Multi-Criteria Argument Selection In Persuasion Dialogues*

T.L. van der Weide
Universiteit Utrecht
The Netherlands
tweide@cs.uu.nl

F. Dignum
Universiteit Utrecht
The Netherlands
dignum@cs.uu.nl

J.-J. Ch. Meyer
Universiteit Utrecht
The Netherlands
jj@cs.uu.nl

H. Prakken
Universiteit Utrecht
The Netherlands
henry@cs.uu.nl

G. A. W. Vreeswijk
Universiteit Utrecht
The Netherlands
gv@cs.uu.nl

ABSTRACT

The main goal of a persuasion dialogue is to persuade, but agents may have a number of additional goals concerning the dialogue duration, how much and what information is shared or how aggressive the agent is. Several criteria have been proposed in the literature covering different aspects of what may matter to an agent, but it is not clear how to combine these criteria that are often incommensurable and partial. This paper is inspired by multi-attribute decision theory and considers argument selection as decision-making where multiple criteria matter. A meta-level argumentation system is proposed to argue about what argument an agent should select in a given persuasion dialogue. The criteria and sub-criteria that matter to an agent are structured hierarchically into a value tree and meta-level argument schemes are formalized that use a value tree to justify what argument the agent should select. In this way, incommensurable and partial criteria can be combined.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods

Keywords

Argumentation, Persuasion, Decision Making, Multi-Criteria

General Terms

Design

1. INTRODUCTION

In many situations agents benefit from sharing their knowledge with each other. For example, agents may disagree about some fact or about what plan to execute. The disagreements may be resolved by combining their resources and knowledge. In everyday life, dialogues are often used to resolve such disagreement. By giving arguments that justify their positions, participants of a dialogue exchange information that may not have been available to all participants. If the receiving agent updates its beliefs, the disagreement may resolve. Otherwise the agent may give an argument justifying why he still does not agree.

The goal of a persuasion dialogue is that the participants can reach agreement about some subject. Typically there are multiple ways how agreement can be reached in a dialogue because agents can choose what argument they give. However, if the only goal of the agent is to reach agreement, then it does not matter whether he gives all arguments he has or only a few before the agreement is reached. Typically agents have other goals in a persuasion dialogue. For example, one agent may want to minimize the duration of the dialogue, a teacher agent may want to be as comprehensive as possible, a benevolent agent may want to help the other agent as much as possible, a secretive agent may want to minimize sharing private information, or a malicious agent may want to give those arguments that require the most processing time of the audience. To determine the effect of an argument with certainty, an agent must know what the audience knows and how the audience will process his argument. This information is typically not available, but agents may have heuristics to make an educated guess about what effect an argument has.

Several heuristics have been proposed that can be used as criteria in argument selection. For example, the heuristic to select the argument using the agent's most important value is proposed in [2]. In [6], a 'desideratabase' is assumed representing how much an agent is interested in certain formulae and it is proposed to use the desideratabase to determine the resonance of an argument. The heuristic to minimise revealing information is proposed in [9], and in [1] several measures are proposed, such as aggressiveness and coherence, to determine the quality of a persuasion dialogue. These measures could be used as heuristics. In [11] the expected utility of dialogue moves in an adjudication dialogue is determined using probabilities that the adjudicator accepts the argument's premises and that the argument is attacked. In [3], agents are assumed to know to which degree formulae can be used as shared knowledge, which could be used as a heuristic for the likelihood that an agent accepts premises.

These are all valid heuristics for selecting arguments and capture aspects that might be important for a particular agent. If what an agent values in a persuasion dialogue is

*The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

Cite as: Multi-Criteria Argument Selection In Persuasion Dialogues, T.L. van der Weide, F. Dignum, J.-J.Ch. Meyer, H. Prakken and G.A.W. Vreeswijk, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. XXX-XXX.
Copyright © 2011, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

represented by multiple of such heuristics, e.g. he wants to minimize attacks and maximize sharing information, then these heuristics need to be combined to form an agent's preferences over arguments. In the field of decision analysis, multiple approaches have been proposed as to how to decompose what an agent values into criteria and sub-criteria. However, these approaches assume that every aspect is commensurable and that every two arguments can be compared from each criterion. Using multi-attribute utility functions requires that the designer specifies many numerical parameters concerning how the multi-attribute utility function works. However, people typically do not feel comfortable giving such quantitative parameters. Designers are comfortable expressing in a qualitative manner as to what an agent should value in a persuasion dialogue. For example, an agent should be friendly, comprehensive, but not give irrelevant arguments. These criteria of friendliness, comprehensiveness and relevancy are general areas of concern, but are too abstract to operationalize. These criteria could then be further decomposed into sub-criteria until operational heuristic can be assigned. For example, the general area of concern of 'friendliness' could be decomposed into 'minimize aggressiveness' and 'maximize using the arguments of the audience'.

First, Section 2 gives a background on argumentation, persuasion dialogues and decision analysis. After giving a general overview of how arguments will be selected, a meta-level argumentation framework is introduced in Section 3 to argue at the meta-level about what argument at the object level an agent should select. The proposed mechanism is based on [15] and decomposes what matters to an agent into a number of criteria and sub-criteria for which heuristics can be used. Next, argumentation is used to recombine those heuristics to determine what argument an agent should select. The proposed formalism allows combining heuristics that are incommensurable and/or partial. Our approach is illustrated with an example in Section 4. We end the paper with some conclusions and recommendations for future work.

