CMNA VI - Computational Models of Natural Argument

The CMNA workshop series has been attracting steadily more submissions and steadily more attendees since its inception in 2001: in this 6th edition we are pleased to have 15 papers, which will be presented, for the fisrt time, in an extended two-day event. As in the past editions, the papers all contribute to the interdisciplinary atmosphere that has characterised the workshop throughout these years: scholars from argumentation theory, philosophy, artificial intelligence, cognitive science, linguistics, computer science promise to offer a rich and intense programme of presentations, together with stimulating and productive discussions.

With this in mind, we would like to take this opportunity to formally announce the 7th workshop of the series, which will be held with IJCAI at Hyderabad, in January 2007: we look forward to an exciting edition.

In the meantime, it is our pleasure to thank all of the authors at this year's meeting, to thank our dedicated Programme Committee, whose diligent work and constant support are key factors for the success of the series, and to welcome everyone who will join us in Riva del Garda for CMNA 2006.

Floriana Grasso
Rodger Kibble
Chris Reed

August 2006

Organization

Program Chairs

Floriana Grasso, University of Liverpool, UK
Rodger Kibble, Goldsmiths College, University of London, UK
Chris Reed, University of Dundee, UK

Program Committee

Leila Amgoud, IRIT, France
Ulises Cortes, UPC, Spain
Fiorella de Rosis, University of Bari, Italy
Tom Gordon, Fraunhofer FOKUS, Berlin, Germany
Nancy Green, University of North Carolina Greensboro, US
Helmut Horacek, University of the Saarland, Saarbrücken, Germany
Anthony Hunter, University College London, UK
Peter McBurney, University of Liverpool, UK
David Moore, Leeds Metropolitan University, UK
Ephraim Nissan, Goldsmiths College, University of London, UK
Henry Prakken, University of Utrecht and University of Groningen, NL
Oliviero Stock, ITC-IRST, Italy
Doug Walton, University of Winnipeg, Canada

Table of Contents

Towards a Formal Argumentation-based Model for Procedural Texts

Leila Amgoud and **Farida Aouladomar** and **Patrick Saint-Dizier** [1]

Abstract.

In this paper, we first present an analysis of the facets of natural argumentation in procedural texts. Next, we extend the formal model proposed in [2] to accommodate these facets. Finally, we outline the main properties of the model.

1 Introduction

Procedural texts are specific forms of discourse, satisfying constraints of economy of means, accuracy, etc. They are in general based on a specific discursive logic, made up of presuppositions, causes and consequences, goals, inductions, warnings, anaphoric networks, etc., and more psychological elements (e.g. *to stimulate a user*). The goal is to optimize a logical sequencing of instructions and to make the user feel safe and confident with respect to the goal(s) he wants to achieve. This type of discourse contains a number of facets, which all are associated in a certain way to argumentation: procedural discourse is indeed informative, narrative, explicative, descriptive, injunctive and sometimes figurative. In fact, argumentation does provide a motivation and an internal coherence to procedural texts: procedural discourse is basically interactive: it communicates, teaches, justifies, explains, warns, forbids, stimulates, evaluates.

Procedural texts consist of a structure goal-subgoals or task-subtasks designed with some accuracy in order to reach an objective (e.g. assemble a computer). In our perspective, procedural texts range from apparently simple cooking receipes to large maintenance manuals (whose paper versions are measured in tons e.g. for aircraft maintenance). They also include documents as diverse as teaching texts, medical notices, social behavior recommendations, directions for use, do-it-yourself and assembly notices, itinerary guides, advice texts, savoir-faire guides, etc. [1].

More precisely, a procedural text is a structure composed of a main *goal* or task, which is decomposed recursively into subtasks. Leaves are elementary tasks, also called instructions. The *tree structure* reflects in general the temporal structure of the system in terms of temporal precedence. To make it more precise, we present below a model that allows for different temporal combinations of tasks (precedence, overlap, etc.). Finally, this tree structure also reveals the modularity of procedures, via the task-subtask decomposition. Therefore, constraints and arguments stated within a subtask, only range over the elements of that subtask.

The backbone of a procedural text is clearly the task-subtasks structure. In most types of procedural texts, procedural discourse has in fact two deeply intertwinned dimensions: an *explicative component*, constructed around rational and objective elements (the task-subtask structure), and a *seduction component* whose goal is (1) to motivate the user by outlining the importance of certain tasks and the necessity to fully realize them, by giving various forms of advices, (2) to make the user understand that the procedure proposed is a safe and efficient way to reach the goal, (3) to prevent the user from making errors of various types via warnings. This seduction component closely associated with the rational elements, forms, in particular, the argumentative structure of the procedural text.

The diversity of procedural texts, their objectives and the way they are written is the source of a large variety of natural arguments. This study is based on a extensive corpus study, within a language production perspective. This approach allows us to integrate logical, linguistic (e.g. [6, 3]) and philosophical views of argumentation.

The aim of this paper is to present a formal model for procedural texts. The model is an extension of a framework developed in [2] for reasoning about agent's desires. The idea is to built plans for achieving those desires, to resolve the conflicts among those plans, and finally to select the set of desires that are achievable together.

In the next sections of this paper, we present some details about a typology of arguments in procedural texts and a motivational example. Then, we present an extension of the formal model developed in [2] for modeling procedural texts and some essential properties.

2 Procedural texts and argumentation

2.1 Role of arguments

In [4], we present in detail the different linguistic and conceptual forms of arguments found in procedural texts. This is a study done for french. Let us review here the 5 major forms of arguments we found frequently in corpora. Verb classes referred to are in general those specified in WordNet:

- Explanations are the most usual ones. We find them in any kind of procedural texts. They usually introduce a set of sequences or more locally an instruction implemented in the "goal" symbol of the grammar. The abstract schemas are the following: (1) purpose connectors-infinitive verbs, (2) causal connectors-deverbals and (3) titles. The most frequently used causal connectors are : pour, afin de, car, c'est pourquoi, etc. (to, in order to) (e.g. to remove the bearings, for lubrification of the universal joint shafts, because it may be prematurely worn due to the failure of another component).
- Warning arguments embedded mostly either in a "negative" formulation. They are particularly rich in technical domains. Their role is basically to explain and to justify. Negative formulation is easy to identify: there are prototypical expressions that introduce the arguments. Negative formulation follows the abstract schemas:

[1] IRIT - CNRS 118, route de Narbonne 31062 Toulouse Cedex 9, France amgoud, aouladom, stdizier@irit.fr

negative causal connectors-infinitive risk verbs; negative causal marks-risk VP; positive causal connectors-VP negative syntaxic forms, positive causal connectors-prevention verbs.

  - negative connectors: sous peine de, sinon, car sinon, sans quoi, etc. (otherwise, under the risk of) (e.g. sous peine d'attaquer la teinte du bois).
  - risk class verbs: risquer, causer, nuire, commettre etc. (e.g. pour ne pas commettre d'erreur).
  - prevention verbs: viter, prvenir, etc. (e.g. afin d'viter que la carte se dchausse lorsqu'on la visse au chssis, gloss: in order to prevent the card from skipping off its rack).
  - Positive causal mark and negative syntaxic forms: de facon ne pas, pour ne pas, pour que ... ne ...pas etc. (in order not to) (e.g. pour ne pas le rendre brillant, gloss: in order not to make it too bright).

- Tip arguments: these arguments are less imperative than the other ones, they guide the user, motivate him, and help to evaluate the quality of the work. They are particularly present in communication texts. The corresponding abstract schemas are: causal connectors-performing NP; causal connectors-performing verbs; causal connectors-modal-performing verbs; performing proposition.
  - performing verbs: e.g. permettre, amliorer, etc. (allow, improve, etc.).
  - performing PPs: e.g. Pour une meilleure finition; pour des raisons de performances (for a better finishing, for performing reasons).
  - performing proposition: e.g. Have small bills. It's easier to tip and to pay your bill that way.
- threatening arguments and reward arguments: these arguments have a strong impact on the user's intention to realize the instruction provided, the instruction is almost made compulsory by using this kind of argument. This is the injunctive form. We could not find any of these types of arguments in procedural texts, except in QA pairs and injunctions texts (e.g. rules) where the author and the adressee are clearly identified. Therefore, in those arguments we often find personal pronouns like "nous" "vous" (we, you). For threatening arguments, it follows the following schemas: otherwise connectors-consequence proposition; otherwise negative expression-consequence proposition :
  - otherwise connectors: sinon.
  - otherwise negative expression: si ... ne ... pas... (e.g. si vous ne le faites pas, nous le primerons automatiquement aprs trois semaines en ligne).
- For reward arguments, the schemas associated are the following : personal pronouns - reward proposition :
  - reward proposition : using possession transfer verbs (gagner, donner, bnficier, etc. (win, give, benefit)

Besides these five main types of arguments, we found some forms of stimulation-evaluation (what you only have to do now...), and evaluation.

2.2 The pragmatic dimensions of argumentative aims

First, it is important to note that arguments associated to a task do not form a homogeneous group. Arguments have different types (as specified above), and range over various facets of the task to carry out. We can talk, similarly to the temporal organization where some tasks may evolve in parallel with little connections, of a polyphony of arguments, to be contrasted with a sequence of arguments jointly operating over the same set of data.

Another important aspect is that, besides their direct use and meaning, arguments or groups of arguments convey several pragmatic effects which are quite subjective. For example, a task may become more salient in the procedural discourse if it is associated with a large number of arguments. Arguments therefore may induce zoom effects on some instructions. Arguments are also often exaggerated, beyond normal expectations, as a way to strengthen them, and to arouse a greater attention from the reader. In texts dedicated to the large public, arguments may be too strong, in particular precautions to take (a form of warning). The result is that the global coherence of arguments over a whole procedural text may not be fully met, while the text remains perfectly 'coherent' from the reader point of view. Finally, a task may be associated with a disjunction of arguments, whose selection depends on the reader's performances and preferences, for example tips and explanations may be tuned to various audiences.

These short considerations are illustrated below. They obviously have an impact on the formal model, in which we will need to introduce temporal dimensions, flexible forms of coherence, locally and globally, and over the various types of arguments (e.g. a tip must not contradict a warning), preferences and salience effects.

3 Illustrative example (Assembling a PC)

Let us illustrate the previous section by means of a simple, real example, extracted from the Web, which will be used throughout the remainder of this paper. The example is about *assembling a PC*. The following instructions are given for that purpose:

Assembling your PC

Material required: Make sure that you have all the below materials before starting: Processors, Motherboard, Hardisk, RAM, Cabinet, Floppy drive, . . .

Precautions: Before starting the actual assembly, the following precautions would help to avoid any mishap during the assembly process:

- be sure to handle all the components with great care, . . .
- use a clean and large enough table, . . .
- avoid the presence of any source of static electricity around,. . .

Procedure:

- *Installing Hardisk*: Ensure that the hard drive is set up to be the master drive on its IDE cable. If so plug it in . . .
- *Floppy Drive*: Plug in the power cable (see picture) carefully since it is quite possible to miss one of the connectors, which will quite possibly cause some damage when the computer is powered on. Then, place drive in slot, . . .

As the reader can note it, procedural texts contain a large number of arguments under the form of advices, warnings, etc. which do help to realize the action. Note also the elliptical style of some titles, with no verbs, but which are nevertheless actions.

4 Logical language

Let $\mathcal{L}$ be a logical language, and $\mathtt{Arg}(\mathcal{L})$ the different arguments that can be built from $\mathcal{L}$. From $\mathcal{L}$, three bases can be distinguished:

- $\mathcal{G}$ contains formulas of $\mathcal{L}$. Elements of $\mathcal{G}$ represent the *subject* or the *goals* to be satisfied through the procedural text. For instance, the goal of the procedural text given in Example 1 is "assembling a PC". Note that, for the same text, the set $\mathcal{G}$ should be consistent.
- $\mathcal{P}$ contains rules having the form $\varphi_1 \wedge \ldots \wedge \varphi_n \rightarrow \varphi$ where $\varphi_1, \ldots, \varphi_n, \varphi$ are elements of $\mathcal{L}$. Such a formula means that the author believes that if the actions $\varphi_1, \ldots, \varphi_n$ are achieved then φ will also be achieved.

Example 1 *In the above example, the different bases contain among others the following information: $\mathcal{G}$ = {Assembling the Computer}. $\mathcal{P}$ = { Check Required Material $\wedge$ Installing Hard Drive $\wedge$ Floppy Drive $\rightarrow$ Assembling the Computer, Processors $\wedge$ Motherboard $\wedge$ Hardisk $\wedge$ RAM $\wedge$ Cabinet $\wedge$ Floppy Drive $\rightarrow$ Check Required Material set up the harddisk $\wedge$ plug in $\rightarrow$ Installing HardDisk, plug in the power cable $\wedge$ place drive in slot $\rightarrow$ Floppy Drive}.*

5 A basic argumentation system

A rational agent can express claims and judgments, aiming at reaching a decision, a conclusion, or informing, convincing, negotiating with other agents. Pertinent information may be insufficient or conversely there may be too much, but partially incoherent information. In case of multi-agent interactions, conflicts of interest are unavoidable. Agents can be assisted by argumentation, a process based on the exchange and the valuation of interacting arguments which support opinions, claims, proposals, decisions, etc. According to Dung [5], an argumentation framework is defined as a pair consisting of a set of arguments and a binary relation representing the defeasibility relationship between arguments.

Definition 1 *(Argumentation framework) An* argumentation framework *is a pair <$\mathcal{A}$, $\mathcal{R}$> where $\mathcal{A}$ is a set of arguments ($\mathcal{A} \subseteq \mathtt{Arg}(\mathcal{L})$), and $\mathcal{R}$ is a binary relation representing a defeasibility relationship between arguments, i.e. $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. $(a, b) \in \mathcal{R}$ or equivalently $a\mathcal{R}b$ means that the argument a defeats b.*

In the above definition, the structure of the argument is unknown. In the remainder of this paper, we do not need to define formally an argument. However, any argument $a \in \mathcal{L}$ is supposed to have a conclusion that is returned by the function Conc.

Since arguments may be conflicting, it is important to know which arguments are considered *acceptable*. Dung has defined different acceptability semantics.

Definition 2 (Defence/conflict-free) *Let $S \subseteq \mathcal{A}$.*

- *S* defends *an argument A iff each argument that defeats A is defeated by some argument in S.*
- *S is* conflict-free *iff there exist no A_i, A_j in S such that A_i defeats A_j.*

Definition 3 (Acceptability semantics) *Let S be a conflict-free set of arguments and let $\mathcal{F}$: $2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$ be a function such that $\mathcal{F}(S)$ = $\{A \mid S$ defends $A\}$.*

- *S is a* complete *extension iff $S = \mathcal{F}(S)$.*
- *S is a* preferred *extension iff S is a maximal (w.r.t set $\subseteq$) complete extension.*
- *S is a* grounded *extension iff it is the smallest (w.r.t set $\subseteq$) complete extension.*

Note that there is only one grounded extension. It contains all the arguments that are not defeated, and the arguments that are defended directly or indirectly by non-defeated arguments.

The last step of an argumentation process consists of determining, among all the conclusions of the different arguments, the "good" ones called *justified conclusions*. Let Output denote this set of justified conclusions. One way of defining Output is to consider the conclusions that are supported by at least one argument in each extension.

Definition 4 (Justified conclusions) *Let $(\mathcal{A}, \mathcal{R})$ be an argumentation system and $\{E_1, \ldots, E_n\}$ be its set of extensions (under a given semantics).* Output = $\{\psi | \forall E_i, \exists A \in E_i$ *such that* $\mathtt{Conc}(A) = \psi\}$.

6 A formal model for procedural texts

In this section we propose a formal framework for procedural texts. This framework builds on a model developed in [2] for reasoning about conflicting desires. The basic idea behind that model is to construct plans for each desire and then to select the set of desires that are achievable together. A plan consists in decomposing the initial desire into sub-desires that are themselves decomposed into other sub-desires. This gives birth to a tree structure, where the leaves of the tree are instructions. In what follows, we will adopt this notion of plan for modeling a procedural text.

The basic concept of our framework is that of *goal*. Indeed, each procedural text is supposed to have a goal. A goal is any element of $\mathcal{G}$, and it may have sub-goals.

Definition 5 (Goal/Sub-goal) *Let us consider the bases <$\mathcal{G}$, $\mathcal{P}$>.*

1. *$\mathcal{G}$ is the set of* goals.
2. *$Sub\mathcal{G}$ is the set of the* sub-goals*: A literal $h' \in Sub\mathcal{G}$ iff there exists a rule $\varphi_1 \wedge h' \ldots \wedge \varphi_n \rightarrow \varphi \in \mathcal{P}$ with $\varphi \in \mathcal{G}$ or $\varphi \in Sub\mathcal{G}$. In that case, h' is a sub-desire of φ.*

A goal can be achieved in different ways. We bring the two notions together in a new notion of *partial plan*.

