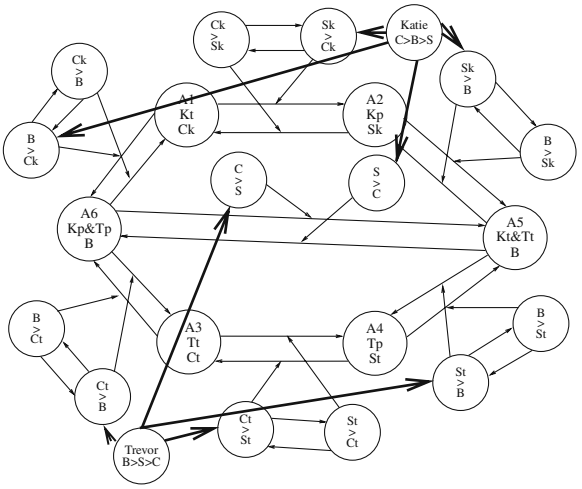

Fig. 3.8 Value based EAF for travel example.

Finally, we apply this to our running example of Trevor and Katie travelling to Paris. The rewritten framework is given in Figure 3.8: note that we have used the EAF style of attacks on attacks rather than the rewrite, for clarity in the diagram.

We have added the audiences Trevor and Katie. Note that, although their preferences differ, these arguments do not conflict, as Trevor and Katie must be allowed to have different preferences: although they are trying to come to a consensus of which arguments to accept, they are free to maintain their own value orders. Trevor's audience attacks the preferences between values in respect of Trevor, and Katie's audience attacks preferences in respect of Katie. Both audiences attack preferences in common.

We evaluate the framework in Figure 3.8 by first removing the arguments attacked by the audiences, and then the attacks attacked by surviving arguments. Reflecting the impact of audiences in this way gives a standard AF, the connected component of which is the same as that shown in Figure 3.5. Now A1 is not attacked, and so the preferred extension will contain the two audiences, the consequent preferences (note that *both* S>C and C>S have been defeated as Trevor and Katie disagree), together with A1, (which is not attacked), A3 and A5 (whose attackers are defeated). Thus, as before, given these preferences both Trevor and Katie choose to travel by train.

7 Summary

Just as deduction is a natural paradigm for justifying beliefs, argumentation is the natural paradigm for explaining and justifying why one course of action is preferred to another, since the notions of defeasibility and individual preference are central to argumentation. We can be coercive about what is the case, but need to be persuasive about what should be the case. But in order to exploit this aspect of argumentation, it is necessary to extend the purely abstract notion of argumentation proposed by Dung to enable individual preferences to explain the choices made in determining which arguments will be accepted by an agent in a particular context. We have discussed such an extension, representing the individual interests and aspirations as values, and individual preferences as orderings on these values.

Using this extension we have shown how different agents can rationally make different choices in accordance with their value orderings, and how in turn these value orderings can emerge from particular situations. In particular we have discussed examples where two agents with different value orderings must agree collectively on what they should do. The range of applications in which reasoning of this sort is required is wide, and we have discussed a number of application areas: law, medicine, politics and moral dilemmas, and an everyday situation. In this chapter we have shown how this important style of reasoning, central to the notion of an autonomous agent, can be captured in a particular form of argumentation framework which, while permitting the expression of individual preferences, retains all the benefits of the clean semantics associated with abstract argumentation frameworks.

References

1. K. Atkinson and T. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10–15):855–874, 2007.
2. K. Atkinson and T. Bench-Capon. Addressing moral problems through practical reasoning. *Journal of Applied Logic*, 6(2):135–151, 2008.
3. K. Atkinson, T. Bench-Capon, and P. McBurney. Arguing about cases as practical reasoning. In *Proc. of the Tenth International Conference on Artificial Intelligence and Law (ICAIL '05)*, pages 35–44, 2005. ACM Press.
4. K. Atkinson, T. Bench-Capon, and P. McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206, 2006.
5. K. Atkinson, T. Bench-Capon, and S. Modgil. Argumentation for decision support. In *Proc. of the Seventeenth DEXA Conference*, LNCS 4080, pages 822–831. Springer, 2006.
6. T. Bench-Capon. Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.
7. T. Bench-Capon, S. Doutre, and P.E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(1):42–71, 2006.
8. D. H. Berman and C. D. Hafner. Representing teleological structure in case-based legal reasoning: the missing link. In *Proc. of the Fourth International Conference on Artificial Intelligence and Law (ICAIL '93)*, pages 50–59, 1993. ACM Press.
9. D. Cartwright and K. Atkinson. Political engagement through tools for argumentation. In P. Besnard, S. Doutre, and A. Hunter, editors, *Proc. of COMMA '08*, pages 116–127, 2008.
10. A. Chorley, T. Bench-Capon, and P. McBurney. Automating argumentation for deliberation in cases of conflict of interest. In P. E. Dunne and T. Bench-Capon, editors, *Proc. of COMMA '06*, pages 279–290. IOS Press, 2006.
11. C. G. Christie. *The Notion of an Ideal Audience in Legal Argument*. Kluwer Academic Publishers, 2000.
12. J. Coleman. *Risks and Wrongs*. Cambridge University Press, 1992.
13. A. S. d'Avila Garcez, D. M. Gabbay, and L. C. Lamb. Value-based argumentation frameworks as neural-symbolic learning systems. *J. of Logic and Computation*, 15(6):1041–1058, 2005.
14. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
15. S. Kaci, L. van der Torre, and E. Weydert. On the acceptability of incompatible arguments. In *Proc. of the Ninth ECSQARU Conference*, pages 247–258, 2007.
16. A. J. P. Kenny. *Practical Reasoning and Rational Appetite*. 1975. Reprinted in [23].
17. S. Modgil. An abstract theory of argumentation that accommodates defeasible reasoning about preferences. In *Proc. of the Ninth ECSQARU Conference*, pages 648–659, 2007.
18. S. Modgil and T. Bench-Capon. Integrating object and meta-level value based argumentation. In P. Besnard, S. Doutre, and A. Hunter, editors, *Proc. of COMMA '08*, pages 240–251, 2008.
19. C. Perelman. *Justice, Law, and Argument*. D. Reidel Publishing Company, Dordrecht, 1980.
20. C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame, IN, USA, 1969.
21. H. Prakken. An exercise in formalising teleological reasoning. In *Proc. of the Thirteenth Annual JURIX Conference*, pages 49–58, 2000. IOS Press.
22. H. Prakken. Combining sceptical epistemic reasoning with credulous practical reasoning. In P. E. Dunne and T. Bench-Capon, editors, *Proc. of COMMA '06*, pages 311–322. IOS Press, 2006.
23. J. Raz, editor. *Practical Reasoning*. Oxford University Press, Oxford, UK, 1978.
24. J. R. Searle. *Rationality in Action*. MIT Press, Cambridge, MA, USA, 2001.
25. M. J. Sergot, F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond, and H. T. Cory. The British Nationality Act as a logic program. *Communications of the ACM*, 29(5):370–386, 1986.
26. D. N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.
27. M. Wooldridge and W. van der Hoek. On obligations and normative ability: Towards a logical analysis of the social contract. *Journal of Applied Logic*, 3:396–420, 2005.