2. BACKGROUND

2.1 Argumentation

We introduce an argumentation system based on [14, 4, 10] to reason defeasibly and in which argument schemes can be expressed. The notion of an argumentation system extends the familiar notion of a proof system by distinguishing between strict and defeasible inference rules. The informal reading of a strict inference rule is that if its antecedent holds, then its conclusion holds without exception. The informal reading of a defeasible inference rule is that if its antecedent holds, then its conclusion tends to hold. A strict rule is an expression of the form $s(x_1, \ldots, x_n)$: $\phi_1, \ldots, \phi_m \rightarrow \phi$ and a defeasible rule is an expression of the form $d(x_1, \ldots, x_n) : \phi_1, \ldots, \phi_m \Rightarrow \phi$, with $m \geq 0$ and $x_1, \ldots, x_n$ all variables in $\phi_1, \ldots, \phi_m, \phi$. We call $\phi_1, \ldots, \phi_m$ the antecedent, ϕ the conclusion, and both $s(x_1, \ldots, x_n)$ and $d(x_1, \ldots, x_n)$ the identifier of a rule.

Definition 1 (Argumentation System). *An argumentation system is a tuple $\mathcal{AS} = \langle \mathcal{L}, \mathcal{R}, ^- \rangle$ with*

- *$\mathcal{L}$ the language of predicate logic,*
- *$\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ such that $\mathcal{R}_s$ is a set of strict and $\mathcal{R}_d$ is a set defeasible inference rules, and*
- *$^-$ a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$.*

For $\phi \in \mathcal{L}$, it is always the case that $\neg\phi \in \overline{\phi}$ and $\phi \in \overline{\neg\phi}$. In this paper, we will assume that $^-$ if $\phi \in \overline{\psi}$, then $\psi \in \overline{\phi}$. In this case, we say that ϕ and ψ are called *contradictory*.

Arguments are defined following [14]. Several functions are defined that return a property of an argument.

Definition 2 (Argument). *Let $\mathcal{AS} = (\mathcal{L}, \mathcal{R}, ^-)$ be an argumentation system. The set* $\mathsf{Args}(\mathcal{AS})$ *denotes the set of all arguments in $\mathcal{AS}$. Arguments are either atomic or compound. An* atomic argument *A is a wff ϕ where*

$\mathsf{conc}(A) = \phi$ $\qquad$ $\mathsf{rules}(A) = \emptyset$
$\mathsf{premises}(A) = \{\phi\}$ $\qquad$ $\mathsf{sub}(A) = \{A\}$
$\mathsf{lastRule}(A) = \mathsf{undefined}$

Let $A_1, \ldots, A_n$ (with $n \geq 0$) be arguments and $r \in \mathcal{R}$ with antecedents $\mathsf{conc}(A_1), \ldots, \mathsf{conc}(A_n)$ and conclusion $\phi \in \mathcal{L}$. A compound argument *A is an argument with*

$\mathsf{conc}(A) = \phi$
$\mathsf{rules}(A) = \{r\} \cup \bigcup_{i=1}^{n} \mathsf{rules}(A_i)$
$\mathsf{premises}(A) = \bigcup_{i=1}^{n} \mathsf{premises}(A_i)$
$\mathsf{sub}(A) = \{A\} \cup \bigcup_{i=1}^{n} \mathsf{sub}(A_i)$
$\mathsf{lastRule}(A) = r$

Arguments can be visualized as inference trees. An argument A is called *strict* if $\mathsf{rules}(A) \cap \mathcal{R}_d = \emptyset$ and *defeasible* otherwise.

Arguments are constructed by applying inference rules to some knowledge base in an argumentation system. A *knowledge base* $\mathcal{K}$ in an argumentation system consists of a set of axioms and a set of ordinary premises. An argument A can be constructed from a knowledge base $\mathcal{K}$ if all A's premises are contained in $\mathcal{K}$. If the premises of argument A only contain axioms, then A is called *firm*. Otherwise, A is called *plausible*.

Typically, agents see arguments as having different strengths. For example, an argument based on imprecise observations is weaker than an argument based on scientific facts. The strength, or conclusive force, of an argument indicates to what degree an agent is convinced of its conclusion. If two arguments have conflicting conclusions and one argument is stronger than the other (or has more conclusive force), then a rational agent should be convinced, *ceteris paribus*, of the conclusion of the stronger argument.

Definition 3 (Argumentation Theory). *An* argumentation theory *is a triple $\langle \mathcal{AS}, \mathcal{K}, \preceq \rangle$, with $\mathcal{AS}$ an argumentation system, $\mathcal{K}$ a knowledge base in $\mathcal{AS}$, and $\preceq$ a binary relation $\preceq$ on $\mathsf{Args}(\mathcal{AS})$ that is reflexive and transitive.*

In [14, 10], argument orderings must satisfy several constraints such as for example that all strict arguments are stronger than defeasible arguments. Although such constraints are rational and useful, we do not want to assume that all agents follow such constraints.

Argumentation Frameworks

Following [10], we distinguish three cases of when an argument attacks another argument. Let $A, B \in \mathsf{Args}(\mathcal{AS})$ be two arguments. Argument A *rebuts* B if A's conclusion contradicts with the conclusion of some defeasible inference rule that was applied in B. Argument A *undermines* B if A's conclusion contradicts with of one of B's non-axiom premises. Argument A *undercuts* B if A concludes an exception to a defeasible inference rule that was applied in B.

Since arguments can differ in strength, not all attacks are successful. The notion of defeat is introduced to denote a successful attack.

DEFINITION 4 (DEFEAT). *Let $\mathcal{AT} = \langle \mathcal{AS}, \mathcal{K}, \preceq \rangle$ be an argumentation theory, $A, B \in \mathsf{Args}(\mathcal{AS})$ be two arguments in $\mathcal{AS}$. A* defeats *B iff (1) A undercuts B, (2) A rebuts B on B' and $A \not\prec B'$, or (3) A undermines B on B' and $A \not\prec B'$.*

Given a set of arguments and the attacks between them, we would like to determine what conclusions are justified. For this we will use argumentation frameworks as defined by Dung [5].