Definition 6 (Partial plan) *A* partial plan *is a pair a = $<h, H>$ such that:*

- *h is a goal or a sub-goal.*
- *H = $\{\varphi_1, \ldots, \varphi_n\}$ if there exists a rule $\varphi_1 \wedge \ldots \wedge \varphi_n \rightarrow h \in \mathcal{P}$, H = $\emptyset$ otherwise.*

The function $\mathtt{Goal}(a)$ = h returns the goal or sub-goal of a partial plan a. $\aleph$ will gather all the partial plans that can be built from <$\mathcal{G}$, $\mathcal{P}$>.

Remark 1 *A goal may have several partial plans corresponding to different alternatives for achieving that goal. Indeed, in procedural texts, it may be the case that for the same goal/sub-goals, several ways for achieving it are provided.*

Remark 2 *Let a = $< h, H >$ be a partial plan. Each element of the support H is a sub-goal of h.*

Definition 7 *A partial plan a = $<h, H>$ is* elementary *iff $H = \emptyset$.*

Remark 3 *If there exists an elementary partial plan for a goal h, then this means that the agent knows how to achieve h directly. This corresponds to the notion of instructions of procedural texts.*

Example 2 *In the above example, the following partial plans can be built: <Assembling the Computer, {Check Required Material, Installing Hard Disk, Floppy Drive}>, <Check Required Material, {Processors, Motherboard, Hardisk, RAM, Cabinet, Floppy Drive}>, <Installing Hard Disk, {Set up the harddisk, plug in}>, <Floppy Drive, {plug in the power cable, place drive in slot}>.*

A partial plan shows the actions that should be performed in order to achieve the corresponding goal (or sub-goal). However, the elements of the support of a given partial plan are considered as sub-goals that must be achieved in turn by another partial plan. The whole way to achieve a given goal is called a *complete plan*. A *complete plan* for a goal d is an *AND* tree. Its nodes are partial plans and its arcs represent the sub-goal relationship. The root of the tree is a partial plan for the goal d. It is an AND tree because all the sub-goals of d must be considered. When for the same goal, there are several partial plans to carry it out, only one is considered in a tree. Formally:

Definition 8 (Complete plan) *A* complete plan *G for a goal h is a finite tree such that:*

- *$h \in \mathcal{G}$.*
- *The root of the tree is a partial plan <h, H>.*
- *A node <h', $\{\varphi_1, \ldots, \varphi_n\}$> has exactly n children $<\varphi_1, H'_1>$, …, $<\varphi_n, H'_n>$ where $<\varphi_i, H'_i>$ is an element of $\aleph$.*
- *The leaves of the tree are elementary partial plans.*

The function Nodes(G) *returns the set of all the partial plans of the tree G. $\mathcal{CP}$ denotes the set of all the complete plans that can be built from $<\mathcal{G}, \mathcal{P}>$. The function* Leaves($G$) *returns the set of the leaves of the tree G.*

Example 3 *In our training example, there is a unique complete plan for the goal "assembling a PC" that is shown in Figure 1.*

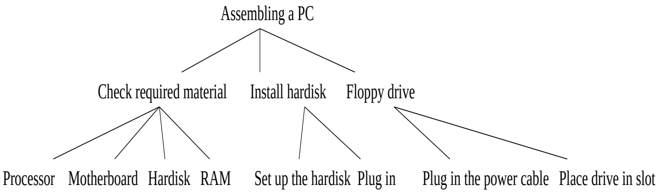


Figure 1. Complete plan

Note that a procedural text may have several complete plans capturing the different ways for achieving the goal of the text.

As said in the introduction, a procedural text may contain arguments for explaining different tasks, and for motivating the reader to behave in a certain way. In [4], we have shown there are mainly tow categories of arguments that are used in procedural texts: *advices* and *warnings*. It is also common that some arguments in a procedural text may defeat other arguments. Indeed, some authors explain a task, and present for that purpose arguments. Since, the authors may expect counter-arguments from the readers, then they introduce those counter-arguments in the text itself and present the counter-attack against them. In sum, the tasks of the procedural text should be justified, and defended in the text.

Now that all the ingredients introduced, we are ready to define formally a procedural text. Indeed, a procedural text has three components: a *goal* that it should satisfy, a *complete plan* for achieving that goal, and an argumentation system that justifies each goal/sub-goal occurring in the complete plan.

Definition 9 (Procedural text) *A* procedural text *is a tuple <g, G, AS> where:*

- *$g \in \mathcal{G}$ is the goal of the procedural text*
- *$G \in \mathcal{CP}$ is a complete plan for g*
- *AS = $<\mathcal{A}, \mathcal{R}>$ is an argumentation system*
- $\forall\, a_i \in \texttt{Node}(G)$, $\texttt{Goal}(a_i) \in \texttt{Outcome}(AS)$

The last condition ensures that the procedural text is coherent in the sense that each goal and sub-goal is justified and correctly supported. This means that the arguments exchanged for supporting a given goal/sub-goal cannot defeat another goal/sub-goal.

7 Conclusion

This paper has proposed a formal model for defining procedural texts. We have mainly shown how these texts can be defined in a more abstract way. Due to the tree structure of procedural texts and their decomposition in terms of tasks and sub-tasks, we have defined a procedural text as a plan for achieving its goal. This formal model makes it possible to easily compare different procedural texts. For instance, a procedural text in which the set $\mathcal{A}$ of arguments is empty is poor, and may be directed towards a professional audience. Therefore further investigations should be carried out in order to study different strategies an author may use according to target audiences.

REFERENCES

[1] J. M. Adam, 'Types de textes ou genres de discours ? comment classer les textes qui disent de et comment faire', *Langages*, **141**, 10–27, (2001).

[2] L. Amgoud, 'A formal framework for handling conflicting desires', in *Proceedings of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2003*, pp. 552–563, (2003).

[3] J.-Cl. Anscombre and O. Ducrot, 'Interrogation et argumentation', *Langue franaise*, **52, L'interrogation**, (1981).

[4] F. Aouladomar, L. Amgoud, and P. Saint-Dizier, 'On argumentation in procedural texts', in *Proceedings of the International Symposium: Discourse and Document*, (2006).

[5] 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).

[6] J. Moeschler, 'Argumentation et conversation, éléments pour une analyse pragmatique du discours', *Hatier - Credif*, (1985).

The Carneades Argumentation Framework – Using Presumptions and Exceptions to Model Critical Questions

Thomas F. Gordon[1] and **Douglas Walton**[2]

Abstract. In 2005, Gordon and Walton presented initial ideas for a computational model of defeasible argument [12, 26], which builds on and elaborates Walton's theory of argumentation [28, 31]. The current paper reports on progress which has been made in the meantime. It presents a formal, mathematical model of argument evaluation which applies proof standards [8] to determine the defensibility of arguments and the acceptability of statements on an issue-by-issue basis. The main original contribution of the Carneades Argumentation Framework is its use of three kinds of premises (ordinary premises, presumptions and exceptions) and information about the dialectical status of statements (undisputed, at issue, accepted or rejected) to model critical questions in such a way as to allow the burden of proof to be allocated to the proponent or the respondent, as appropriate. Both of these elements are required for this purpose: presumptions hold without supporting argument only so long as they have not been put at issue by actually asking the critical question.

1 Introduction

The work in this paper flows from previous attempts to solve a key problem common to AI and argumentation theory concerning the using of the device of critical questions to evaluate an argument. Critical questions were first introduced by Arthur Hastings [15] as part of his analysis of presumptive argumentation schemes. The critical questions attached to an argumentation scheme enumerate ways of challenging arguments created using the scheme. The current method of evaluating an argument that fits a scheme, like that for argument from expert opinion, is by a shifting of the burden of proof from one side to the other in a dialog [30]. When the respondent asks one of the critical questions matching the scheme, the burden of proof shifts back to the proponent's side, defeating or undercutting the argument until the critical question has been answered successfully. At least this has been the general approach of argumentation theory. Recently, however, it was observed [3] that critical questions differ with respect to their impact on the burden of proof. These observations led to two theories about the shifting of the burden of proof when critical questions are asked. According to one theory, when any critical question is asked, the burden shifts to the proponent's side to answer the question and, if no answer is given, the argument fails. According to the other theory, merely asking a critical question is not enough to shift the burden of proof back to the proponent. On this theory, to make the argument fail, the question needs to be supported by further argument. Some critical questions fit one theory better, while others fit the other theory better. This duality has posed a recurring problem for the project of formalizing schemes.

In this paper, we put forward a new model for evaluating defeasible arguments that solves this problem, continuing work we began in 2005 [12, 26]. The current paper presents a formal, mathematical model of argument evaluation which applies proof standards [8] to determine the defensibility of arguments and the acceptability of statements on an issue-by-issue basis. The formal model is called the Carneades Argumentation Framework, in honor of the Greek skeptic philosopher who emphasized the importance of plausible reasoning [6, vol. 1, p. 33-34].

Arguments in Carneades are identified, analyzed and evaluated not only by fitting premise-conclusion structures that can be identified using argumentation schemes. Arguments also have a dialectical aspect, in that they can be seen as having been put forward on one side or the other of an issue during a dialog. The evaluation of arguments in Carneades depends on the stage of the dialog. Whether or not a premise of an argument holds depends on whether it is undisputed, at issue, or decided. One way to raise an issue is to ask a critical question. Also, the proof standard applicable for some issue may depend on the stage of the dialog. In a deliberation dialog, for example, a weak burden of proof would seem appropriate during brainstorming, in an early phase of the dialog. The Carneades Argumentation Framework is designed to be used in a layered model of dialectical argument [19] for various kinds of dialogs, where higher layers are responsible for modeling such things as speech acts, argumentation protocols and argument strategies.

The rest of the paper is structured as follows. The next two sections formally define the Carneades Argumentation Framework. Section 2 defines the structure of arguments and illustrates this structure with examples from related work by Toulmin, Pollock and others. Section 3 formally defines how arguments are evaluated in terms of the acceptability of statements, the defensibility of arguments, and the satisfiability

[1] Fraunhofer FOKUS, Berlin, Germany, email: thomas.gordon@fokus.fraunhofer.de

[2] Department of Philosophy, University of Winnipeg, Winnipeg, Manitoba, Canada, email: d.walton@uwinnipeg.ca

of proof standards. Section 4 illustrates argument evaluation with an example from the AI and Law literature. The paper closes in Section 5 with a brief discussion of related work and some ideas for future work.

2 Argument Structure

We begin by defining the structure of arguments. Unlike Dung's model [5], in which the internal structure of arguments is irrelevant for the purpose of determining their defensibility, our model makes use of and depends on the more conventional conception of argument in the argumentation theory literature, in which arguments are a kind of conditional linking a set of premises to a conclusion. Intuitively, the premises and the conclusion of arguments are statements about the world, which may be accepted as being true or false. In [12] the internal structure of statements was defined in such a way as to enable the domain of discourse to be modeled in a way compatible with emerging standards of the Semantic Web [2]. These details, however, need not concern us here. For the purpose of evaluating arguments, the internal structure of statements is not important. We only require the ability to compare two statements to determine whether or not they are equal.

Definition 1 (Statements) *Let* $\langle \text{statement}, = \rangle$ *be a structure, where* statement *denotes the set of declarative sentences in some language and* $=$ *is an equality relation, modeled as a function of type* statement × statement → boolean.

Next, to support defeasible argumentation and allow the burden of proof to be distributed, we distinguish three kinds of premises.

Definition 2 (Premises) *Let* premise *denote the set of premises. There are three kinds of premises:*

1. *If* s *is a* statement*, then* premise(s) *is a premise. These are called* ordinary premises*. As a notational convenience, we will use a statement* s *alone to denote* premise(s)*, when the context makes it clear that the statement is being used as a premise.*
2. *If* s *is a* statement*, then* $\bullet s$*, called a* presumption*, is a premise.*
3. *If* s *is a* statement*, then* $\circ s$*, called an* exception*, is a premise.*
4. *Nothing else is a premise.*

Now we are ready to define the structure of arguments.

Definition 3 (Arguments) *An* argument *is a tuple* $\langle c, d, p\rangle$*, where* c *is a* statement*,* $d \in \{\text{pro}, \text{con}\}$ *and* $p \in \mathcal{P}(\text{premise})$*. If* a *is an argument* $\langle c, d, p\rangle$*, then* conclusion(a) $= c$, direction(a) $= d$ *and* premises(a) $= p$*. Where convenient, pro arguments will be notated as* $p_1, \ldots, p_n \rightarrow c$ *and con arguments as* $p_1, \ldots, p_n \multimap c$*.*

This approach, with two kinds of arguments, pro and con, is somewhat different than the argument diagramming model developed by Walton in [28] and implemented in Araucaria. There counterarguments are modelled as arguments pro some statement which has been asserted to be in conflict with the conclusion of the other argument, called a *refutation*. Our approach, with its two kinds of arguments, is not uncommon in the literature on defeasible argument [18, 22, 14, 13].

We assume arguments are asserted by the participants of a dialog. We have specified and implemented a simple communication language and argumentation protocol to test Carneades, but that is a subject for another paper. For our purposes here, it is sufficient to note that argument moves, i.e. speech acts, are modelled as functions which map a state of the dialog to another state. (Again, this is a purely functional model, so states are not modified.) A dialog state is a tuple $\langle t, h, G\rangle$, where t is a statement, the *thesis* of the dialog, h is a sequence of moves, representing the history of the dialog, and G is an *argument graph*.[3]

It is these argument graphs which concern us here. An argument graph plays a role comparable to a set of formulas in logic. Whereas in logic the truth of a formula is defined in terms of a (consequence) relation between sets of formulas, here we will define the *acceptability* of statements in argument graphs. An argument graph is not merely a set of arguments. Rather, as its name suggests, it is a finite graph. There are two kinds of nodes, statement nodes and argument nodes. The edges of the graph link up the premises and conclusions of the arguments. Each statement is represented by at most one node in the graph.

To illustrate argument graphs, suppose we have the following (construed) arguments from the domain of contract law:

a1. agreement, $\circ$ minor → contract
a2. oral, $\bullet$ estate $\multimap$ contract
a3. email → oral
a4. deed $\multimap$ agreement
a5. $\bullet$ deed → estate

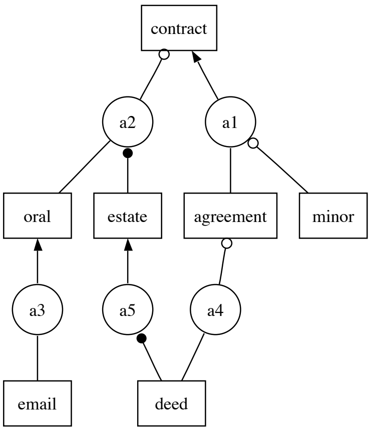


Figure 1. Argument Graph

The argument graph induced by these arguments is shown in Figure 1. In this figure, statements are displayed as boxes and arguments as circles. Different arrowhead shapes are used to distinguish pro and con arguments as well as the three

[3] In prior work [11, 13], Gordon has referred to argument graphs as *dialectical graphs*.

kinds of premises. Pro arguments are indicated using ordinary arrowheads; con arguments with open-dot arrowheads. Ordinary premises are represented as edges with no arrowheads, presumptions with closed-dot arrowheads and exceptions with open-dot arrowheads. (The direction of the edge is implicit in the case of ordinary premises; the direction is always from the premise to the argument.) Notice that the premise type cannot be adequately represented using statement labels, since argument graphs are not restricted to trees. A statement may be used in multiple arguments and as a different type of premise in each argument. The above example illustrates this point. The fourth and the fifth arguments each use the statement 'deed' in a premise. In the fourth argument it is used in an ordinary premise but in the fifth it is used in a presumption. Walton has called this use of shared premises a *divergent argument structure* [28, p. 91].

Although argument graphs are not restricted to trees, they are not completely general; we do not allow cycles. This restriction assures the decidability of the defensibility and acceptabilty properties of arguments and statements, respectively.

Definition 4 (Argument Graphs) *An* argument-graph *is a labeled, finite, directed, acyclic, bipartite graph, consisting of* argument *nodes and* statement *nodes. The edges link the argument nodes to the statements in the premises and conclusion of each argument.*

This completes the formal definition of the structure of arguments and argument graphs. Let us now discuss briefly the expressiveness of this model, beginning by comparing our approach with Toulmin's model [21]. Recall that arguments in Toulmin's model consist of a single premise, called the *datum*; a conclusion, called the *claim*; a kind of rule, called the *warrant*, which supports the inference from the premise to the conclusion of the argument; an additional piece of data, called *backing*, which provides support for the warrant; an exception, called a *rebuttal*; and, finally, a *qualifier* stating the probative value of the inference (e.g. presumably, or necessarily). Of these, the datum and conclusion are handled in a straightforward way in our model. The set of premises of an argument generalizes the single datum in Toulmin's system. Claims are modeled comparably, as conclusions. Rebuttals are modeled with con arguments. The probative weight of an argument is handled as part of our model of proof standards, as will be explained shortly.

This leaves our interpretation of warrants and backing to be explained. Our model does not directly allow arguments about other arguments. (The conclusion of an argument must be a statement.) Rather, the approach we prefer is to add a presumption for the warrant to the premises of an argument. If an argument does not have such a presumption, the argument graph can first be extended to add one. We leave it up to the argumentation protocol of the procedural model to regulate under what conditions such *hidden premises* may be *revealed*. In effect, the datum and warrant are modelled as minor and major premises, much as in the classical theory of syllogism. Backing, in turn, can be modelled as a premise of an argument supporting the warrant.