DEFINITION 5 (ARGUMENTATION FRAMEWORK). *An* argumentation framework *(AF) in argumentation theory $\mathcal{AT} = \langle \mathcal{AS}, \mathcal{K}, \preceq \rangle$ is a tuple $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Defeat} \rangle$ with Args arguments in $\mathcal{AS}$ that can be constructed from $\mathcal{K}$ and Defeat a binary relation on Args as defined in Definition 4.*

Given the defeat relations between arguments, different semantics have been proposed for what conclusions are acceptable [5]. An argument is called *justified* (w.r.t. stable semantics) iff it is 'in' in all stable assignments, *overruled* iff it is 'out' in all stable assignments, and *defensible* if it is 'out' in some but not all stable assignments.

Similarly, a formula $\phi \in \mathcal{L}$ is called *justified* iff there is a justified argument that concludes ϕ, *defensible* iff ϕ is not justified but there is a defensible argument concluding ϕ, *overruled* iff ϕ is not justified and not defensible but there is an overruled argument concluding ϕ, and lastly *unknown* iff there is no argument concluding ϕ.

2.2 Persuasion Dialogue

For simplicity, this section describes a persuasion dialogue as in [1], in which only argument games can be played. Let Agents denote the set of all agents.

DEFINITION 6 (DIALOGUE CONTEXT AND MOVES). *A* dialogue context *is a tuple $D = \langle P, \mathcal{AS} \rangle$ with $P \subseteq \mathsf{Agents}$ a set of participants and $\mathcal{AS}$ an argumentation system. A* move *in a dialogue context $\langle P, \mathcal{AS} \rangle$ is a tuple $\langle \alpha, A \rangle$, where $\alpha \in P$ and $A \in \mathsf{Args}(\mathcal{AS})$. If $m = \langle \alpha, A \rangle$, then $\mathsf{loc}(m) = A$, $\mathsf{speaker}(m) = \alpha$ and the* audience *of m is $P \setminus \{\alpha\}$.*

Persuasion dialogues are defined as a sequence of moves in a dialogue context.

DEFINITION 7 (PERSUASION DIALOGUE). *A* persuasion dialogue *is a tuple $\delta = \langle D, (m_0, m_1, \ldots, m_n) \rangle$ consisting of a dialogue context D and a non-empty sequence of moves in D. The subject of δ is $\mathsf{subject}(\delta) = \mathsf{loc}(m_0)$ and the length of δ, denoted $|\delta|$, is $n + 1$.*

There may be a dialogue protocol that governs what moves participants can make when, but in this paper we do not focus on that. A protocol can be seen as a filter on moves that each participant can make in a given persuasion dialogue.

The goal of a persuasion dialogue is to reach agreement about its subject among the participants. However, participants typically have other goals that they want to achieve such as minimizing the duration or maximizing sharing information.

2.3 Decision Analysis

In complex decisions, there are many aspects of what an agent values. Various approaches have been proposed in the decision theory literature how to decompose what an agent values. In [7], what matters to an agent is decomposed into an objective hierarchy. An objective is characterized by a decision context, an object and a direction of preference. For example, in the decision context of persuasion dialogues, some objectives are to maximize persuasiveness and minimize duration. An agent's motivation is decomposed into so-called fundamental objectives, which are then further decomposed into means-objectives until they are operational.

In a similar fashion, [13] decomposes what an agent values into a so-called value tree. A value tree hierarchically relates general areas of concern, intermediate objectives, and specific evaluation criteria defined on measurable attributes. The purpose of a value tree is to explicate and operationalize higher level values.

When using the Analytical Hierarchical Process (AHP) [12], what an agent values is decomposed in a hierarchy of criteria and sub-criteria . Next, the agent makes judgments about the importance of the elements. These judgments are then quantified and used to determine what decision is best.

Decision Analysis And Argument Selection

What argument to select in a persuasion dialogue is a complex decision if there are multiple sides to what an agent values. Consequently, the techniques developed in the field of decision analysis are useful for this purpose. In this paper, we will refer to these techniques as the 'quantitative approaches'.

EXAMPLE 1. *A teacher agent could decompose what he values in a persuasion dialogue into the following general areas of concern: persuasiveness and friendliness. The area of concern 'persuasiveness' could be decomposed into the specific evaluation criteria 'maximize promoting audience's values' (as in [2]) and 'maximize impact' (as in [6]). Friendliness could be decomposed into the specific evaluation criteria 'minimize aggression' and 'maximize loan' (with 'aggression' and 'loan' as in [1]).*

However, when using quantitative approaches for argument selection several problems arise. Firstly, these quantitative approaches require that all criteria and sub-criteria are commensurable. However, designers of agents may be uncomfortable specifying quantitatively how incommensurable criteria should be combined. For example, a teacher agent may want to maximize persuasiveness and friendliness, but it is difficult to specify exactly to what degree persuasiveness is more important than friendliness. People are often comfortable giving qualitative statements concerning criteria. For example, criterion 1 is unimportant, criterion 2 is more important than criterion 3, the less attacks, the higher the persuasiveness (without exactly specifying how much).

Secondly, criteria may depend on information that is not available fully. For example, the persuasiveness of an argument depends on what knowledge the audience has. If only parts of the information required by a criterion is available, then not all arguments can be compared using this criterion. Furthermore, some criteria cannot be used by nature to compare all arguments. For example, if there is a criterion that measures the beauty of an argument, then it may be possible that the beauty of two arguments cannot be compared. Concluding, there is a need to allow criteria that result in a partial ordering of arguments.

Lastly, if an agent uses a quantitative approach, then the explanation of why a certain argument was selected consists of showing the calculation, which is not intuitive or easy to understand. For certain applications agents are required to explain to human users why a certain argument was selected. For example, if agents are used to train communication skills in a serious game, then they should explain to a student why a certain argument should be selected. If agents select arguments based on a quantitative utility function, then the explanation is not very intuitive. Arguments on the other hand are intuitive.

3. ARGUMENT SELECTION

This section proposes an argumentation-based approach inspired by multi-attribute decision theory for argument selection in persuasion dialogues. First, a general description is given of criteria in argument selection. Next, Section 3.2 proposes a meta-level argumentation mechanism that allows arguing about what argument to select if there are incommensurable and/or partial criteria. Finally, several properties of the proposed formalism are discussed. Section 4 then illustrates the proposed formalism with an example that combines several heuristics found in the literature.