For example, here is a version of Toulmin's standard example about British citizenship.

Datum. Harry was born in Bermuda.

Claim. Harry is a British subject.

Warrant. A man born in Bermuda will generally be a British subject.

Backing. Civil Code §123 provides that persons born in Bermuda are generally British subjects.

Exception. Harry has become an American citizen.

The argument can be reconstructed in our framework as illustrated if Figure 2.

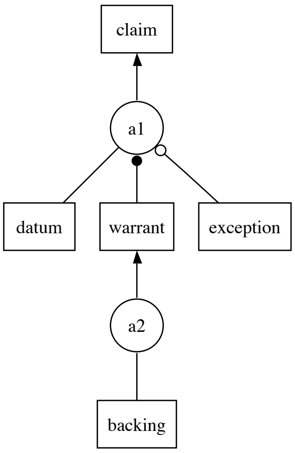


Figure 2. Reconstruction of Toulmin Diagrams

This approach generalizes Toulmin's model, by supporting arguments pro and contra both warrants and backing, using the same argumentation framework as for arguments about any other kind of claim. Indeed, Toulmin appears to have overlooked the possibility of arguing against warrants or making an issue out of backing claims.

Our model of argument is rich enough to handle Pollock's concepts of rebuttal, premise defeat and undercutting defeaters [18]. Rebuttals can be modeled as arguments in the opposite direction for the same conclusion. (If an argument a_1 is *pro* some statement s, then some argument a_2 *con* s is a rebuttal of a_1, and vice versa.) Premise defeat can be modeled with arguments con an ordinary premise or presumption, or pro an exception.

Undercutting defeaters are a bit trickier. The idea of an undercutting defeater is to argue against the argument itself, or the rule or warrant which was applied to create the argument. We model undercutting defeaters by revealing and then attacking premises, similar to the way we handled warrants in the reconstruction of Toulmin's system. Consider Pollock's example of things which look red but turn out to be illuminated by a red light:

Red. The object is red.

Looks Red. The object looks red.

Applicable. The general rule "Things which look red are red." applies to this object.

Illuminated. The object is illuminated by a red light.

An argument graph for this example is shown in Figure 3. Rather than undercutting argument a_1 (the object is red because it looks red) directly, with an argument contra a_1, we undercut the argument by first revealing a presumption (about the general rule being applicable in this case) and then assert an argument contra this presumption. Notice by the way that another presumption is still implicit in this example, namely a presumption for the "warrant" about things which look red being red.

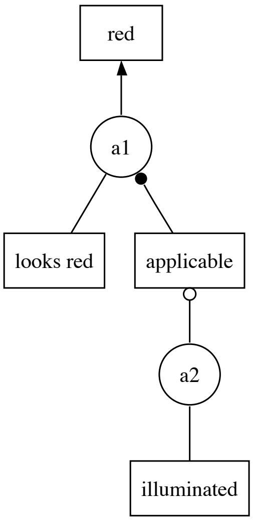


Figure 3. Undercutting Defeater Example

Walton [28] distinguishes two kinds of arguments, called *convergent* and *linked* arguments. Convergent arguments provide multiple reasons for a conclusion, each of which alone can be sufficient to accept the conclusion. Convergent arguments are handled in our approach by multiple arguments for the same conclusion. Linked arguments, on the other hand, consist of two or more premises which all must hold for the argument to provide significant support for its conclusion. Linked arguments are handled in our approach by defining arguments to consist of a set of premises, rather than a single premise, and defining arguments to be defensible only if all of their premises hold. (The concept of argument defensibilty is formally defined below.)

Presumptions and exceptions are a refinement of Walton's concept of *critical questions* [29]. Critical questions enumerate specific ways to defeat arguments matching some argument scheme. But so long as an issue has not been raised by actually asking some critical question, we would like to be able to express which answer to presume. The distinction between presumptions and exceptions here provides this ability.

Consider the scheme for arguments from expert opinion [25]:

Major Premise. Source E is an expert in the subject domain S containing proposition A.

Minor Premise. E asserts that proposition A in domain S is true.

Conclusion. A may plausibly be taken as true.

The scheme includes six critical questions:

CQ1. How credible is E as an expert source?
CQ2. Is E an expert in the field that A is in?
CQ3. Does E's testimony imply A?
CQ4. Is E reliable?
CQ5. Is A consistent with the testimony of other experts?
CQ6. Is A supported by evidence?

When the scheme for arguments from expert opinion is instantiated to create a specific argument, the critical questions can be represented, in our model, as presumptions and exceptions. Whether a presumption or exception is appropriate depends on the burden of proof. If the respondent, the person who poses the critical question, should have the burden of proof, then the critical question should be modeled as an exception. If, on the other hand, the proponent, the party who used the schema to construct the argument, should have the burden of proof, then the critical question should be modeled as a presumption.[4]

Our model does not require that premises for critical questions be made explicit at the time the argument is first made. Rather, they can be *revealed* incrementally during the course of the dialog. The conditions under which a premise may be left implicit or revealed raise procedural issues which need to be addressed in the protocol for the type of dialog. Our contribution here is to provide an argumentation framework which can be used for modeling such protocols.

3 Argument Evaluation

By argument evaluation we mean determining whether a statement is *acceptable* in an argument graph. As we will see soon, this in turn will depend on the *defensibility* of arguments in the graph. Notice that our terminology is somewhat different than Dung's [5], who speaks of the acceptability of arguments, rather than their defensibility. Also, for those readers familiar with our preliminary work on this subject in [12], please notice that the terminology and other details of the current model are different, even though the basic ideas and general approach are quite similar.

The definition of the acceptability of statements is recursive. The acceptability of a statement depends on its *proof standard.* Whether or not a statement's proof standard is *satisfied* depends on the defensibility of the arguments pro and con this statement. The defensibility of an argument depends on whether or not its premises *hold.* Finally, we end up where we began: Whether or not a premise holds can depend on whether or not the premise's statement is acceptable. Since the definitions are recursive, we cannot avoid making forward references to functions which will be defined later.

To evaluate a set of arguments in an argument graph, we require some additional information. Firstly, we need to know the current *status* of each statement in the dialog, i.e. whether it is accepted, rejected, at issue or undisputed. This status information is pragmatic; the status of statements is set by speech acts in the dialog, such as asking a question, asserting an argument or making a decision. Secondly, we assume that a proof standard has been assigned to each statement. We do

[4] We agree with Verheij [24] that critical questions which are entailed by the premises of the argument schema are redundant and may be omitted. This is arguably the case in the example for the first three critical questions.

not address the question of how this is done. Presumably this will depend on domain knowledge and the type of dialog. Finally, one of the proof standards we will define, *preponderance of the evidence*, makes use of numerical weights, comparable to conditional probabilities. To use this proof standard, we require a weighing function.

Let us formalize these requirements by postulating an *argument context* as follows.

Definition 5 (Argument Context) *Let $\mathcal{C}$, the argument context, be a tuple* $\langle G, \text{status}, \text{proof-standard}, \text{weight}\rangle$, *where G is an* argument-graph*,* status *is a function of type* statement $\rightarrow$ $\{\text{accepted}, \text{rejected}, \text{undisputed}, \text{issue}\}$, proof-standard *is a function of type* statement $\rightarrow$ $\{\text{SE}, \text{PE}, \text{DV}, \text{BRD}\}$ *and* weight *is a function of type* statement $\times$ statement $\rightarrow \{0, \ldots, 10\}$

Intuitively, a statement which has been used in a dialog is initially undisputed. Later in the dialog, an issue can be made out of this statement. Presumably after arguments pro and con have been collected for some period of time, a decision will be taken and the statement will be either accepted or rejected. The details of how this is done need not concern us further here. These are matters which need to be addressed fully when modeling protocols for dialogs.

Definition 6 (Acceptability of Statements) *Let* acceptable *be a function of type* statement $\times$ argument-graph $\rightarrow$ boolean*. A statement is acceptable in an argument graph if and only if it satisfies its proof standard in the argument graph:* $\text{acceptable}(s, ag) = \text{satisfies}(s, \text{proof-standard}(s), ag)$.

Definition 7 (Satisfaction of Proof Standards) *A proof standard is a function of type* statement $\times$ argument-graph $\rightarrow$ boolean*. Let f be a proof standard.* $\text{satisfies}(s, f, G) = f(s, G)$

Four proof standards are defined in this paper.

SE. A statement meets this standard iff it is supported by at least one defensible pro argument.

PE. A statement meets this standard iff its strongest defensible pro argument outweighs its strongest defensible con argument. This standard balances arguments using probative weights.

DV. A statement meets this standard iff it is supported by at least one defensible pro argument and none of its con arguments are defensible.

BRD. A statement meeets this standard iff it is supported by at least one defensible pro argument, all of its pro arguments are defensible and none of its con arguments are defensible.

The names of three of these standards are meant to suggest three legal proof standards: scintilla of evidence, preponderance of the evidence and beyond a reasonable doubt. However, we do not claim that the definitions of these standards, above, fully capture their legal meanings. What these standards have in common with their legal counterparts is their relative strength. If a statement satisfies a proof standard, it will also satisfy all weaker proof standards.

The name of the DV proof standard is an acronym for *dialectical validity*, a term used by Freeman and Farley [8]. They defined five proof standards. In addition to the four we have defined here, they included a fifth, called *beyond a doubt*, which was defined to be an even stronger standard than *beyond a reasonable doubt*.

The preponderance of evidence (PE) standard compares the weight of arguments. The weight of an argument is defined to be the same as the weight of its *weakest premise*, i.e., to be precise, the same as the weight of the premise with the lowest weight. Recall we assume a weighing function, weight, as part of the context to provide this information. The weight of a premise p for a conclusion c is $\text{weight}(p, c)$. Other proof standards which aggregate and compare weights are conceivable. For example, one could sum the weights of the arguments pro and con and compare these sums.

We have defined weights to be natural numbers in the range of 0 to 10. We originally considered using real numbers in the range of 0.0 to 1.0, as in probability theory. However, on the assumption that the weights will be estimated by human users, we prefer to use a simpler ordinal scale, since we are skeptical that users can estimate such weights with a greater degree of accuracy.

All of the proof standards defined above depend on a determination of the *defensiblity* of arguments. Defensibility is defined next.

Definition 8 (Defensibility of Arguments) *Let* defensible *be a function of type* argument $\times$ argument-graph $\rightarrow$ boolean*. An argument α is defensible in an argument graph G if and only if all of its premises* hold *in the argument graph:* $\text{defensible}(\alpha, G) = \text{all}(\lambda p.\, \text{holds}(p, G))(\text{premises}\,\alpha)$.[5]

Finally, we come to the last definition required for evaluating arguments, for the holds predicate. This is where the status of a statement in the argument context and the distinction between ordinary premises, presumptions and exceptions come into play. Accepted presumptions and ordinary premises hold. Rejected presumptions and ordinary premises do not hold. Undisputed presumptions hold. Undisputed ordinary premises do not hold. An exception, $\circ s$, holds only if $\text{premise}(s)$ does not hold.

Definition 9 (Holding of Premises) *Let* holds *be a function of type* premise $\times$ argument-graph $\rightarrow$ boolean*. Let* $\sigma = \text{status}(s)$*. Whether or not a premise holds depends on its type (ordinary, presumption, or exception). Thus, there are the following three cases:*

If p is an ordinary premise, $\text{premise}(s)$*, then*

$$\text{holds}(p, G) = \begin{cases} true & \textit{if } \sigma = \text{accepted} \\ false & \textit{if } \sigma = \text{rejected} \\ acceptable(s, G) & \textit{if } \sigma = \text{issue} \\ false & \textit{if } \sigma = \text{undisputed} \end{cases}$$

If p is a presumption, $\bullet s$, then

[5] Here ‘all’ is a higher-order function, not a quantifier, applied to an anonymous function, represented with λ, as in lambda calculus.

$$\text{holds}(p, G) = \begin{cases} \mathit{true} & \mathit{if}\ \sigma = \text{accepted} \\ \mathit{false} & \mathit{if}\ \sigma = \text{rejected} \\ \mathit{acceptable}(s, G) & \mathit{if}\ \sigma = \text{issue} \\ \mathit{true} & \mathit{if}\ \sigma = \text{undisputed} \end{cases}$$

Finally, if p is an exception, $\circ s$, then

$$\text{holds}(p, G) = \neg\,\text{holds}(\text{premise}(s), G)$$

The important thing to notice is that whether or not a premise holds depends in this model not only on the arguments which have been asserted, but also on the kind of premise (ordinary, presumption, or exception) and the status of the premise's statement in the argument graph (undisputed, at issue, accepted, or rejected). We assume that the status of a statement progresses in the course of the dialog:

1. Initially, statements used in arguments are undisputed. Whether or not a premise which uses this statement holds at this stage of the dialog depends on the kind of premise. Ordinary premises do not hold; presumptions do hold. *This is the only semantic difference between ordinary premises and presumptions in our model.* An exception holds at this stage only if it would not hold if it were an *ordinary premise.* Notice that exceptions are not the dual of presumptions. As undisputed presumptions hold, an undisputed exception would not hold if we had defined exceptions to hold only if they would not hold if they were presumptions. But this is not the semantics we want. Rather, both undisputed exceptions and undisputed presumptions hold.
2. At some point a participant may make an issue out of a statement. Now ordinary premises and presumptions which use this statement hold only if they are acceptable, i.e. only if the statement meets its proof standard, given the arguments which have been asserted. Exceptions at issue hold only if the statement is not acceptable. We presume that arguments will be exchanged in a dialog for some period of time, and that during this phase the acceptability of statements at issue will be in flux.
3. Finally, at some point a decision will be made to either accept or reject some statement at issue. The model does not constrain the discretion of users to decide as they please. Unacceptable statements may be accepted and acceptable statements may be rejected. This remains transparent however. Any interested person can check whether the decisions are justified given the arguments made and the applicable proof standards. Anway, after a decision has been made, it is respected by the model: Accepted statements hold and rejected statements do not hold, no matter what arguments have been made or what proof standards apply.

4 An Example

Although our model of argument is rather simple, we claim, it is nonetheless rather difficult to illustrate all of its features, or indeed validate the model, with just a few examples. We have rather ambitious aims for the model. It should be sufficient for use as the *argumentation framework* layer [19] in procedural models of protocols for a wide variety of dialog types [31]. It should be sufficient as a basis for formal models of argument schemes, including critical questions. The distinction between the three kinds of premises should be adequate for allocating the burden of proof. It should be capable of being extended to handle other proof standards, such as more adequate models of legal proof standards. And of course it should yield intuitive results when applied to real examples of natural arguments. We have begun the work of testing and validating the model, but much work remains. Here we can only present a couple of examples to illustrate its main features.

As we are particularly interested in legal applications, we have reconstructed several examples from the Artificial Intelligence and Law literature [11, 17, 24, 1]. Some of these [11, 17] are procedural models of argumentation. Our reconstruction of these examples makes use of a procedural model of persuasion dialogs, based on the argumentation framework presented here. For lack of space, we will instead illustrate the model with one of the other examples which does do require us to address these procedural aspects.

We have selected one of Verheij's main examples [24, p. 69], which he calls the "grievous bodily harm" example. The example consists of the following statements.

8 years. The accused is punishable by up to 8 years in imprisonment.

bodily harm rule. Inflicting grievous bodily harm is punishable by up to 8 years imprisonment.

Article 302. According to article 302 of the Dutch criminal code, inflicting grievous bodily harm is punishable by up to 8 years imprisonment.

bodily harm. The accused has inflicted grievous bodily harm upon the victim.

10 witnesses. 10 pub customers' testimonies: the accused was involved in the fight.

accused's testimony I was not involved in the fight.

broken ribs not sufficient. Several broken ribs do not amount to grievous bodily harm.

precedent 1. The rule that several broken ribs does not amount to grievous bodily harm, explains precedent 1.

lex specialis. The rule explaining precedent 2 is more specific than the rule explaining precedent 1.

sufficient with complications. Several broken ribs with complications amount to grievous bodily harm.

precedent 2. The rule that several broken ribs with complications amount to grievous bodily harm, explains precedent 2.

hospital report. The victim has several broken ribs, with complications.

The arguments are displayed, together with their evaluation, in Figure 4. We've made some assumptions about the context, for the purposes of illustration:

- The status of statements is indicated in the diagram via a suffix: A question mark (?) means the statement is at issue; A plus sign (+) means it has been accepted; a minus sign (-) indicates it has been rejected; and the lack of a suffix means the statement is undisputed. The *lex specialis* and *10 witnesses* statements have been accepted. The statements of other leaf nodes are undisputed. All the other statements are at issue.
- The DV proof standard (dialectical validity) applies to all statements. This is closest to the evaluation criteria of Ver-

heij's model of argumentation, which does not support multiple proof standards.
- Weights are irrelevant in this example, since the PE proof standard (preponderance of the evidence) is not used.