3.1 Criteria in Argument Selection

Criteria and heuristics that can be used as criteria require some description of the state. The state should capture information about what has been said in the persuasion dialogue upon until now and information about the audience, e.g. what values the audience finds important. We will generalize from how the state is represented exactly, but we will assume that the set of all states is denoted with S. Furthermore, we will use Args as the set of object-arguments that the persuasion dialogue allows the agent to give.

Definition 8 (Perspective on Arguments). *A* perspective on arguments in Args *is a binary relation* $\leq$ *over* Args *that is reflexive and transitive. The set of all perspectives on a set of arguments is denoted with* $\mathcal{P}$.

A criterion is now defined as a function that maps a state to a perspective on arguments. For example, according to criterion c, argument A is better than argument B in state s_1, whereas A and B are equally good in state s_2. According to another criterion, A and B may be equally good in both s_1 and s_2.

Definition 9 (Criterion). *A* criterion *is a function* $\mathsf{c} : \mathsf{S} \to \mathcal{P}$.

- *A criterion function c is called* complete *if $c(s)$ is a complete ordering for all $s \in \mathsf{S}$. Otherwise, c is called* partial.
- *A criterion function is called* total *if c is complete and for all $A, B \in \mathsf{Args}$ is is true that either $(A, B) \in c(s)$ and $(B, A) \notin c(s)$ or it is true that $(B, A) \in c(s)$ and $(A, B) \notin c(s)$.*

For example, let c be a criterion that orders arguments by the number of arguments in the dialogue that they attack. Because for every argument and dialogue it can be determined how much arguments are attacked but it is possible that two arguments attack the same number of arguments, c is complete but not total.

Note that criteria that map states to real numbers can easily be transformed into criteria that map states to an argument ordering.

3.2 Arguing about Arguments

Meta-level argumentation is required to argue about what argument should be selected. In [16], first-order hierarchical meta-languages are used for argumentation and [8] reasons about object-level arguments on a meta-level. To use the structure of arguments as described in Section 2.1, a meta-argumentation system is proposed on the basis of an (object) argumentation system. The meta argumentation system can refer to formulae, inference rules and arguments in the object argument system and therefore these things are defined as terms in the meta-language.

Definition 10 (Meta-Argumentation System). *A* Meta-Argumentation System *on the basis of argumentation system $\mathcal{AS} = (\mathcal{L}, \mathcal{R}, ^-)$ is an argumentation system $\mathcal{AS}' = (\mathcal{L}', \mathcal{R}', ^-)$ such that*

- *each formula ϕ in $\mathcal{L}$ is a term in $\mathcal{L}'$*
- *each inference rule $r \in \mathcal{R}$ is a term in $\mathcal{L}'$,*
- *each argument $A \in \mathsf{Args}(\mathcal{AS})$ is a term in $\mathcal{L}'$, and*
- *the functions defined on arguments (see Definition 2) are function symbols in $\mathcal{L}'$.*

A meta-argumentation system is a special class of argumentation systems. Therefore, a meta-argumentation system can be used in an argumentation theory and argumentation framework as described in Section 2.1. To distinguish meta-arguments from object-arguments, meta-arguments are denoted with monospace font, e.g. `A'`, `B'` and `C'`.

Perspectives

The meta-language will now be instantiated with several relations based on [15] to be able to argue about what object-argument should be selected. Each perspective in $\mathcal{P}$ is a term in the meta-language $\mathcal{L}'$.

Definition 11 (Perspective). *For each perspective $p \in \mathcal{P}$, a binary predicate $\leq_p$ over* Args *is introduced in $\mathcal{L}'$ that is reflexive and transitive.*

If $(A, B) \in \leq_p$, then we write $A \leq_p B$ and say that argument B is weakly preferred to argument A from perspective p. Strict preference $<_p$ and equal preference $\equiv_p$ are defined in the standard way for each perspective. Furthermore, the contrariness function is such that $A <_p B$ is contradictory with both $A \equiv_p B$ and $B <_p A$ for all perspectives and object-arguments.

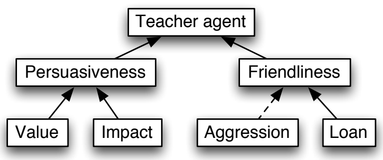
Each criterion is now associated with a perspective. Namely, if c_p is a criterion and s is the current state, then $c_p(s)$ is referred to as perspective p.

Influence

Influence is a binary relation between perspectives that is transitive and irreflexive. The binary predicates $\uparrow$ and $\downarrow$ over perspectives are introduced in $\mathcal{L}'$. If $(p, q) \in \uparrow$, then we write $p \uparrow q$ and say that perspective p positively influences perspective q. Similarly, if $(p, q) \in \downarrow$, then we write $p \downarrow q$ and say that perspective p negatively influences perspective q. Intuitively, 'p positively influences q' can be read as 'the better from p, the better from q, and 'p negatively influences q' can be read as 'the better from p, the worse from q'.

Example 2 (Value Tree). *Consider a teacher agent with a value tree as in Example 1. The agent's preferences, general areas of concern and the specific evaluation criteria used can all be represented as perspectives with influences between as visualized in Figure 1 (where a node represents a perspective, a normal arrow positive influence, and a dotted arrow negative influence).*

Figure 1: Influence graph of a teacher's value tree



Influence between perspectives is used to propagate value from the influencing perspective to the influenced perspective. The argument scheme *perspective p positively influences perspective q, argument B is strictly preferred to argument A from p, therefore, presumably B is strictly preferred to A from q* propagates value using positive influence between perspectives.