Some further assumptions about the types of the premises have been made, to illustrate many features of the system with this one example. The result of the evaluation has been indicated in the diagram by filling in the nodes for acceptable statements and defensible arguments with a gray background. All the other statements are not acceptable and all other arguments are not defensible. Let us now try to explain the result, for each issue:

- The main issue, or thesis, that the accused is punishable by up to 8 years in prison, is acceptable. This is because both premises of the argument a_1 are acceptable and there are no rebuttals to consider.
- The statement about the *bodily harm rule* is acceptable, because it is supported by one defensible argument, a_2, and there are no counterarguments. Argument a_2 is defensible, because its single premise, about Article 302, is an undisputed presumption.
- The claim that the accused has inflicted bodily harm is acceptable, because it is supported by a defensible argument, a_3, and neither of the two counterarguments are defensible. The supporting argument, a_3, is defensible because its premise has been accepted.
- Argument a_4 is not defensible, because its premise, regarding the accused's testimony, in which he claims not to have been involved in a fight, is at issue and not acceptable.
- The accused's testimony is not acceptable for two reasons: 1) it is successfully countered by the argument a_6, with the testimony of 10 witnesses who claim to have seen the fight. (This testimony has been accepted with no further argument or evidence.) 2) It is not supported by at least one defensible pro argument, as required by the DV proof standard.
- The statement about broken ribs not being sufficient to amount to grievous bodily harm is not acceptable both because its only pro argument, a_9, is not defensible and also because its counterargument, a_7, is defensible. That is, the statement would not have met the *DV* proof standard even if its supporting argument had been defensible, since it is countered by a_7.
- The statement about several broken ribs with complications being grievous bodily harm is acceptable, because it is supported by a defensible argument, a_8, and has no counterarguments. The argument a_8 is defensible, because its only premise, about the second precedent, has been presumed and is not at issue.
- Finally, argument a_9 is not defensible, although it is supported by an undisputed premise, about the first precedent, because the lex specialis exception has been revealed (we assume) and accepted. Notice how lex specialis, which provides a reason to prefer precedent 2 over precedent 1, can be modeled even though our argumentation framework does not explicitly provide a way to order arguments.

One important function of an argumentation framework is to provide a basis for clear and comprehensible explanations or justifications of decisions. Argumentation framework which depend on a deep understanding of mathematics (e.g. fixed points) or formal logic (e.g. entailment from minimal subsets of hypotheses, as in some models of abduction) for justifying decision do not meet this requirement. We hope the Carneades system is sufficiently simple that explanations, such as the above, can be quickly appreciated and understood by people with no formal background in logic or mathematics.

5 Discussion

The idea of developing a computer model for managing support and justification relationships between propositions goes back to research on "truth" or reason maintenance systems in Artificial Intelligence [4, 16]. The first author's prior work on the Pleadings Game [11] included a formal model of dialectical graphs, for recording various kinds of support and defeat relationships among arguments. The concept of an *argumentation framework* was introduced by Henry Prakken [19] as part of a three-layered model for dialectical systems. As noted previously, Freeman and Farley [8] were the first to our knowledge to develop a computational model of burden of proof.

The Zeno Argumentation Framework [13] was based on Horst Rittel's Issue-Based Information System (IBIS) model of argumentation [20]. The Carneades Argumentation Framework, in contrast, uses mainstream argumentation theory as its starting point. Also, Zeno did not provide a foundation for modeling argument schemes with critical questions, and was not as well suited as the current system for modeling persuasion dialogs.

Verheij's work in [23] was the source of inspiration for distinguishing between different kinds of critical questions, which we have called presumptions and exceptions. Verheij's book, Virtual Arguments [24], includes an enlightening comparison of several theories of defeasible argumentation. Verheij compared them with regard to whether and, if so, how each system modeled 1) pro and con arguments; 2) warrants, in Toulmin's sense; 3) argument evaluation; and, finally 4) theory construction. We have already explained how our formal model handles the first three of these dimensions. In our model, the set of statements found to be acceptable can be viewed as a theory constructured collaboratively by participants in a dialog. Indeed, the first author, influenced by Fiedler [7], has long viewed reasoning explicitly as a theory construction process [9, 10] and was first attracted to argumentation theory precisely for this reason.

One key element of our theory construction approach is the idea of revealing hidden or implicit premises during a dialog. This approach was illustrated during the discussion of Toulmin and Pollock, for example, where warrants and undercutting defeaters where modelled as implicit presumptions revealed during dialog. Walton and Reed have done some recent work showing how argument schemes can be used to reveal implicit premises [27].

The formal model has been fully implemented, in a declarative way using a functional programming language, and tested on a number of examples from the Artificial Intelligence and Law literature, thus far yielding intuitively acceptable results. This validation work is continuing. More work is required to validate the models of the various proof standards, in particular the model of prepondernance of the evidence, which uses weights. For this purpose, we plan to reconstruct examples

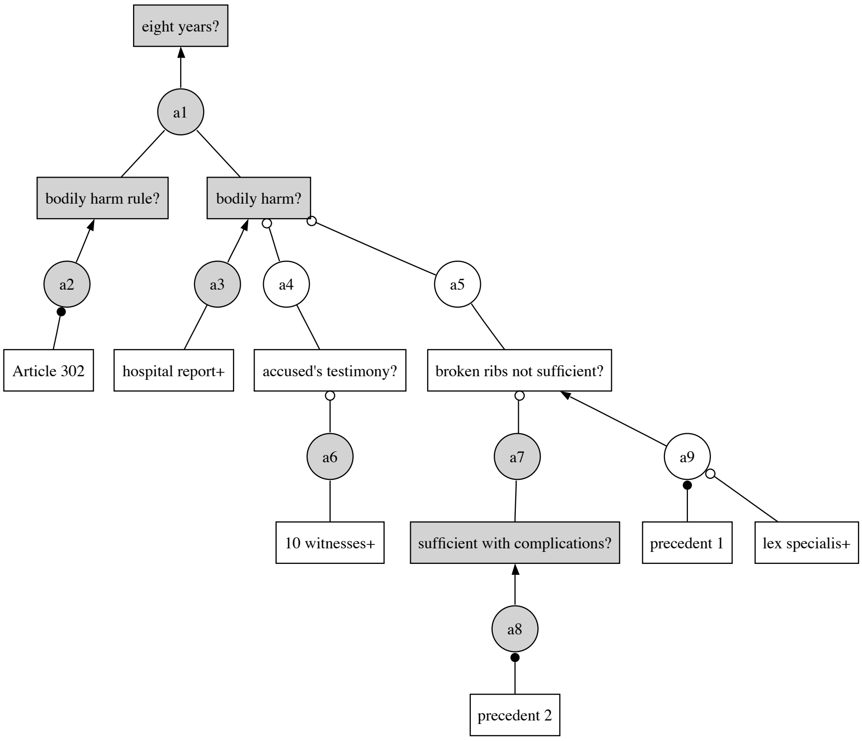


Figure 4. Reconstruction of Verheij's Grievous Bodily Harm Example

of reasoning with evidence. When completed, Carneades will support a range of argumentation use cases, including argument construction, evaluation and visualization. Although the focus of this paper was argument evaluation, it contains some hints about the direction we are heading to support argument visualization. One of our next tasks will be to refine the diagramming method used here to illustrate the argumentation framework.

REFERENCES

[1] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney, 'Arguing about cases as practical reasoning', in *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pp. 35–44, Bologna, Italy, (2005).

[2] Tim Berners-Lee, James Hendler, and Ora Lassila, 'The semantic web', *Scientific American*, **284**(5), 34–43, (May 2001).

[3] Floris Bex, Henry Prakken, Chris Reed, and Douglas Walton, 'Towards a formal account of reasoning with evidence: Argumentation schemes and generalizations', *Artificial Intelligence and Law*, **11**(2-3), (2003).

[4] Jon Doyle, 'A truth maintenance system', *Artificial Intelligence*, **12**, 231–272, (1979).

[5] Phan Minh Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artificial Intelligence*, **77**(2), 321–357, (1995).

[6] Paul Edwards, *The Encyclopedia of Philosophy*, volume 1, Macmillian and Free Press, 1972.

[7] Herbert Fiedler, 'Expert systems as a tool for drafting legal decisions', in *Logica, Informatica, Diritto*, eds., Antonio A. Martino and Fiorenza Socci Natali, 265–274, Consiglio Nazionale delle Richere, Florence, (1985).

[8] Kathleen Freeman and Arthur M. Farley, 'A model of argumentation and its application to legal reasoning', *Artificial Intelligence and Law*, **4**(3-4), 163–197, (1996).

[9] Thomas F. Gordon, 'The argument construction set — a constructive approach to legal expert systems', Technical report, German Research Institute for Mathematics and Data Processing (GMD), (1988).

[10] Thomas F. Gordon, 'A theory construction approach to legal document assembly', in *Expert Systems in Law*, ed., Antonio A. Martino, 211–225, Amsterdam, (1992).

[11] Thomas F. Gordon, 'The Pleadings Game – an exercise in computational dialectics', *Artificial Intelligence and Law*, **2**(4), 239–292, (1994).

[12] Thomas F. Gordon, 'A computational model of argument for legal reasoning support systems', in *Argumentation in Artificial Intelligence and Law*, eds., Paul E. Dunne and Trevor Bench-Capon, IAAIL Workshop Series, pp. 53–64. Wolf Legal Publishers, (2005).

[13] Thomas F. Gordon and Nikos Karacapilidis, 'The Zeno argumentation framework', in *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, 10–18, Melbourne, Australia, (1997).

[14] Jaap Hage, 'A theory of legal reasoning and a logic to match', *Artificial Intelligence and Law*, **4**(3-4), 199–273, (1996).

[15] Arthur C. Hastings, *A Reformulation of the Modes of Reasoning in Argumentation*, Ph.D. dissertation, Northwestern University, Evanston, Illinois, 1963.

[16] Johan de Kleer, 'An assumption-based TMS', *Artificial Intelligence*, **28**, (1986).

[17] Arno R. Lodder, *DiaLaw — On Legal Justification and Dialogical Model of Argumentation*, Springer, 1998.

[18] John L. Pollock, 'How to reason defeasibly', *Artificial Intelligence*, **57**, 1–42, (1992).

[19] Henry Prakken, 'From logic to dialectic in legal argument', in *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, 165–174, Maryland, (1995).

[20] Horst W.J. Rittel and Melvin M. Webber, 'Dilemmas in a general theory of planning', *Policy Science*, **4**, 155–169, (1973).

[21] Stephan E. Toulmin, *The Uses of Argument*, Cambridge University Press, 1958.

[22] Bart Verheij, *Rules, Reasons, Arguments. Formal Studies of Argumentation and Defeat*, Ph.d., Universiteit Maastricht, 1996.

[23] Bart Verheij, 'Dialectical argumentation with argumentation schemes: An approach to legal logic', *Artificial Intelligence and Law*, **11**(2-3), 167–195, (2003).

[24] Bart Verheij, *Virtual Arguments*, TMC Asser Press, The Hague, 2005.

[25] Douglas Walton, *Argumentation Methods for Artificial Intelligence in Law*, Springer, 2005.

[26] Douglas Walton and Thomas F. Gordon, 'Critical questions in computational models of legal argument', in *Argumentation in Artificial Intelligence and Law*, eds., Paul E. Dunne and Trevor Bench-Capon, IAAIL Workshop Series, pp. 103–111, Nijmegen, The Netherlands, (2005). Wolf Legal Publishers.

[27] Douglas Walton and Chris A. Reed, 'Argumentation schemes and enthymemes', *Synthese*, **145**, 339–370, (2005).

[28] Douglas N. Walton, *Argument Structure : a Pragmatic Theory*, Toronto studies in philosophy, University of Toronto Press, Toronto ; Buffalo, 1996. Douglas Walton. ill. ; 24 cm.

[29] Douglas N. Walton, *Argumentation Schemes for Presumptive Reasoning*, Erlbaum, 1996.

[30] Douglas N. Walton, *Appeal to Expert Opinion*, Penn State Press, University Park, 1997.

[31] Douglas N. Walton, *The New Dialectic: Conversational Contexts of Argument*, University of Toronto Press, Toronto; Buffalo, 1998. 24 cm.

Promises and Threats in Persuasion

Marco Guerini [1] and **Cristiano Castelfranchi** [2]

Abstract. In this paper we analyse Promises and Threats (P/T) use in persuasion. Starting from a general definition of P/T based on the concepts of speech act and social commitment we focus on Conditional Influencing P/T (CIP/T): those incentive-based P/T used to persuade the addressee, rooted on dependence and power relations. We argue that in CIP/T class the concepts of promise and threat are strictly connected: the promise act is necessarily accompanied by a threat act and vice versa. Thus we discuss the problem of why the CIP/T are credible even if the speaker is supposed to be a rational agent and analyse some asymmetries between CIP and CIT. We also identify - beyond the rhetorical presentation - a deeper difference between substantial promises and substantial threats. Throughout the article is given a pre-formal model of these concepts.

1 INTRODUCTION

In this paper (based on a bigger research on P/T [8]) the concepts of promises and threats are analysed in order to gain some insight on their nature and their relations. The aim is to study P/T use in persuasion.

Starting from the concepts of speech act and social commitment we briefly show that not all P/T are for persuasion or conditional in their nature (like in "*if you do your homework I will bring you to the cinema*"): four different typologies of P/T are possible.

We then focus on Conditional Influencing P/T (CIP/T): those P/T used to persuade the addressee. In our analysis CIP/T are incentive-based influencing actions, rooted on dependence and power relations. These communicative actions affect the practical reasoning of the receiver by adding "artificial" consequences to the required action.

Finally we argue that in CIP/T class the concepts of promise and threat are two faces of the same coin. The deep logical form of these social acts is an IFF: the promise act is always and necessarily accompanied by a threat act ("*if you do not do your homework I will not bring you to the cinema*"), and vice versa.

Thus we discuss the problem of why the CIP/T are credible even if the speaker is supposed to be a rational agent and analyse some asymmetries between CIP and CIT. We also identify - beyond the rhetorical presentation - a deeper difference: a substantial threat, consisting in a choice between two losses, compared with substantial promises where the choice is between a gain and a missed-gain.

Throughout the article is given a pre-formal model for a computational treatment of these concepts. We adopt the Beliefs, Desires, Intentions (BDI) model as a reference framework [9, 10]. In the context of negotiating agents some simplified formalizations of CIP/T has been put forward, see for example [16, 1, 23]. Still, here we will focus on the implicit negotiational nature of CIP/T and not on their use in negotiation.

Hereafter variable x indicates the sender, and variable y the receiver, of the message.

2 PROMISES AND THREATS

2.1 What is a 'promise'

A Promise is, from a general point of view, a speech act that consists in the declaration, by x, of the *intention* of performing a certain action ax, under the pre-condition that ax is something wanted by y, with the aim of entering into an obligation (*social commitment*) of doing ax [20, 2, 22, 18]. A similar definition can be also found in the Webster Dictionary.

Intention = the notion of internal-commitment (intention) as defined by Bouron [3] establishes a relation between two entities: the agent x and the action ax.

$$INTEND(x\ ax) = GOAL(x\ DOES(x\ ax)) \qquad (1)$$

This formula defines the intention of x to perform ax as the goal of x to perform the action in the next time interval (for a thorough definition see [10]).

Social commitment = the notion of social commitment (S-commitment) [5] involves four entities: the agent x, the action ax (that x has the intention to perform, for which he takes the responsibility), the agent y for which action ax has some value, and an agent z before whom x is committed (the witness).

$$S - COMMITED(x\ y\ ax\ z) \qquad (2)$$

In the definition of S-commitment the key point is that x is committed to do ax because y is interested in ax. So a S-commitment is a form of goal adoption[3], and P/T are a particular form of social commitment.

When x promises something (ax) to y she is committing herself to do ax. This is not simply an internal commitment that stabilize x's choices and actions [4], and it is not simply a 'declaration of a personal intention'. In intention declaration x is committed about the action only with herself and she can change her mind. Instead in

[1] Itc-Irst, Istituto per la Ricerca Scientifica e Tecnologica, 38050 - Trento, ITALY, email: guerini@itc.it

[2] National Research Council - ISTC - Institute of Cognitive Sciences and Technologies via San Martino della Battaglia 44, 00185 - Roma, ITALY, email: c.castelfranchi@istc.cnr.it

[3] By '(Social) Goal-Adoption' we mean the fact that x comes to have a goal because and until she believes that it is a goal of y. x has the goal to 'help' y, or better (since 'help' is just a sub-case of social goal-adoption) x has the goal that y realizes/obtains his goal $GOAL(y\ p)$, thus decides to act for y by generating $GOAL(x\ p)$. This can be for various motives and reasons: personal advantages (like in exchange), cooperation (common higher goals), altruism, norms, etc. [11].

promises she is committed with the other, x has an interpersonal obligation - $OBL(x\ y\ DOES(x\ ax))$ - and creates some ‘rights’ in the other (entitled expectation & reliance/delegation, checking, claiming, protesting).

Moreover, being sincere in promising (i.e. being internally committed) is not necessary for a P/T to be effective. This commitment has an interpersonal and non-internal nature, there is a real created and assumed ‘obligation’ (see also [24]).