Similarly, the argument scheme *p negatively influences q, B is strictly preferred to A from p, therefore, presumably B is strictly preferred to A from q* propagates value using negative influence. Finally, the argument scheme *p either positively or negatively influences perspective q and arguments A and B are equally preferred from a perspective p, therefore presumably A and B are equally preferred from q* propagates value in the case of equal preference. These three argument schemes are formalized by adding the following three defeasible inference rules to $\mathcal{R}'_d$.

$$d_{\uparrow}(p,q,A,B): \quad p{\uparrow}q,\ A <_p B\ \Rightarrow A <_q B$$
$$d_{\downarrow}(p,q,A,B): \quad p{\downarrow}q,\ A <_p B\ \Rightarrow B <_q A$$
$$d_{\equiv}(p,q,A,B): \quad p{\downarrow}q \vee p{\uparrow}q,\ A \equiv_p B\ \Rightarrow A \equiv_q B$$

Note that these inference rules are defeasible, which means that they only create a presumption for their conclusion. Consequently, an argument that applies a defeasible inference rule can be attacked on the conclusion of the defeasible inference rule and the application of the defeasible inference rule can be undercut when there is an exceptional situation.

Relative Importance of Perspectives

Not all perspectives that influence a perspective p need to be equally important for p. For example, for the perspective of friendliness it may be more important to minimize aggressiveness than to maximize lending the audience's arguments. To represent importance of perspectives relative to the perspective that they influence, the following is introduced.

Definition 12 (Relative Importance of Perspectives). Relative importance of perspectives *is a ternary relation* $\trianglelefteq \subseteq \mathcal{P}^3$ *such that:*

- *if* $(p_1, p_2, q) \in \trianglelefteq$ *and* $(p_2, p_3, q) \in \trianglelefteq$, *then* $(p_1, p_3, q) \in \trianglelefteq)$ *for all* $p_1, p_2, p_3, q \in \mathcal{P}$,
- $(p, p, q) \in \trianglelefteq$ *for all* $p, q \in \mathcal{P}$,
- *if p does not influence r and q does influence r, then* $(p, q, r) \in \trianglelefteq$ *and* $(q, p, r) \notin \trianglelefteq$.

If $(p, q, r) \in \trianglelefteq$, then we write $p \trianglelefteq_r q$ and say that perspective q is weakly more important for perspective r than perspective p. The relative importance of perspective will now be used to determine the strength of meta-arguments. The strength of arguments is used in an argumentation theory to determine what attacks are successful, i.e. what arguments defeat other arguments, as explained in Section 2.1.

Definition 13 (Strength of Meta-Arguments). *Let* $\langle \mathcal{AS}', \mathcal{K}', \preceq' \rangle$ *be an argumentation theory with* $\mathcal{AS}'$ *a meta argumentation system on the basis of* $\mathcal{AS}$. *For all* $\mathtt{A'}, \mathtt{B'} \in \mathsf{Args}(\mathcal{AS}')$: $\mathtt{A'} \preceq' \mathtt{B'}$ *if*

- $\mathsf{lastRule}(\mathtt{A'}) = d_X(p, r, A, B)$,
- $\mathsf{lastRule}(\mathtt{B'}) = d_Y(q, r, A, B)$, *and*
- $p \leq_r q$.

with $X, Y \in \{\uparrow, \downarrow, \equiv\}$, $p, q, r \in \mathcal{P}$ *and* $A, B \in \mathsf{Args}(AS)$.

Note that if two meta-arguments infer value to a different perspective, then their strength is incomparable. For example, $\mathtt{A'}$ infers value from p to p' and $\mathtt{B'}$ infers value from q to q' such that $p' \neq q'$, then the strength of $\mathtt{A'}$ and $\mathtt{B'}$ is incomparable. Since such meta-arguments have conclusions concerning different perspectives, they never conflict and thus their relative strength is never required to determine defeat.

Using the framework

The argumentation mechanism proposed in this section takes as input a state, a number of criteria, an influence graph describing how these criteria influence an agent's perspective, and the relative importances of perspectives. Given this input, the output is an argument ordering from each perspective that can be justified. In this sense, our approach is a criterion itself that requires that the state contains a set of criteria, influences and importances.

Suppose the current state in the dialogue is $s \in \mathsf{S}$ and that the agent wants to select the best argument in the set Args of object-arguments in argumentation system $\mathcal{AS}$. Furthermore, we have the set perspectives $\mathcal{P}$ with a special perspective α denoting the perspective of the agent. The positive influence relation $\uparrow$ and negative influence relation $\downarrow$ between $\mathcal{P}$ are used to capture α's value tree and the relative importance between $\mathcal{P}$ is captured by $\trianglelefteq$.

Let $\mathcal{AS}' = \langle \mathcal{L}', \mathcal{R}', ^- \rangle$ be a meta-argumentation system based on $\mathcal{AS}$ such that $\mathcal{L}'$ contains the influence predicates between perspectives and the binary relations $\leq_p$, $<_p$ and $\equiv_p$ for each perspective $p \in \mathcal{P}$ and such that $\mathcal{R}'$ contains the defeasible inference rules as introduced in this section. Furthermore, let $\mathcal{K}'$ be a knowledge-base in $\mathcal{AS}'$ such that

- if perspective p positively / negatively influences perspective q, then $p{\uparrow}q \in \mathcal{K}'$ and $p{\downarrow}q \in \mathcal{K}'$ respectively, and
- if c is a criterion associated to perspective p, then $A \leq_p B \in \mathcal{K}'$ if $(A, B) \in c(s)$ for all A and B object-arguments.