Let us better represent these features of a Promise:

a) x declare to y his intention to do ax

$$UTTER(x\ y\ INTEND(x\ ax)) \tag{3}$$

b) that is assumed to be in y’s interest and as y likes,

$$GOAL(y\ DOES(x\ ax)) \tag{4}$$

c) in order that y believes and expects so

$$BEL(y\ INTEND(x\ ax)) \tag{5}$$

d) and y believes also that x takes a commitment to y, an obligation to y to do as promised.

$$BEL(y\ S-COMMITED(x\ y\ ax)) \tag{6}$$

e) The result of a promise is y’s belief about ax, the public ‘adoption’ by x of a goal of y, y’s right and x’s duty about x doing ax.

$$BEL(y\ DOES(x\ ax)) \tag{7}$$

Finally, a promise presupposes the (tacit) agreement of y to be effective, i.e. to create the obligation/right. It is not complete and valid, for example, if y refuses (see section 2.4).

2.2 What is a ‘threat’ and P/T asymmetry in commitments

A threat is, from a general point of view, the declaration, by x, of the intention of performing a certain action ax, under the pre-condition that ax is something not wanted by y. Analytically, the situation is similar to promises apart from:

b1) ax is assumed to be against y’s interest and what y dislikes,

$$GOAL(y\ \neg DOES(x\ ax)) \tag{8}$$

d1) x takes a commitment, an obligation to y to do as threaten.

In the threatening case, ax is something y dislikes (b1), and the consent or agreement of y is neither presupposed nor required. It is important to note that it is not strictly necessary that conditions (b) and (b1) hold before the P/T utterance. It is sufficient that ax is wanted (or not wanted) after that the P/T is uttered: P/T can be based on the elicitation or activation of a non-active goal of y[4].

P creates an obligation of x toward y, and corresponding rights of y about x’s promised action. But this looks counter intuitive for T cases where ax is something y does not want[5]. To find an answer, we have to differentiate the two S-commitments that P creates.

S1) A S-commitment about the **truth** of what x is declaring (he takes responsibility for this) and this is the kernel of ‘promising’

S2) A S-commitment on a future event under x’s control. This is about the action that x has to accomplish in order to **make true** what he has declared.

In T the first commitment (S1) is there: y can blame and make fun of x for not keeping his word on what threatened: the reputation of x is compromised. But for the second more important social-commitment to do ax, there is an important asymmetry between P and T (conditions (d) and (d1)) that we will adjust in section 4.3.

2.3 Promises as public goal adoption

Our analysis, so far, basically converges with Searle’s one, but in our view Searle missed the “adoption” condition, which is entailed by the notion of S-commitment (condition (d)). In order to have a promise, it is not enough (as seems compatible with his 4th condition and not well expressed in his 5th condition) that:

- x declares (informs y) to have a give intention to do action ax - condition (a) of our analysis
- x and y believe that y likes (prefers) that x does such an action - condition (b) of our analysis.

This is not a promise. For example:

E1) for his own personal reasons x has to leave, and informs y of his intention, and he knows that y will be happy for this; but this is not a ‘promise’ to y, since x do not intend to leave because y desires so.

While promising something to y, x is adopting a goal/desire of y. x intends to do the action since and until she believes that it is a goal for y; x’s intention is “relativized” to this belief (see formula below).

$$REL-GOAL(x\ DOES(x\ ax)GOAL(y\ DOES(x\ ax))) \tag{9}$$

2.4 Y’s agreement

The commitment, and the following ‘obligations’, of x to do ax is relativised to ax being a goal of y. So, for a felicitous promise the (tacit) acceptance of y is crucial; it is this (tacit) agreement that actually creates the obligation and the obligation vanishes if y does no (longer) desires/requires ax (condition (b)). This analysis is also valid for the threatening case, but in a reverse sense: the consent/acceptance is presupposed not to be given. The paradoxical joke of the sadist and the masochist, in example E2, points out clearly this case:

E2) Sadist: “*I will spank you!*” Masochist: “*Yes please!*” Sadist: “*No*”

But y, in declaring she does not want x to perform ax, is not necessarily negating her need for ax: there are different reasons that can bring y to reject x’s help (e.g. not to feel in debt).

[4] We thank Andrew Ortony for suggesting us to make this explicit and clear. On goal-activation see [6].

[5] One might also claim - for the sake of uniformity and simplicity - that in fact there are such a ‘right’ for y and such an obligation for x, but y will never exercise his rights and claim for them. One might support the argument with the example of the masochist (E2): if pain is a pleasure for y he can expect for x’s ‘promised’ bad action, and can in fact claim for it, since x has committed himself on it.

2.5 The notion of persuasion

There is a strong relation between P/T and persuasion; P/T are often used as persuasive means. We think there is a lack of theory on their relation. To analyse it we need a theory of persuasion (some preliminary ideas can be found in [15, 14]).

According to Perelman [19], persuasion is a skill that human beings use in order to make their partners perform certain actions or collaborate in various activities, see also [17]. This is done by modifying - through communication (arguments) - the other's intentional attitudes. In fact, apart from physical coercion and the exploitation of stimulus-response mechanisms, the only way to make someone do something is to change his beliefs [6].

We propose two different formalizations of "goal of persuading" (formulae 10 and 11). Formula 10 implies formula 11 when y is an autonomous agent (i.e. every action performed by an agent follows from an intention).

$$PERSUADE(x\ y\ ay) \rightarrow INTEND(x\ DOES(y\ ay)) \quad (10)$$

$$PERSUADE(x\ y\ ay) \rightarrow INTEND(x\ INTEND(y\ ay)) \quad (11)$$

Considering formula 11, in persuasion the speaker presupposes that the receiver is not already performing or planning the required action ay. In a more strict definition it can also be presupposed that the receiver has some *barriers* against ay: y wouldn't spontaneously intend to do so. Persuasion is then concerned with finding means to overcome these barriers by conveying the appropriate beliefs to y.

The relation between persuasion and dissuasion is non-trivial, though, here we will consider dissuasion as persuasion to not perform a given action.

$$DISSUADE(x\ y\ ay) \rightarrow PERSUADE(x\ y\ ay) \quad (12)$$

In analyzing the notion of 'intention', three cases must be considered. The intention of performing ay (formula 13), the intention of not performing ay (formula 14), and the lack of intention (formula 15).

$$INTEND(y\ ay) \quad (13)$$

$$INTEND(y\ \neg ay) \quad (14)$$

$$\neg INTEND(y\ ay) \quad (15)$$

Following the definitions from 13 to 15 we can model two different notions of persuasion and dissuasion:

- the *weak notion* captures the idea that the receiver is not already planning to perform the required action (formula 15);
- the *strong notion*, captures not only the idea that y is not already planning to perform ay, but also that he has some specific barriers against the action (y has some reason for not doing ay).

The terms "barriers/reasons" indicate those dispositions - of the receiver - that are against ay. In our approach barriers are modelled as contrary intentions: for any given action ay, the contrary intention is the intention of performing $\neg ay$ (formula 14). P/T, when used as persuasive means, refer only to the strong cases of persuasion (see section 3.3).

2.6 The main classes of P/T

There are four main classes of promises and threats. The distinction can be made along two dimensions: (a) presence of a conditional part in the P/T message, (b) presence of a persuasive aim in x (see table 1).

a) Some promises are conditional in their nature (e.g. "*If tomorrow is sunny I will bring you to the zoo*", "*If you do your homework I will bring you to the cinema*"). This dimension refers to the presence or the absence of a conditional part in the message

b) The second dimension refers to the presence or the absence, in the speaker of the intention to influence the hearer. If the predicate $PERSUADE(x\ y\ ay)$ holds, we are in the influencing class. This dimension is the most important in the division of P/T. In this paper we will focus on conditional-influencing class, central from a persuasive perspective.

	INFLUENCING	NON-INFLUENCING
CONDITIONAL	"*If ay then ax*" (CIP/T)	"*If c then ax*" (CP/T)
NON-CONDITIONAL	"*I will ax*" (IP/T)	"*I will ax*" (P/T)

Table 1. Main classes of promises and threats

3 THE INFLUENCING CLASSES

3.1 General Structure

The key question is: why should x perform an action positive or negative for y? And why x should want to communicate this to y?

This is done exactly with the aim of inducing y to perform (not to perform) some other action (ay). This is obtained by artificially linking a new effect (ax) to the action ay. This is the very nature of Influencing P/T (IP/T).

The two classes of IP/T can be considered both as conditional, because this is entailed by the influencing nature of IP/T, and we will refer to both as CIP/T. In non conditional cases, simply, x leaves implicit the conditional part for pragmatic reasons. The structure of the utterance is:

"*If ay then ax*"

In CIP/T structure, the condition of the utterance ("*if ay*") is equal to the achievement or avoidance goal of the act.

- In P the condition expresses what y has to 'adopt'. x is proposing an 'exchange' of reciprocal 'adoption': "*if you adopt my goal (ay) I will adopt your goal (ax)*".
- In T the condition is what x wants to avoid and he is prospecting a 'reciprocation' of damages: "*if you do what I dislike (ay), I will harm you (ax)*".

Generically, a CIP has a higher goal that ay, and the message is aimed at this goal. More precisely: when x utters the sentence, he has the goal that y believes that x is going to favour him ($G1$) with the super-goal ($G2$) to induce in y the intention to do ay. Finally $G2$ has another super-goal ($G3$) to induce y to perform ay (which is the ultimate goal of CIP/T). The cognitive structure is depicted in figure 1.

A CIT has the same structure, except that the influencing goals ($G2$ and $G3$) are the opposite of the condition of the utterance: $\neg ay$ and ay (for additional important differences in the plan, see section 2.7). The distinction between goals $G2$ and $G3$ is motivated by the two definitions of PERSUADE: to induce someone to act (formula 11), by creating the corresponding intention (formula 10). This distinction is necessary in those cases where CIP/T are used only to create an intention, as in example E3.

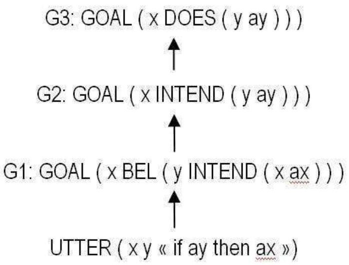

G3: GOAL (x DOES (y ay)))
G2: GOAL (x INTEND (y ay)))
G1: GOAL (x BEL (y INTEND (x ax)))
UTTER (x y « if ay then ax »)

Figure 1. The goal structure of a CIP speech-act

E3) x, a lackey of a Mafia boss, promises to y, another lackey of the boss, to give him a huge money reward (ax) if he kills the boss (ay). But x wants to show to the boss that y is not loyal. The overall goal of his promise is just that y intends to kill the boss ($G2$), and not that he actually does it ($G3$).

3.2 The relation between persuasion/dissuasion and IP/T

In common sense, promises are for persuading and threats are for dissuading (see for example [12, 25]), but this is not true. The complete spectrum is depicted in table 2 ("+" means a benefit for y, "-" means a disadvantage).

	A. Persuading PP: $\neg INTEND(y\ ay)$ Gx: $INTEND(y\ ay)$	**B. Dissuading** PP: $INTEND(y\ ay)$ Gx: $\neg INTEND(y\ ay)$
1. Promise: y prefers ax	*"If ay then ax+"* (CIP/T)	*"If not ay then ax+"* (CP/T)
2. Threat: y prefers $\neg ax$	*"If not ay then ax -"* (IP/T)	*"If ay then ax -"* (P/T)

Table 2. The relation between Persuasion/Dissuasion and IP/T

In 1A and 1B, x is meaning: "*if you change your mind, I will give you a prize*"; i.e. the condition of the CIP is the opposite of the presupposition. While in 2A and 2B x is meaning: "*if you persist, do not change your mind, I will punish you*"; i.e. the condition of the CIT coincides with the presupposition.

3.3 CIP/T as "commissive requests"

Using Searle's terminology, CIP/T represent a *request* speech act by means of a *commissive* [22]. A set-based description of the various classes is given in figure 2.

There are different communicative acts (like "asking for", argumenting) with different "costs" that can be used to persuade. CIP/T are the most "expensive". In fact, given that every action has a cost, if y carries out ay, then x is committed to carry out ax (on this, see section 3.6). Why not simply asking for ay, or argumenting on the advantages, for y, to perform ay? If successful, x does not have any additional cost.

The answer relies on the necessity (following x) of using rewards (defined as "incentives", see section 3.4) and on the different presuppositions that lead to different persuasive acts.

1. In a simple request (lowest cost for x) y is presupposed to have no contrary intentions on ay (or that y's internal reward - like satisfaction, reciprocation - may suffice for overcoming y's barriers)

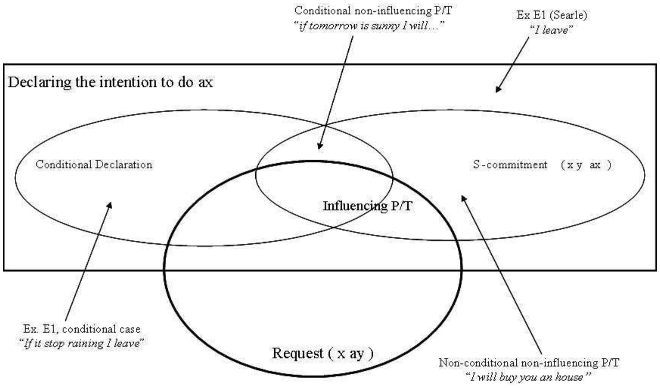

Conditional non-influencing P/T
"if tomorrow is sunny I will..."
Ex E1 (Searle)
"I leave"
Declaring the intention to do ax
Conditional Declaration
S-commitment (x y ax)
Influencing P/T
Ex. E1, conditional case
"If it stop raining I leave"
Request (x ay)
Non-conditional non-influencing P/T
"I will buy you an house"

Figure 2. A set based description of the various classes of P/T and related concepts

2. In argumenting the presupposition is that, even if y can have some contrary intentions, when he will know all the outcomes of ay he will perform it.
3. In P/T (highest cost for x) instead the presupposition is not only that y has some contrary intentions, but also that there is no purely argumentative way to make him change his mind.

So, an influencing promise is a sort of combination between two different (linguistic) acts, an *offer* (*commissive* offer) of ax and a *request* for *ay*. In particular the offer is conditioned to the request.

3.4 Artificial consequences and incentives

In argumentation x can persuade y by prospecting "natural" positive or negative consequences of ay. But in CIP/T x has additional ways to persuade y to do ay:

- through the prospect of positive outcomes (whose acquisition is preferable) due to x's intervention (ax), not natural consequence of ay
- through the prospect of negative outcomes (whose avoidance is preferable) due to x's intervention (ax), not natural consequence of ay[6].

In CIP/T outcomes are linked to ay in an artificial way: "artificial" means that the consequence is under the control (direct or indirect) of x and will not happen without his intervention. With CIP/T arguments are "built" and not "found". This definition includes also the case in which ax is performed by a third, delegated, agent z. The fact is that this third agent will perform ax only if requested, and because delegated, by x. Let us consider the following examples:

E4) y's schoolmate: "*if you finish your homework your mother will bring you to the cinema*"
E5) y's mother: "*if you finish your homework I will tell your aunt to bring you to the cinema*"

These two examples show that being *natural* or *artificial* is strictly context dependent and the presence of an agent in the delivering of the outcome does not discriminate the two cases. In example E4 the same consequence of E5 (to be bring to the cinema) is used by the

[6] It is important to remark that 'not doing a' is an action (when is the output of a decision). Thus x can induce y to not doing something.

speaker in an argumentative way, by making the other believe or consider some benefits coming from her own action.

We consider CIP/T as social acts based on the prospect of incentives, where "incentives" are precisely those artificial consequences that are delivered - by x to y - in order to influence y. These incentives can be positive (*prizes*) or negative (*punishments*). In particular:

a) If ax is something given because is wanted by y, then it is a prize:

$$GOAL(y\ ax) \rightarrow PRIZE(ax) \tag{16}$$

b) If ax is something given because is not wanted by y, then it is a punishment:

$$GOAL(y\ \neg ax) \rightarrow PUNISHMENT(ax) \tag{17}$$

In table 3, we have a summary of the different typologies of outcomes of ay with the corresponding term to indicate them (similar to the distinction proposed in [12] between *conditionals inducements* and *conditional advices* classes). Incentives, promises and threats are on line B; prospected natural outcomes, instead, are on line A.

	POSITIVE OUTCOMES	NEGATIVE OUTCOMES
A. Natural Consequences	Advantages	Disadvantages/ Drawbacks
B. Artificial Consequences	**Prizes**	**Punishments**

Table 3. Different typologies of ay outcomes

3.5 Credibility, preferability pre-conditions and the power of x

Many pre-conditions of the P/T act have to be met in order to have a felicitous communication: a P/T must be *credible* and convincing (*preferable*).