Given how the operational perspectives influence a perspective p, meta-arguments are constructed for how object-arguments compare from p. Because the influencing perspectives may disagree about how arguments should compare from p, some of these arguments may attack each other. Definition 13 defines $\leq'$, which is used to construct argumentation theory $\mathcal{AT}' = \langle \mathcal{AS}', \mathcal{K}', \leq' \rangle$. From $\mathcal{AT}'$ the argumentation framework $\mathcal{AF}' = \langle \mathsf{Args}', \mathsf{Defeat}' \rangle$ is constructed with Args' all arguments in $\mathsf{Args}(\mathcal{AS}')$ that can be constructed from $\mathcal{K}'$ and Defeat' the defeat relations between arguments in Args' as defined by Definition 4. The justified conclusions of $\mathcal{AF}'$ then induce an ordering over object-arguments from perspective p. Consequently, this argumentation mechanism is a criterion: if $A \leq_p B$ is a justified conclusion, then $(A, B) \in p$, if $A <_p B$ is a justified conclusion, then $(A, B) \in p$ and $(B, A) \notin p$, and if $A \equiv_p B$ is a justified conclusion, then $(A, B) \in p$ and $(B, A) \in p$.

3.3 Properties

In the previous subsection, an argumentation-based approach was proposed to argue about what argument an agent should prefer. Agents are prescribed to select the argument that they prefer maximally. To determine what argument is maximally preferred from the perspective of the agent, it is useful if all arguments are comparable from the perspective of the agent.

The following proposition concerns whether an argument can be constructed comparing two arguments from a perspective.

Proposition 1. *Let A and B be two object arguments. If there is a perspective p from which A and B can be compared and p influences perspective q, then a meta-argument can be constructed concerning how A and B compare from perspective q.*

Proof. Because p positively or negatively influences q, $p \uparrow q$ or $p \downarrow q$ is true. Furthermore, because A and B can be compared from p, either $A <_p B$, $A \equiv_p B$ or $B <_p A$ is true. If $A \equiv_p B$ is true, then $d_\equiv$ can be applied concluding that $A \equiv_q B$. Otherwise, if $p \uparrow q$ is true, then the defeasible inference rule $d_\uparrow$ can be applied and if $p \downarrow q$ is true, then $d_\downarrow$ can be applied. Both inference rules conclude how A and B compare from perspective q. □

Consequently, if there is a complete criterion p that influences perspective q, then for each combination of object-arguments a meta-argument can be constructed concluding how they compare from q. This does however not mean that all these meta-arguments are justified or even defensible. They could be attacked by other arguments.

Similarly, if A and B are incomparable from every perspective influencing perspective p, then no meta-arguments can be constructed concluding how A and B compare from perspective p. Consequently, the justified conclusions of the corresponding argumentation framework do not induce a complete ordering of arguments from perspective p.

We will now investigate possible attack relations between meta-arguments. Recall from Section 3.2 that because of the contrariness function, a meta-argument concluding $A <_p B$ attacks a meta-argument concluding $B <_p A$.

Proposition 2. *Let $\mathcal{AT}' = \langle \mathcal{AS}', \mathcal{K}', \leq' \rangle$ be an argumentation theory with $\mathcal{AS}'$ a meta-argumentation system and $M \subseteq \mathsf{Args}(\mathcal{AS}')$ a set of meta-arguments such that each argument can be constructed from $\mathcal{K}'$ and concludes how object-arguments A and B compare from perspective p. For all `A'`, `B'` $\in M$, if argument `A'` attacks `B'`, then `A'` either rebuts `B'` on conc(`B'`) or on the conclusion of a non-atomic sub-argument of `B'`.*

Proof. If p is the perspective of a criterion, then the meta-arguments are atomic. Because by definition there can be no conflicts in the perspective of a criterion, it is not possible that `A'` and `B'` attack each other. Consequently, the meta-arguments cannot be undermined. Otherwise, p is not the perspective of a criterion, but is on a higher level in the influence graph. In that case, the meta-arguments have applied the defeasible inference rules $d_\uparrow$, $d_\downarrow$ or $d_\equiv$. Because no undercutters have been introduced for these defeasible inference rules, it is not possible to undercut such a meta-argument. Finally, it is possible to rebut the conclusion of `B'` because there may be multiple perspective from which value can be inferred to p. The same reason holds for sub-arguments of `B'` that are not atomic. □

Now that we understand possible attack relations between meta-arguments better, we want to investigate the conclusions. The relative importance of perspectives is used to determine the argument strength. Argument strength is used to determine what attacks are successful (i.e. defeats) and what attacks are unsuccessful. In other words, the set of defeats is a subset of or equal to the set of attacks between arguments.

Proposition 3. *Let $\mathcal{AF}' = \langle \mathsf{Args}', \mathsf{Defeat}' \rangle$ be an argumentation framework of a meta-argumentation system. If Args' contains one or more meta-arguments that conclude how object-arguments A and B compare from perspective p, then there is either a defensible or justified conclusion concerning how A and B compare from p.*

Proof. Because of Proposition 2, if the meta-arguments attack each other, then they either rebut a conclusion or rebut a sub-argument's conclusion. In both cases, the attacks are bi-directional and originate from that value is inferred from different perspectives. If the different perspectives are equally important for p, then the corresponding meta-arguments are equally strong resulting in that all arguments are defensible. On the other hand, if some perspective p' is more important than another for p, then the argument using p' is stronger than the other and consequently, it defeats the other and becomes a justified argument. □

We will now investigate a particular instantiation of influences and importances that results in a complete perspective on arguments that is justified.

Proposition 4. *Let the perspectives in set P all influence perspective q. If there is a complete perspective of a criterion $p \in P$ such that for all $p' \in P$ it is true that if $p \neq p'$ then $p' \lhd_q p$, then it is always the case that the corresponding argumentation framework has a justified conclusion concerning how A and B compare from q.*

Proof. Because p is a perspective of a criterion, the meta-arguments inferring value from p to q do not rebut. Furthermore, because p is complete, a meta-argument will be constructed for every two object-arguments. If meta-arguments are constructed from other perspectives that influence q that conflict with how value is inferred from perspective p, then the p-based meta-argument defeats the other argument because p is more important for q than any other influencing perspective. □

4. EXAMPLE

This section illustrates the approach of the previous section by combining several criteria found in the literature. Suppose that an agent is in a certain point of a persuasion dialogue where he can choose from only two arguments: $\mathsf{Args} = \{A, B\}$. The state s captures the persuasion dialogue until now and some information about what values the audience finds important.

In [1], the criterion of 'aggressiveness' is used which is based on the number of arguments uttered by the audience that an argument attacks. Because we use structured argumentation, three different kinds of attacks have been distinguished in Section 2.1, so three different attack criteria can be distinguished: the number of undermining attacks, rebutting attacks and undercutting attacks denoted by criterion c_{umine}, c_{rbut} and c_{ucut} respectively (with $\mathsf{umine}, \mathsf{rbut}, \mathsf{ucut}$ the associated perspectives). Note that the these criteria are complete because for every argument it can be determined how many arguments of the audience it attacks.

Also in [1] the criterion of 'loan' is used which is based on counting how many formulae in an argument have already been uttered by the audience. The criterion of loan is denoted with c_{loan}. Let the set $X \subseteq \mathcal{L}$ be the set of formulae such that for all $\phi \in X$ there is an argument A the audience has uttered with $\phi \in \mathsf{premises}(A)$. Then $(A, B) \in c_{\mathsf{loan}}$ iff $\mathsf{premises}(B) \cap X$ is as much or more than $\mathsf{premises}(A) \cap X$. Note that c_{loan} is complete because for every argument it can be determined how many premises it lends.

In [2], the criterion is proposed to select the argument promoting the value that the audience finds most important. The criterion of using the argument promoting the most important value is denoted as c_{val}. Given that the state captures the value ordering of the audience at least partially, $(A, B) \in c_{\mathsf{val}}$ if and only if the audience finds the value promoted by argument B weakly more important than the value promoted by argument A. Note that c_{val} is not necessarily complete because the agent may not know the audience's complete ordering over values.

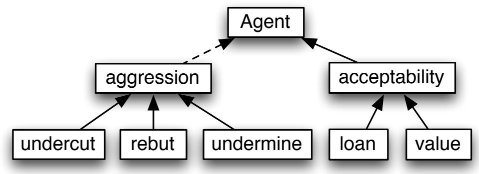
Decomposing What Matters To An Agent

Because all criteria result in a perspective on arguments in Args, the set $\mathcal{P}$ contains a perspective for each criterion. Also, $\mathcal{P}$ contains the perspective α denoting the perspective of the agent who is deciding what argument to select. Because the agent has decomposed what matters into two general areas of concern 'aggressiveness' and 'acceptability', two perspectives are added to denote those general areas of concern. Consequently, the set of perspectives is the following: $\mathcal{P} = \{\alpha, \mathsf{aggr}, \mathsf{accpt}, \mathsf{ucut}, \mathsf{rbut}, \mathsf{umine}, \mathsf{loan}, \mathsf{val}\}$.

Undercutting, rebutting and undermining positively affect aggressiveness, i.e. the more arguments of the audience an argument A undermines, the more aggressive A is. Therefore, $\uparrow$ contains $(\mathsf{ucut}, \mathsf{aggr})$, $(\mathsf{umine}, \mathsf{aggr})$, and $(\mathsf{rbut}, \mathsf{aggr})$. The more premises an argument A lends from the audience, the more likely the audience will accept A. Furthermore, the more important the audience finds the value promoted by argument A, the more likely the audience accepts A. Therefore, $(\mathsf{loan}, \mathsf{accpt})$ and $(\mathsf{val}, \mathsf{accpt})$ are elements of $\uparrow$.

The agent, denoted with perspective α, wants to minimize aggression and maximize acceptability of the arguments that he gives. Therefore, $\downarrow = \{(\mathsf{aggr}, \alpha)\}$ and $(\mathsf{accpt}, \alpha) \in \uparrow$. These influences are visualized in the influence graph in Figure 2 (where a node represents a perspective, a normal directed edge denotes positive influence and a dotted directed edge denotes negative influence). Note that other agents may care about different criteria in different ways, e.g. an aggressive agent may be positively influenced by aggressiveness and may not care about acceptability at all.

Figure 2: Influence graph for the agent



Relative Importances Of Influences

Undermining an argument of the audience is more important for aggressiveness than undercutting or rebutting an argument of the audience. Namely, undermining an argument means that its premises are attacked, whereas undercutting an argument means that there is an exceptional situation in which some defeasible inference rule cannot be applied. Because the premises of an audience's arguments are likely in the audience's knowledge base, undermining is more important for aggressiveness than rebutting and undercutting. Consequently, $\mathsf{ucut} \lhd_{\mathsf{aggr}} \mathsf{rbut} \lhd_{\mathsf{aggr}} \mathsf{umine}$.

Because the designer did not want to specify whether aggressiveness or acceptability is more important for the agent, these two perspectives are incomparable with respect to importance for the agent.

Constructing Meta-Arguments

As described in the previous section, the meta-argumentation system $\mathcal{AS}' = \langle \mathcal{L}', \mathcal{R}', ^- \rangle$ is initialized on the basis of the object-argumentation system $\mathcal{AS}$.

A knowledge base $\mathcal{K}'$ in $\mathcal{AS}'$ is then initialized with $p{\uparrow}q \in \mathcal{K}'$ iff p positively influences q and $p{\downarrow}q \in \mathcal{K}'$ iff p negatively influences q. Furthermore, if c is criterion and p the perspective associated to c, then $A <_p B \in \mathcal{K}'$ iff $(A, B) \in c(s)$ and $(B, A) \notin c(s)$ and $A \equiv_p B \in \mathcal{K}'$ iff $(A, B), (B, A) \in c(s)$.