1) *Credibility pre-conditions*: The fact that the loss or gain for y is due to x's decision and intervention, explains why, in order to have a "credible" promise or threat, it is crucial that y believes that x is in condition to favour or to damage her. Thus when x announces his promise or threat he also has the goal that y believes that x has the "power of" ax; this belief is y's "trust" in x and it can be based on x reputation, on previous experience, on some demonstration of power, etc.[7]

Thus in order to have true promises or threats, x must have some power over y; the power of providing to y incentives (or at least y must believe so). More analytically:

- x has some *power of* doing ax

$$CAN - DO(x\ ax) \tag{18}$$

- y depends on x, and more precisely on his action ax, as for achieving some goal Gy;

$$DOES(x\ ax) \rightarrow Gy \tag{19}$$

$$DEPEND(y\ x\ ax\ Gy) \tag{20}$$

This means that:

- x gets a *power over* y's goal Gy, the power of giving incentives or not to y by the realization of Gy;

$$POWER - OVER(x\ y\ Gy) \tag{21}$$

- both x and y believe so[8];

$$BMB(x\ y\ POWER - OVER(x\ y\ Gy)) \tag{22}$$

on such a basis:

- x gets a power of influencing y to do ay while using the promise of Gy (performing ax) as an incentive [9].

$$PERSUADE(x\ y\ ay) \tag{23}$$

$$PRIZE(ax) \tag{24}$$

That is, x can make y believe that "if y performs ay (adopts the goal of x) then x will reward her by performing ax (adopting y's goal)".

2) *Preferability pre-conditions*: The above conditions represent the applicability conditions for P/T, but there is still another condition to be met in order to make CIP/T effective:

- If x has the power to jeopardise (or to help achieve) a goal Gy of y, and the goal has a higher value than the value of the action (ay), then x can threaten y to jeopardise the goal if he does not perform ay (or promise to help him realise his goal if he performs ay).

$$V(Gy) > V(ay) \tag{25}$$

Preferability conditions regard only the effectiveness of the message. "*If you carry that heavy bag for five kilometres I will give you 20 cents*": this is a true and credible promise, but ineffective (not preferable), because x has the *power of* giving 20 cents to y but the value of ay (carrying the heavy bag for five kilometres) is much greater the value of Gy (gaining 20 cents).

3.6 Scelling's plan asymmetry and inefficacy paradox in CIP/T

Plan asymmetry: in order to be efficacious the promised or threatened action ax must have an higher value than the requested action ay (in y's perspective)[10]: $V(ax) > V(ay)$. On the other side (in x's perspective), the promised action ax (that is: x's cost) has to have less value than ay: $V(ax) < V(ay)$. It represents x's costs. However, there is an asymmetry between P and T under this respect (considering those P/T where ax is an action to be performed and not the abstaining from an action).

- In Promises, x - if sincere - plans (intends) to do ax in order to obtain ay. In case of a successful P it is expected that x performs ax.
- In Threats, x plans the *non* execution of ax. It should be executed only in case of failure and y's refusal[11].

[7] This is why a mafia's warning is not usually limited to a simple (verbal) message, but is a concrete harm (beating, burning, etc.). This is a 'demonstrative' act (that is communication) but with the advantage to directly show and make credible the threatening power of the speaker [7]. On the use of fear and scare tactics in threats see also [26].

[8] We do not address here the problem of false P/T, like in the case of an armed robbery with a fake gun.

[9] The power of influencing y to do something can based not only on incentive power, but also on imitation, reactive elicitation, normative endowment, etc.

[10] $V(ax)$ for y is equivalent to $V(Gy)$ since $ax \rightarrow Gy$

[11] This is the genial intuition of Schelling [21] (p.36, especially note 7, p. 123) but within an not enough sophisticated theory of P/T.

This difference is especially important in substantial P vs. substantial T (see later). Under this respect a T looks more convenient than a P: a successful T has only communication/negotiation costs.

Though, there are serious limits in this 'convenience', not only from the point of view of social capital and collective interest, but also from x's point of view. In fact in those kinds of relationships y is leaning to exit from the relation, to subtract herself from x (bad) power and influence. It requires a lot of control and repression activity for maintaining people under subjection and blackmail.

Inefficacy paradox: in threats, ax (detrimental for y) should be executed only in case of failure/inefficacy of the threat, but why x should perform it and having useless costs? [21]. Surely not for achieving the original goal - $DOES(y\ ay)$ -. Thus, it seems irrational to do what has been threatened.

Moreover, that this action would be useless for x should be clear also to y, and this makes x's threat non credible at all: y knows that x (if rational) will not do as threaten if unsuccessful; so why accepting?

Analogously, the promised action (beneficial for y) usually[12] has to be performed by x in case of success, so why should x spend his resources when he already obtained his goal? But this is known by y and should make x's promise not very credible.

As Shelling suggests, threats (and promises) should be performable in steps: the first steps are behavioural messages, demonstration of the real power of x, warnings or "lessons". However, this is just a sub-case; the general solution of this paradox has to be found in *additional and different reasons and motives of* x.

Let's consider threats. In keeping threats after a failure, x aims at giving a "lesson" to y, at making y learning (for future interactions with x or with others) that (x's) threats are credible. This can be aimed also at maintaining the reputation of x as a coherent and credible person. Another motive can be just rage and the desire of punishing y; TIT for TAT. In keeping promises after success - a part from investing in reputation capital - there might be 'reciprocation' motives, or fairness, or morality, etc.

If these additional motives are known by y, they make x's P/T credible; but it is important to have clarified that:

- if x performs what he promised it is **not** in order to obtain what he asked for.

4 THE JANUS NATURE OF CIP/T

4.1 Logical form of CIP/T

No P/T of the form "*if ay I will ax*" would be effective if it does not also mean "*if not ay I will not ax*", that is: if it would mean "*if ay I will* ax*, and also if not ay*". x can either plan for persuading y to ay or for dissuading y from not ay. He can say: "*if ay I will give you a positive incentive*" (promise) or "*if not ay I will give you a negative incentive*" (threat).

In these cases, one act is only the implicit counterpart of the other and the positive and negative incentives are simply one the negation of the other ("*I will do ax*" vs. "*I will not do ax*"). Also for this reason, one side can remain implicit. A threat is aimed at inducing an avoidance goal, while a promise is aimed at eliciting attraction, but they co-occur in the same influencing act[13]. Though the two P/T are not an identical act they are two necessary and complementary parts of the same communicative plan.

Despite the surface IF-THEN form of CIP/T, our claim is that the deep logical form is an IFF[14]. There is no threat without promise and vice versa. In the (intuitive) equivalence between: "*if you do your homework I will bring you to the cinema*" (promise) and "*if you do not do your homework I will not bring you to the cinema*" (threat), the logical IF-THEN interpretation doesn't work:

$$(ay \rightarrow ax) \neq (\neg ay \rightarrow \neg ax) \tag{26}$$

while this is the case for the IFF interpretation:

$$(ay \leftrightarrow ax) = (\neg ay \leftrightarrow \neg ax) \tag{27}$$

4.2 Deep and surface CIP/T

Only a pragmatic difference seems to distinguish between P and T as two faces of the same act (here we will not address the problem of how x decides which face to show). However, common sense and language have the intuition of something deeper. What is missed is an additional dimension, where promises refer to real gains, while threats refer to losses and aggression. We need to divide CIP/T along two orthogonal dimensions: the deep and surface one.

1. The deep (substantial) dimension regards the "gain" and "losses" for the receiver related to speaker's action.

   Gain: the fact that one realizes a goal that he does not already have, passing from the state of $Goal\ \ p\ \ \&\ \ not\ \ p$, to the state that $Goal\ \ p\ \ \&\ \ p$ (the realization of an 'achievement' goal in Cohen-Levesque terminology); in this case the welfare of the agent is increased.
   Losses: the fact that one already has p and has the goal to continue to have p ('maintenance' goals in Cohen-Levesque terminology); in case of losses one passes from having p - as desired - to no longer having p; in this case the welfare of the agent is decreased.

2. The surface dimension regards the linguistic form of the CIP/T: the use of the P or T face.

In table 4, on the columns we have losses and gains (with regard to ax in y's perspective). These two columns represent:

- deep threatening (loss): a choice between two losses ("harm or costs?" no gain),
- deep promises (gain): a choice between a gain (greater then the cost) or a missed gain.

On the rows we have the surface form of the corresponding communicative acts: in the case of surface promise what is promised is a missing loss or a gain, while in the case of surface threat what is promised is a loss or a missing gain. The distinction (for a same deep structure) is granted by the IFF form of CIP/T.

What is explained in table 4 is the general framework, but, for example we must distinguish "defensive" promises/threats (defensive from x's perspective: x does not want ay and uses ax to stop y) from "aggressive" ones (in which ay is something wanted by x).

[12] There are promises of this form: "*I will do ax if you promise to do ay*". In this case the promised action ax has to be performed before ay. In such conditions there is no reason for x to defeat.

[13] It is also possible to have independent and additional positive and negative incentives, in a strange form of double Threat-Promise act like the following one: "*If you do your homework I will bring you to movie; if you do not do your homework I will spank you*".

[14] We mean that the correct logical representation of the intended and understood meaning of the sentence is an IFF. One can arrive to this either via a pragmatic implicature [13] or via a context dependent specialized lexical meaning (see later).

	Deep T: Loss (scenario A)	**Deep P: Gain** (scenario B)
Surface Promise	If ay then *not-loss* "*If you do the homework I will not spank you*"	If ay then $gain$ "*If you do the homework I will bring you to the cinema*"
Surface Threat	If *not-ay* then $loss$ "*If you do not do the homework I will spank you*"	If *not-ay* then *not-gain* "*If you do not do the homework I will not bring you to the cinema*"

Table 4. Deep and surface P and T

4.3 CIP/T and their commitments

The analysis just introduced on the logical structure of CIP/T allows us, now, to define the different kinds of commitments entailed by promises and threats (points d and d1 of our analysis, see section 1.3). As we already saw (section 2.2 and note 5) apparently, threats seem to fall out of our analysis in terms of S-commitment. In threats the committed action is not, superficially, a y's goal. If x does not keep his commitment, y won't protest. But, given that every threat entails a promise - at least for CIP/T - the asymmetry can be solved: the S-commitment in threats is taken on the corresponding promise form. So:

- Promise: $(COMMITTED\ x\ y\ \underline{ax}\ z)$ where ax is "I will bring you to the cinema"
- Threat: $(COMMITTED\ x\ y\ \underline{\neg ax}\ z)$ where ax is "I will spank you"

In the first case y can protest if x does not perform the action, in the second, instead, y can protest if x performs the action[15].

But the commitment structure of CIP vs. CIT is even more complex: we need the concept of "Pact" - or "Mutual S-commitment" - in which the commitment of x with y is conditioned to the commitment of y with x and vice versa. In fact any P presupposes the 'agreement' of y (see section 2.4), a tacit or explicit consent, or a previous request by y. This means that y takes a S-Commitment toward x to accept his 'help' and to rely on his action [5]. x will protest (and is entitled to) if y solves the problem on his own or ask someone else.

In our view an accomplished promise is a Multi-Agent act, it requires two acts, two messages and outputs with two commitments. It seems necessary to go - thank to the notion of conditional reciprocal goal-adoption - beyond the enlightening notion of Reinach [20] (cited and discussed in [18]) of 'social act' as an act which is etherodirected, that needs the listening and "grasping" of the addressee.

Moreover, there's the need of a distinction between "negative pacts" (based on threats) and "positive pacts" (based on promises), they entail different S-commitments.

- In CIP x proposes to y to 'adopt' her goal (ax) if y adopts his own goal (ay); he proposes a **reciprocal goal-adoption**, and exchange of favors.
- In prototypical CIT we have the complementary face. x is proposing to y an **exchange of abstentions from harm** and disturb. The reciprocal S-commitments are formulated and motivated by avoidance, in both x and y.

5 CONCLUSIONS

In this paper we analysed the persuasive use of Promises and Threats. Starting from the definition of P/T as "speech acts creating social-commitments" and the definition of persuasive goal, we showed that not all P/T are for persuasion or conditional in their nature.

We then focused exactly on those conditional P/T that are intended to influence - persuade - the addressee (CIP/T). In our analysis CIP/T are incentive-based influencing actions for overcoming y's resistance to influence; they are based on x's power over y's goals.

We claimed that in CIP/T class the concepts of promise and threat are two faces of the same coin: a promise act is always and necessarily accompanied by an act of threat, and vice versa.

We also identified - beyond the rhetorical presentation - a deeper difference: a substantial threat and a substantial promise (independent of the presented 'face'). A plan asymmetry between P/T and a paradox of CIP/T, that should be non-credible in principle, were also introduced.

The aim of this work was to give a pre-formal model of P/T as a basis for a computational treatment of these concepts.

ACKNOWLEDGEMENTS

This work was partially supported by the HUMAINE project.

We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] L. Amgoud and H. Prade, 'Threat, reward and explanatory arguments: generation and evaluation', in *Proceedings of the ECAI Workshop on Computational Models of Natural Argument*, Valencia, Spain, (August, 2004).

[2] J.L. Austin, *How to do things with words*, University Press, Oxford, 1962.

[3] T. Bouron, *Structures de Communication et dOrganisation pour la Coopration dans un Univers Multi-agent*, Ph.D. dissertation, Universit Paris 6, 1992.

[4] M. Bratman, *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, Mass., 1987.

[5] C. Castelfranchi, 'Commitments: from individual intentions to group and organizations', in *Proceedings of the First International Conference on Multi-Agent Systems*, S. Francisco, (1995).

[6] C. Castelfranchi, 'Reasons: Beliefs structure and goal dynamics', *Mathware & Soft Computing*, **3(2)**, 233–247, (1996).

[7] C. Castelfranchi, 'Silent agents. from observation to tacit communication.', in *Proceedings of the workshop 'Modeling Other Agents from Observations' (MOO 2004).*, NY, USA, (2004).

[8] C. Castelfranchi and M. Guerini, 'Is it a promise or a threat?', Technical report, ITC-Irst Technical report T06-01-01, (January 2006).

[9] P. R. Cohen and H. J. Levesque, *Intentions in Communication*, chapter Rational interaction as the basis for communication, 221–256, The MIT Press, Cambridge, MA, 1990.

[10] P. R. Cohen and H. J. Levesque, *Intentions in Communication*, chapter Persistence, Intention, and Commitment, 33–69, MIT Press, Cambridge, MA, 1990.

[11] R. Conte and C. Castelfranchi, *Cognitive and Social Action*, UCL Press, London, cognitive and social action edn., 1995.

[12] J. St. B. T. Evans, 'The social and communicative function of conditional statements', *Mind & Society*, **4(1)**, 97–113, (2005).

[13] H. P. Grice, *Speech Acts*, chapter Logic and conversation, 4158, New York: Academic Press, 1975.

[14] M. Guerini, *Persuasion models for multimodal message generation*, Ph.D. dissertation, University of Trento., 2006.

[15] M. Guerini, O. Stock, and M. Zancanaro, 'Persuasion models for intelligent interfaces', in *Proceedings of the IJCAI Workshop on Computational Models of Natural Argument*, Acapulco, Mexico, (2003).

[15] Even from a threatening point of view is counterproductive for x not to respect the "promise" after a successful threat. In fact x would be perceived as unfair if she were to spank the kid after he did his homework.

[16] S. Kraus, K. Sycara, and A. Evenchik, 'Reaching agreements trough argumentation: a logic model and implementation', *Artificial Intelligence Journal*, **104**, 1–69, (1998).

[17] B. Moulin, H. Irandoust, M. Belanger, and G. Desordes, 'Explanation and argumentation capabilities: Towards the creation of more persuasive agents', *Artificial Intelligence Review*, **17**, 169–222, (2002).

[18] K. Mulligan, *Speech Act and Sachverhalt*, chapter Promises and others social acts: Constituents and Structure, 29–90, Dordrecht, 1987.

[19] C. Perelman and L. Olbrechts-Tyteca, *The new Rhetoric: a treatise on Argumentation*, Notre Dame Press, 1969.

[20] A. Reinach, *Samtliche Werke, 2 Bde*, chapter Die apriorischen Grundlagen des burgerlichen Rechtes, Philosophia, Munchen, 1989.

[21] T. Schelling, *The Strategy of Conflict*, Harvard University Press, Cambridge, 1960.

[22] J. Searle, *Speech Acts*, Cambridge University Press, Cambridge, 1969.

[23] C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons, 'A framework for argumentation-based negotiation', *Lecture Notes in Computer Science*, **1365**, 177–192, (1998).

[24] M.P. Singh, 'Social and psychological commitments in multiagent systems', in *Proceedings of Knowledge and Action at Social & Organizational Levels, AAAI Fall Symposium Series*, ed., California Menlo Park, pp. 104–106. American Association for Artificial Intelligence, Inc., (1991).

[25] V. A. Thompson, J. St. B. T. Evans, and S. J. Handley, 'Persuading and dissuading by conditional argument', *Journal of Memory and Language*, **53(2)**, 238–257, (2005).

[26] D. Walton, *Scare Tactics: Arguments that Appeal to Fear and Threats*, Kluwer, Dordrecht, 2000.