Suppose that object-argument B undercuts an argument of the audience and object-argument A does not, but A undermines an argument of the audience while B does not. In that case, $A <_{\mathsf{uc}} B$ and $B <_{\mathsf{umine}} A \in \mathcal{K}'$ are in the meta knowledge base $\mathcal{K}'$. Using this information, the following two meta-arguments can be constructed.

$$\mathtt{A'} = \cfrac{\mathsf{aggr}{\downarrow}\alpha \qquad \cfrac{\mathsf{uc}{\uparrow}\mathsf{agr} \quad A <_{\mathsf{uc}} B}{A <_{\mathsf{agr}} B}\, d_{\uparrow}}{B <_{\alpha} A}\, d_{\downarrow}$$

$$\mathtt{B'} = \cfrac{\mathsf{aggr}{\downarrow}\alpha \qquad \cfrac{\mathsf{umine}{\uparrow}\mathsf{agr} \quad B <_{\mathsf{umine}} A}{B <_{\mathsf{agr}} A}\, d_{\uparrow}}{A <_{\alpha} B}\, d_{\downarrow}$$

Further suppose that A and B both do not loan any premises of the audience and that it is not known which of the values promoted by A and B the audience finds important. In that case, $A \equiv_{\mathsf{loan}} B$ is in $\mathcal{K}'$ and A and B are incomparable from the perspective val.

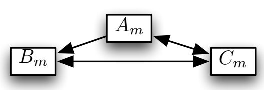
$$\texttt{C'} = \cfrac{\mathsf{accpt}\uparrow\alpha \qquad \cfrac{\mathsf{loan}\uparrow\mathsf{accpt} \qquad A \equiv_{\mathsf{loan}} B}{A \equiv_{\mathsf{accpt}} B}\, d_{\equiv}}{A \equiv_{\alpha} B}\, d_{\equiv}$$

Because A and B are incomparable from val, no argument can be constructed using how A and B compare from val.

Determining The Justified Conclusions

Arguments `A'` and `B'` attack each other, but because undermining is more important for aggressiveness than undercutting, i.e. $\mathsf{ucut} <_{\mathsf{aggr}} \mathsf{umine}$ is true, `A'` defeats `B'`. Also `C'` and `A'` attack each other and so do `A'` and `B'`. Because neither acceptability nor aggressiveness is more important for α, the strengths of `A'` and `C'` are incomparable. Figure 3 visualizes the corresponding argumentation framework $\mathcal{AF} = \langle\{\texttt{A'},\texttt{B'},\texttt{C'}\},\{(\texttt{A'},\texttt{B'}),(\texttt{A'},\texttt{C'}),(\texttt{C'},\texttt{A'}),(\texttt{C'},\texttt{B'})\}\rangle$ that is constructed from argumentation theory $\langle\mathcal{AS}',\mathcal{K}',\preceq'\rangle$ following Definition 5. Both `A'` and `C'` are defensible arguments and `B'` is an overruled argument. Consequently, $B <_\alpha A$ and $A \equiv_\alpha B$ are defensible conclusions. Therefore the agent should conclude that he should weakly prefer B to A.

Figure 3: Defeats between the arguments visualized.



5. CONCLUSION

In this paper we have proposed a formalism to argue on a meta-level about what argument an agent should select in a given persuasion dialogue. Inspired by techniques in decision analysis, what matters to an agent in a persuasion dialogue is decomposed into criteria and sub-criteria. Several argument schemes are formalized to combine criteria that are incommensurable or partial.

The advantages of our approach are that it is (1) easier for the designer than a purely quantitative approach, (2) it allows using criteria that are partial and/or incommensurable, and (3) the agent can explain why a certain argument is selected in a more intuitive way. The main disadvantage of our approach is that it does not always result in all arguments being comparable, which is inconvenient for deciding what argument to select.

Whether to take a purely quantitative approach or the approach proposed in this paper depends on the application. If it is possible to describe how an agent should select arguments quantitatively, then this should be done because it results in a complete ordering over arguments and requires less computation. If on the other hand it is impossible to take a quantitative approach or it requires choosing many parameters in an arbitrary manner, then the approach of this paper may offer the best of both worlds.

For future work we would like to investigate the properties of the persuasion dialogue when agents use different criteria. For example, what effect does aggressiveness have on the duration of persuasion dialogues? This papers assumes that the influences and importances are given, however, it also possible that an agent determines these dynamically using the state. For example, if the audience has attacked one of the agent's arguments, then minimizing aggressiveness does not matter anymore to the agent. Finally, it would be interesting to explore more refined influence and importance relations.

6. REFERENCES

[1] L. Amgoud and F. de Saint Cyr. Measures for persuasion dialogs: A preliminary investigation. In *Computational Models of Argument. Proceedings of COMMA 2008*, pages 13–24, 2008.

[2] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[3] E. Black and A. Hunter. A Relevance-theoretic Framework for Constructing and Deconstructing Enthymemes. *Journal of Logic and Computation*, 2009.

[4] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[5] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[6] A. Hunter. Towards higher impact argumentation. *Proc. of the 19th American National Conf. on Artificial Intelligence (AAAI 2004), MIT Press*, pages 275–280, 2004.

[7] R. Keeney and H. Raiffa. *Decisions with Multiple Objectives.* Wiley, New York, 1976.

[8] S. Modgil and T. J. M. Bench-Capon. Metalevel argumentation. *Journal of Logic and Computation*, 2010.

[9] N. Oren, T. Norman, and A. Preece. Information based argumentation heuristics. *Argumentation in Multi-Agent Systems*, pages 161–174, 2007.

[10] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[11] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: A game-theoretical investigation. In *Computational Models of Argument. Proceedings of COMMA 2008*, pages 324–335, 2008.

[12] T. Saaty. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1):83–98, 2008.

[13] D. Von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research.* Cambridge University Press, 1986.

[14] G. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90(1-2):225–279, 1997.

[15] T. v. d. Weide, F. Dignum, J.-J. C. Meyer, H. Prakken, and G. A. W. Vreeswijk. Arguing about preferences and decisions. In *Proc. of the 7th Int. Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2010)*, 2010.

[16] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In *Argumentation in Multi-Agent Systems 2005*, volume 4049/2006 of *LNCS*, pages 42–56, 2005.