Argument Understanding and Argument Choice A Case Study

Helmut Horacek[1]

Abstract. Several categories of discourse moves and sequences have been studied in formal models of disputes. However, most of these models make two simplifications that neglect important factors in argumentation: (1) raising an argument is typically done by introducing one or several new facts in the dispute, assuming that the associated warrant is self-evident, and (2) variants of arguments addressing the same issue are rarely assessed in terms of their benefits and drawbacks. In this paper, we illustrate these two points by studying the role of alternative arguments in explaining the solution to a rather simple, but not so easily understandable problem. Arguments may differ in terms of the effort needed to communicate them, the confidence they achieve, and requirements on knowledge of the audience, which makes their relative benefit task- and context-dependent.[1]

1 INTRODUCTION AND MOTIVATION

In the literature, several categories of argumentative moves have been studied in formal models of disputes, including arguments based on perception, statistics, and causality (see the sources of *prima facie* reasons in [7]). Arguments are examined in terms of their logical grounding [7], their role and contribution to progress in the discourse [6], and their potential to defend against attacks as raised by critical questions in argumentation schemata [10]. However, most models of argumentation include simplifications concerning the comprehensibility and variation of arguments. On the one hand, raising an argument is typically done by introducing one or several new facts in the dispute, assuming that the associated warrant is self-evident. Making the underlying reasoning more precise and explicit aims at uncovering implicit assumptions and potential sources for critical questions rather than addressing the comprehensibility of an argument. On the other hand, alternatives in arguments addressing the same issue are rarely considered, although benefits and drawbacks may vary significantly among possible tasks and contexts. We are convinced that studying these factors is likely to improve the understanding of driving forces underlying natural argumentation and associated skills significantly.

In this paper, we address the role of knowledge and purpose in argument choice in a case study, by examining the role of several categories of arguments in explaining the solution to the so-called *goat problem*. This problem constitutes a superficially simple task, but this task is not easily understandable at first, so that it gives rise to a variety of arguments providing sources of explanations. Arguments may differ in terms of the effort needed to communicate them, the confidence they achieve, and requirements on knowledge of the audience. Typical scenarios where the choice among such arguments and their presentation plays a prominent role include teaching reasoning in tutorial systems and argumentation within qualitative economic models.

This paper is organized as follows. First, we introduce the goat problem and its solution. Then we describe variants of arguments justifying that solution over the typically occurring misconception and discuss benefits and drawbacks. Finally, we sketch an operationalization of these concepts.

2 RUNNING EXAMPLE – THE *GOAT PROBLEM*

The goat problem is a superficially simple problem that originates from a game show. The problem comprises two consecutive guesses to be made by a candidate, with an apparently hidden dependency. The scenario consists of three doors, a car, and two goats. Behind each of the doors there is either the car or one of the goats, and the goal of the candidate is to guess where the car is (see Figure 1). In the starting position, the candidate makes an – apparently arbitrary – guess and picks one of the doors behind which he hopes the car being located. Then the showmaster opens one of the other two doors, unveiling one of the goats behind this door. Then the candidate is to make the second and final choice, in which he can stick to his original guess or alter it. The crucial question in the whole problem is whether one of these alternatives is superior to the other – and why – or whether the second choice offered is also a pure guess.

When confronting people with this problem, it turns out that not only finding but even understanding the solution is surprisingly difficult. The overwhelming majority of people unfamiliar with the problem believes that both alternatives in the second choice have the same likelihood to win, but this view is simply wrong. In contrast, changing the original choice is superior by a significant margin, winning two out of three times per average. The reason basically lies in the difference between the situation when the candidate first picks the door with the car behind it and the complementing situations when the candidate first picks a door with a goat behind it. In the second case, the showmaster has *no* choice, since he must present the only remaining goat and open the door in front of it. In the first case, however, the showmaster can pick any of the two remaining doors, and we can assume that he takes one or the other with equal likelihood. Hence, the second case occurs twice as often as the first case, so that altering the original choice is significantly superior.

[1] Universität des Saarlandes, FB 14 Informatik, Postfach 1150, D-66041 Saarbrücken, B.R.D., email: horacek@ags.uni-sb.de

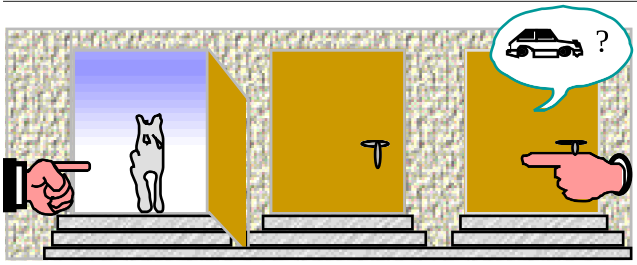


Figure 1. Running example scenario – the goat problem

3 CATEGORIES OF ARGUMENTS

Despite its superficial simplicity, the goat problem proves to be difficult to understand for humans. Therefore, several attempts have been undertaken to find illustrative explanations for the reasoning required to solve the problem. In addition, arguing in favor of the correct solution in a dispute may be of interest. Achieving a concession may not necessarily involve complete understanding on behalf of the other person, although this may also be beneficial for related pruposes, such as strengthening confidence. Consequently, there are several ways of arguing in favor of the solution, including the categories illustrated in Table 1:

1. As in many other situations, the simplest argument type is appeal to expert opinion. While this type of argument appears to be convincing to some extent, the confidence in it is limited, since the goat problem is fully accessible to well-justified logical reasoning.
2. Justification by statistics is probably the most convincing argument – in the given problem, this argument is not defeasible, since the task is to find out about the better chances in general, and not in an individual situation. There is an anecdote, that even a famous researcher in probability theory failed to understand the rationale behind the goat problem and did not believe in the solution. It was only due to simulations carried out by his students, which made his mind change – he knows perfectly well about the likelyhood of deviations from expected outcomes. The contribution to understanding the underlying rationale, however, is also not present in this argument category.
3. An extremely suitable argument is reference to an analogous problem, since a good deal of prior understanding can be exploited in this manner. The goat problem has a perfect counterpart in the game of bridge, namely the problem of restricted choice, where one of the opponents is in a situation that is isomorphic to the situation of the quizmaster in the goat problem. Unfortunately, referring to this analogy requires quite specific expertise, that is, being acquainted with the game of bridge at a non-trivial level. Whenever this argument is meaningfully applicable, its explanatory effect is very high.

North
♠ A 10 x x x
West ♠ x x ? ♠ J/Q ? East
♠ K x x x
South

	South	West	North	East
First trick:	♠ K	♠ x	♠ x	♠ J/Q
Second trick:	♠ x	♠ x	♠ ? (A or 10)	

Other things being equal, a singleton jack or queen in the East hand is twice as likely as queen and jack doubleton.

Figure 2. The principle of restricted choice in the game of bridge

4. The most commonly occurring argument is the exposition of a causal reason. In contrast to the other ones, an argument in this category provides a perspective on the rationale behind it, at least on some coarse-gained level. As the example texts in Table 1 demonstrate, even a short version is significantly longer than the arguments in the other three categories.

Since the rationale behind the solution to the goat problem is not easily to understand for humans, most of these arguments, specifically the reference to analogy and the causality exposition can be given in varying degrees of detail, the texts in Table 1 being on the short end of the scale. The analogy reference can also be formulated as a hint ("Consider how the problem of restricted choice in bridge can be related to the goat problem"), with a variety of adds-ons about the relation between the two problems. Moreover, the correspondence between the problems can be elaborated explicitly in an explanation, identifying the quizmaster with the defender playing the honor card in the first round, and mapping the associated ocurrences and decision preferences. Versions of the causality argument can differ even more in terms of detail and perspective, making the significance of the first choice evident, elaborating its consequences.

Like varying degrees of detail in the associated exposition, the suitability of categories of arguments justifying the solution to the goat problem depends on a number of contextual factors. One crucial factor is presence of specific knowledge that is required for using the analogy argument in a meaningful manner. Occasionally, testing the expertise of the audience prior to choosing an argument category may be beneficial to check the applicability of an efficient argument. Another factor is the goal of the discourse, which may range somewhere between the aim of just winning a dispute to the goal of enhancing the experience of the audience, as in a tutorial setting. If "winning" is the primary concern, a "hard" and comparably short argument such as appeal to expert opinion or reference to statistical results is probably preferable. When explanation is the primary concern, such references can only be accomponying arguments to a causally-based exposition. Moreover, this exposition needs to be tailored in an appropriate degree of detail according to the knowledge of the audience. Finally, even when winning a dispute is of some interest, this may be associated with a long-term goal of being assessed as a reliable arguer who deserves confidence. Under such circumstances, investigating in explanations that do not only convince the audience to some degree, but also enhance its understanding of the underlying rationale is likely to bear secondary benefits.

Table 1. Argument categories instantiated for the goat problem

1. Expert assessment
Informed experts recommend to change the original choice.
2. Justification by statistics
Simulations strongly favor changing the original choice.
3. Reference to analogical situation
The choice among the remaining doors works analogously to the problem of restricted choice in the game of bridge.
4. Causal reasoning
Altering the original choice is superior to staying with the original one. When the car is behind the door not previously pointed at, the quizmaster was forced to open the door he did, whereas he had a choice when the car is behind the door the candidate pointed at in his first guess.

4 TOWARDS AN OPERATIONALIZATION

In most approaches to formal models of natural argumentation, a warrant justifying the inference $p \rightarrow q$ (or, more general $P \vdash q$) is treated as a "unit". When it is introduced in the dispute, it is provisionally accepted, and may be attacked later. The assumption is that the inference itself is understood, otherwise accepting or attacking it is not meaningful. In contrast, we make a crucial distinction between degrees of *understanding* and degrees of *confidence*, to assess the effectiveness of an argument. Sufficient degrees of both components are required to make the argument acceptable.

The confidence in an inference depends primarily on the category of the underlying warrant. For some categories, degrees of understanding are also relevant. In order to address the understanding component in argumentation, we require arguments to be modeled in varying degrees of detail, for use in communication. While it is normally assumed that an argument $P \vdash q$ is also raised in precisely that form, we introduce expansions of arguments that make the underlying derivation more explicit. Thus, communicating an argument can either be done directly by *Say*($P\rightarrow q$), or an expanded form is introduced in the dispute, through Say($^{P}\nabla_{q}$) where $^{P}\nabla_{q}$ is a derivation tree underlying the argument $P \vdash q$ that makes some of its intermediate results explicit.

Exposing arguments in appropriate degrees of detail to meet the mental capabilities of an audience is a common topic at the intersection of the areas of deductive system and natural language presentation. Arguments in communication are frequently much more concise than in a mechanical proof [1], exploiting discourse expectations and background knowledge [3], which also holds for everyday discourse in comparison to underlying logical patterns [2, 8]. In contrast, some cognitively difficult reasoning patterns, such as modus tollens and disjunction elimination need to be exposed in more detail in order to support proper understanding [5, 9]. Hence, there are significant variations in terms of degrees of detail, which strongly influence degrees of comprehension, in accordance with the purpose of an expository explanation (full-depth, summary, sketchy idea [4]).

Table 2. Argument categories and understanding and confidence

	Understanding	*Confidence*
Expert assessment:	low	reasonable, but limited
Statistical justification:	mediocre	depending on the task
Analogy reference:	depends on related knowledge	
Causal assessment:	depends on thematic knowledge	

Based on these options, there are several factors which contribute to assessing the effectiveness of an argument, when raised in some chosen degree of detail:

- Degrees of *confidence* in the argument
- Degrees of *understanding* of the argument
- *Communicative effort* needed to expose the argument
- *Learning* of inferences through a detailed exposition

The last factor constitutes a kind of "investment" in subsequent sections of the dispute, with the idea that increasing the understanding of the other conversant may enable the beneficial use of causal or even analogical arguments with less communicative effort. The communicative effort is proportional to the size of the derivation tree that corresponds to the degree of detail in which the argument is to be presented. The degrees of understanding and confidence depend on the argument category, as sketched in Table 2. For an argument appealing to experts opinion, the degree of understanding is generally low, since a deeper understanding would require expertise. Moreover, a certain, but limited degree of confidence is present, in comparison to easier understandable arguments. Moreover, the degree of confidence depends on whether there is general agreement among experts about the issue at stake, or whether the expert opinion referred to is challenged by others. For an argument relying on statistics, the degree of understanding is similar, but it can be increased when more details are given about how the statistical procedure is used. The degree of confidence, in turn, may be increased when details about the strength of the statistical results are exposed. For the remaining argument categories, reference to analogy and causal assessment, the knowledge accessible to follow the causality in enough detail is the decisive factor. For analogy reference, that knowledge refers to the issue related through the analogy. In contrast to the other categories, the possible range in the degrees of understanding and confidence may vary significantly – they are virtually zero, if the causality (or analogy) is not understood, and maximal in case of full understanding.

In order to select among competing arguments from different categories and with varying degrees of detail, the domain in which the dispute takes place must be elaborated in two ways. Firstly, arguments must be made available in several version distinguished in their degrees of detail, or a mechanism must be provided which allows for such a construction. Secondly, a user model must be elaborated which allows assessing the knowledge of the other conversant in terms of the items appearing in different versions of arguments. Moreover, on the side of proper argumentation, the benefits of argument categories must be put in a precise relation to each other, including partial success, when arguments are not exposed to the degree of detail needed, as well as some contributions for the communicative effort and for the "learning component". Once these prerequisites are fulfilled, argument selection can proceed according to the following lines: for each argument candidate, the most compact version is picked and evaluated. Those arguments which are assumed not being fully understood by the addressee are successively expanded in relevant aspects according to the variations available. This process is continued for each argument until one of the following holds: (1) no more expansions are possible, (2) the argument is considered comprehensible in the degree of detail considered, or (3) the communicative effort is considered to be on its limit. From all argument versions generated this way, the one that scores best is chosen.

In an advanced version, such a system requires a full-fledged natural language generation approach, at least for text planning, when abstracting from surface realization. The task is then to express a communicative intention – here, making an argument, given a repertoire of alternatives in varying details, to meet assumptions about the intended audience, which in some sense appears to be a classical text planning task. The only extension in terms of assessing the relative merits of the alternatives available lies in judging the role of making an 'investment' through providing detailed expositions, which may make subsequent atrgumentation easier or which may even be necessary to pursue some future line of argumentation. Similar considerations proved to be problematic in dialog systems when playing the role of an agent with certain interest.

5 CONCLUSION

In this paper, we have studied the role of competitive arguments and requirements on knowledge to understand these arguments. In a case study, we have discussed the benefit of arguments in terms of their context and task-dependency, including tutorial purposes, dispute winning, and long-term goals aiming at establishing confidence. In the preliminary state of this work, the associated formalization is still on an abstract level only, that requires task- and domain-specific interpretation for an operational application.

Apparently, the example chosen for our case study is idealized in comparison to real argumentative scenarios. The available choice and variations in detail may be more limited in several realistic situations and, most importantly, arguments might be defeasible or, at least, it may be possible to weaken their strength. Apart from tutorial applications, scenarios where the considerations raised in the paper are important, are discussions with unbalanced levels of expertise, specifically when the role of a referee is more prominent than in most formal models of dispute. A typical application would be an expert discussion in television, arguing in favor or disfavor of competing strategies, such as economic models to improve the emploiment situation. In formal reconstructions of argumentative situations, such as cases at the court, benefits consist in uncovering implicit assumptions through raising critical questions. In addition to that, formal reconstruction of argumentation in more knowledge-intensive scenarios may also uncover missing knowledge required for following the course of the argumentation, through focusing on warrants that require a more detailed exposition. These additions, in turn, may lead to uncovering more deeply hidden implicit assumptions which improves not only the understanding, but also the reliabilty of the argumentation.

REFERENCES

[1] R. Cohen. 'Analyzing the Structure of Argumentative Discourse'. *Computational Linguistics* **13**(1-2): 11-24, (1987).

[2] H. Horacek. 'Generating Inference-Rich Discourse Through Revisions of RST-Trees', in *Proc. of AAAI-98,* pp. 814-820, (1998).

[3] H. Horacek. 'Presenting Proofs in a Human-Oriented Way', in *Proc. of CADE-99*, pp. 142-156, Trento, Italy, (1999).

[4] H. Horacek. 'Tailoring Inference-Rich Descriptions Through Making Compromises Between Conflicting Principles'. *International Journal on Human Computer Studies* **53**:1117-1146, (2000).

[5] P. Johnson-Laird, R. Byrne. Deduction. Ablex Publishing, (1990).

[6] H. Prakken. 'On Dialogue Systems with Speech Acts, Arguments, and Counterarguments', in Proc. of *7th European Workshop on Logic for Artificial Intelligence (JELIA'2000),* Springer Lecture Notes in AI (LNAI) 1919, 224-238, Springer, Berlin, (2000).

[7] J. Pollock. 'Defeasible Reasoning'. *Cognitive Science* **11**:481-518, (1987).

[8] M. Thüring, K. Wender. 'Über kausale Inferenzen beim Lesen'. In *Sprache und Kognition* **2**:76-86, (1985).

[9] Marilyn Walker. 'The Effect of Resource Limits and Task Complexity on Collaborative Planning in Dialogue'. In *Artificial Intelligence* **85**:181-243, (1996).

[10] D. Walton. *Argumentation Schemes for Presumptive Reasoning,* Mahwah, N.J., Erlbaum, (1996).

Argumentation-based Decision Support in Naval Command & Control

Hengameh Irandoust & Abder Rezak Benaskeur[1]

Abstract. Threat Evaluation and Weapons Assignment (TEWA), a process which is at the heart of tactical naval Command & Control (C2) process, comprises a number of operations that must be performed under time and resource constraints. This article discusses the challenges of decision making in this context, and more particularly the critical issue of target engagement, and shows how this process can be supported by an argumentation-based Decision Support System (DSS). It is shown how the information gathered and analyzed during the execution of the engageability assessment, defined and formalized for the purpose of the paper, can be exploited by an argumentation module. Based on a dialectical model and affording both proactive and reactive interaction modes, the module enables the DSS to anticipate and respond to the operator's objections to its recommendations, and thus substantially enhance the accuracy of its argumentation in a time-constrained decision support context.

Keywords : decision support, argumentation, explanation, threat evaluation, weapons assignment, engageability assessment, Toulmin's model

1 INTRODUCTION

Advances in threat technology, the increasing difficulty and diversity of open-ocean and littoral scenarios, and the volume and imperfect nature of data to be processed under time-critical conditions pose significant challenges for future shipboard Command & Control Systems (CCSs). Among other functionalities, the CCS provides capabilities to allow operators to evaluate the threat level of the different objects within the Volume of Interest (VOI), and when deemed necessary, use the shipboard combat resources to respond to them. This is commonly referred to as the Threat Evaluation & Weapons Assignment (TEWA) problem. It provides a time and resource-constrained application that involves both human and software decision-makers.

Current operational systems generally provide little support for tactical decision making. The need for such support is all the more pressing given the current emphasis on littoral warfare, including asymmetrical threats, that results in reduced reaction time and the need to deal quickly and correctly with complex Rules Of Engagement (ROEs).

The proposed Decision Support System (DSS) is based on a decision-centered perspective. The system assists the operator in making timely, error-free and effective decisions while reducing his cognitive workload. Yet, given the complexity of the problem he has to address, the high level of stress he is exposed to, and finally the fact that he knows that he will be held responsible for his decisions, the operator may discard the system's recommendation if he does not fully understand the underlying rationale, or if the recommendation is different from the solution he had foreseen. To overcome the operator's reluctance or lack of trust, the system has to convince him that its recommendation is based on sound reasoning. To do so, it needs to both retrieve the relevant knowledge structures and present them to the operator in a meaningful manner.

In this paper, we focus on the problem of target engagement, which is one of the most important decision making issues in TEWA. We introduce and define the*engageability assessment* process and show its usefulness in building trust in the system's information processing capability (Section 2). We then propose to organize the engageability assessment's data and results into an argument structure. This is first illustrated using Toulmin's inferential model of argument (Section 3). We then propose a dialectical model that can warrant the system's conclusion by anticipating and responding to the operator's objections to its arguments. Finally, we describe an argumentation module which based on this model, and by affording both proactive and reactive interaction modes, can substantially enhance the accuracy of the system's argumentation in a time-constrained decision support context (Section 4).

2 NAVAL TEWA

Naval Command & Control (C^2) is a very complex problem, and often this complexity arises from the multitude, the heterogeneity and the inter-relationships of the systems and resources involved. The tactical naval C^2 process can be decomposed into a set of generally accepted functions that must be executed within some reasonable delays to ensure mission success. A high-level description of those functions includes surveillance (*i.e.*, detection, tracking, and identification) and Threat Evaluation and Weapons Assignment (TEWA). In this paper, the focus will be on the TEWA process (see Figure 1), and more specifically the engageability assessment functionality, which concerns the evaluation of the feasibility of own-force's engagement options against non-friendly entities within the VOI.

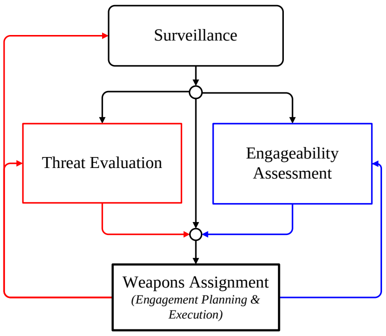


Figure 1. Global view of TEWA process

[1] Decision Support Systems Section, Defence R&D Canada - Valcartier, Canada, email: {Hengameh.Irandoust, Abderrezak.Benaskeur}@drdc-rddc.gc.ca

2.1 Threat Evaluation

Within the TEWA process, *threat evaluation* establishes the intent and the capability of potential threats within the VOI. The process results in a list $(rank_T)$ of entities ranked according to the level of threat they pose. For two objects O_i and O_j, $rank_T(O_i, t) < rank_T(O_j, t)$ means that O_i is more threatening, at time instant t, than O_j. $\mathcal{O}$ is the set of all objects O_i within the VOI.

2.2 Weapons Assignment

Weapons assignment makes decisions on how to deal with the identified threats. It can be seen as a real-time and constrained resource management problem. During this process, weapons are designated to engage threats. Also are assigned the supporting resources (*e.g.*, sensors, communications, etc.) required for each and every one-to-one engagement. This process results in a ranked list $(rank_E)$ that gives the recommended order of engagements for the threats, *i.e.*, the solution to the TEWA problem. For two objects O_i and O_j, $rank_E(O_i, t) < rank_E(O_j, t)$ means that, at time instant t, decision has been made to engage O_i before O_j. For a single weapon configuration, this boils down to a scheduling problem.

2.3 Engageability Assessment

The common definition of the TEWA process includes, as discussed above, the *threat evaluation* and *weapons assignment*. Nevertheless, one important issue that needs to be addressed is target engageability. Engageability assessment (see Figure 1) can support the *weapons assignment* module by eliminating candidate solutions that violate one or more of the problem constraints, and which for this reason will not be feasible. Several factors can be taken into consideration during this process, such as Rules Of Engagement (ROEs), pairing appropriateness[2], window (range, time, . . .), blind zones, ammunition availability, etc. (see Figure 2).

The engageability assessment outputs a list of objects ranked according to their engageability score E_s. The latter reflects the availability and feasibility of own-force options against all the non-friendly objects within the VOI. For two objects O_i and O_j, $E_s(O_i, t) > E_s(O_j, t)$ means that own-force has more options, at time t, against O_i than against O_j. Note that the engageability score is non-negative, that is $E_s(O_i, t) >= 0$. $E_s(O_i, t) = 0$ means that there is no solution (option) for engaging O_i at time instant t.

3 ARGUMENTATION-BASED DSS

The TEWA process can be seen as a dynamic decision-making process aimed at the successful exploitation of tactical resources (*e.g.* sensors, weapons) during the conduct of C^2 activities. From this perspective, decision support is defined to be a capability that is meant to assist operators in making well-informed and timely decisions while providing robust error detection and recovery. The DSS must be designed as to reduce the operator's cognitive overload and improve the overall effectiveness of the process [7].

However, the complexity of the TEWA problem, the issues that are at stake, the high level of stress induced by resource and time constraints, the effects of stress and fatigue on attentional resources, and most important of all, the sense of responsibility with regard to one's decisions, can all lead to a situation of under-confidence, where the operator becomes overly concerned with the perils of a course of action [6]. In such a situation, it is unlikely that the operator will accept the system's recommendation if he does not fully understand it or if the recommendation is different from the alternatives he had considered [11], a phenomenon referred to as an *expectation failure*[3].

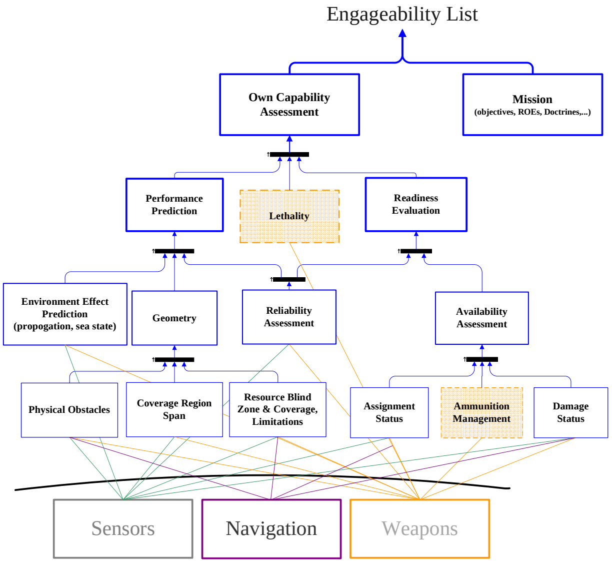


Figure 2. Engageability Assessment Inferential Model

To be acceptable by the user, the information provided to him needs to be presented in a comprehensible and convincing manner. Indeed, it is not only the quality of the recommendation made by the DSS that needs to be improved through more optimized processing, but also the user's *interpretation* of the quality of the decision [6].

This interpretation can be substantially improved if the system has the capacity to expose its rationale using sound arguments. To address this problem, we need to use an argumentative structure that can capture the inferential nature of reasoning used in TEWA, and more specifically in the engageability assessment process[4]. Toulmin's model of argument [8] or argumentative schemes [9] seem appropriate for this purpose. However, our approach requires a different mechanism since in this context what determines the strength of a support for a claim is how well it can respond to specific objections, and not, for example, how widely accepted it is. In the following, we first show how Toulmin's general model can be used to outline an argument based on the information provided by the engageability assessment. Then we show how the basic inferential structure can be augmented with a dialectical component which is more adapted to a time-constrained decision support context.

3.1 Toulmin's model

Toulmin proposes an argument structure that reflectsthe natural procedure by which claims can be argued for. The model is composed of six elements that depict the move from a set of premises to a conclusion.

In addition to the premise-conclusion structure, Toulmin identifies several components that support the inferential relation. The warrant

[2] Ensure that the weapon selection corresponds to the threat type.

[3] See Section 4.2 for a more detailed discussion of expectation failure.

[4] Solutions are inferred from the intermediary results input by lower-level processes, as shown in Figure 2.

has the function of a rule of inference, licensing the conclusion on the basis of the arguer's data or grounds. The arguer can invoke a backing if the warrant is challenged or insufficient. The modal qualifier is a word or phrase that indicates the force of the warrant. Finally, the rebuttal accounts for the fact that some exception-making condition might be applicable [3].

The model expresses plausible reasoning, captures inferetial mechanisms, can outline a decision situation and preserve it for future use, and finally, can be used as a basis for explanation facilities [10]. Useless to say that Toulmin's model has been extensively cited in argument studies[5], particularly informal logic, as well as in artificial intelligence, and has even been applied to military problems such as theater missile defense [2].

3.2 Example of Application of Toulmin's Model

Table 1 presents an example of the application of Toulmin's Model to the TEWA problem. The example is based on the concept of engageability assessment, formalized in Section 2.3. The results of engageability assessment, based on constraints violation avoidance, are used as intermediary results to justify recommendations for the weapons assignment phase.

Data	Two objects (O_i, O_j) have been detected within VOI and assessed hostile to ownship. Object O_j has been assessed more threatening than O_i. Options against both objects have been evaluated. As a result, the engagement order (O_j, O_i) has been deemed non-feasible, while (O_i, O_j) offers options.
Qualifier	Supports
Claim	The weapons assignment module recommends the engagement order (O_i, O_j).
Warrant	Since by the end of engagement of O_j, O_i will enter the Fire Control Radar (FCR) blind zone, while by the end of engagement of O_i, O_j will still be within the FCR coverage area.
Backing	The Anti-Ship Missile (ASM) nature of threats requires the use of Surface-to-Air Missile (SAM) to counter them. FCR support is mandatory for the SAM's guidance and threat illumination.
Rebuttal	Unless probability of kill (P_k) on O_i is much lower than for O_j.

Table 1. Example of Toulmin model's application

The controversial nature of the claim requires that the inferential relation be licensed with a warrant. In Toulmin's model, a warrant is a general law ('major premise' in Walton's argumentation schemes) which licenses the move from data to a claim. Here, the system has to warrant the recommendation with specific information. Also, the domain knowledge provided in the backing will be of little use for the operator who will rather want to know what are exactly the factors that the system has considered. As a matter of fact, the warrant may be challenged, not because the reason it provides is not good enough, but because the operator may object that the conditions under which that warrant holds can be modified (see Section 4.2).

Based on these remarks, we propose to augment the premise-conclusion structure with a dialectical component that will enable the DSS to handle such situations.

4 INFERENTIAL MODEL OF ARGUMENT WITH A DIALECTICAL COMPONENT

The functional account of Toulmin's model is a deductive, rather than a dialectical model of argumentation in that it does not take into account the beliefs, opinions or reasoning schemes of the audience it is addressed to. In a dialectical scheme, the arguer has to consider possible counter-arguments. In Toulmin's model, although the rebuttal accounts for the possibility of the defeat of the argument, it simply shows that an exception-making condition might be applicable. This is a condition that the arguer contemplates, but it is not a condition that he considers as being the object of his audience's belief. Reasoning on the beliefs of the audience is the core of dialectical reasoning. As Johnson [5] has argued, because the conclusion may not meet the initial beliefs of the audience, an arguer will need to do more than put forward some supporting statements. He or she will need to respond to objections and alternative positions.

4.1 Model of dialectical argumentation

The dialectical component can be viewed as an *argument-objection-response to objection* sequence. This justificatory triad warrants the inference from data to a claim, which in the case of a decision support system is a solution or recommendation. This is illustrated in Figure 3.

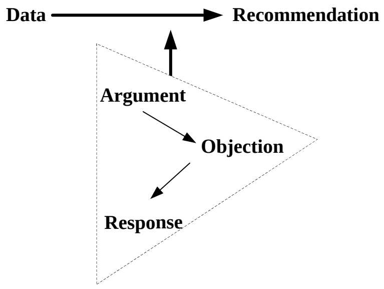


Figure 3. Inferential Model of Argument With a Dialectical Component

Using this model, we propose to design the DSS so that it can anticipate possible objections on the part of the operator and prepare its responses to those objections. This concept is illustrated in the following using the engageability assessment process, where the constraints violation avoidance principle is used as a basis for argument/response generation.

4.2 Use of constraints for argumentation

Most of the time, decision problems such as TEWA that have to be solved under constraint lead to sub-optimal solutions. The set of constraints defines the feasibility space in which the system will have to search for the best solution. The harder are the constraints, the smaller is this space, and the farther can be the solution from the optimal[6]. For the TEWA problem, the feasibility of different options is defined by means of the engageability assessment. The smaller is the engageability score E_s of the objects in the VOI, the smaller will be

[5] See the recent OSSA's conference theme.

[6] Since the optimal may not belong to the feasible solution space.

the solution space for weapons assignment, and the more distant will be the engagement plan from the operator's expected plan, hence the increasing risk of an *expectation failure*.

An expectation failure generally happens when the solution proposed by the system is different from the one the user had predicted. Given the very limited number of constraints he can consider at a time, a human operator often works on simplified representations of problems that capture only a subset of the actual constraints. A DSS, which is not as limited as the human operator in its working memory, can handle a much larger number of constraints. This difference can lead to a situation where the solution foreseen by the operator is closer to the optimal than the one recommended by the DSS. The discordance between the two solutions can be justified by the number and the nature of constraints that would be violated if the DSS tried to get closer to the optimal in order to meet the operator's expectations.

The engageability assessment concept can be used to illustrate the idea. Since engageability assessment is about the evaluation of the feasibility of engagement plans, it mainly boils down to a Constraint Satisfaction Problem (CSP). Examples of such constraints are given in Table 2, among which some are relaxable (considered as soft constraints for which solutions may exist) and some non-relaxable (considered as hard constraints for which no solution exists).

One case where the expectation failure situation may happen is the following. For two objects O_i and $O_j (i \neq j)$

$$rank_T(O_i,t) > rank_T(O_j,t)\ \&\ rank_E(O_i,t) < rank_E(O_j,t)$$

which means that O_j is more threatening than O_i, yet O_i is judged as being of higher priority from the engagement perspective. This situation can be problematic because the operator will be more likely to rely on the threat list ranking ($rank_T$) for the engagement prioritization[7]. Such engagement order cannot be presented to the operator without the support of some credible reasons. The engageability assessment module can justify this outcome. A typical case that can explain the controversial recommendation above is as follows. For two objects O_i and O_j, if

$$rank_T(O_i,t) > rank_T(O_j,t) \tag{1}$$

that is, O_j is more threatening than O_i, and

$$\begin{aligned} E_s([O_j,O_i],t) &= E_s([O_j],t) \times E_s([O_i],t+d_j) \\ &< E_s([O_i],t) \times E_s([O_j],t+d_i) = E_s([O_i,O_j],t) \end{aligned}$$

which means that the engagement sequence (O_i,O_j) offers more possibilities to own-force than (O_j,O_i). A special case is where $E_s([O_j,O_i],t) = 0$, while $E_s([O_i,O_j],t) \neq 0$, which means that the sequence (O_j,O_i) is not feasible. This can be caused by the loss of opportunity on O_i during the engagement of O_j.

The more and the harder are the constraints that define the feasibility space, the more difficult it will be for the DSS to bridge the gap between the two solutions. In anticipation of the operator's dissatisfaction, those constraints that would be violated if the DSS deviated from its solution, are stored at run-time during the engageability assessment. These are later presented to the operator by the argumentation module (see Section 4.3) in response to his objections.

4.3 Argumentation module

The proposed argumentation module is depicted in Figure 4. The engageability assessment process evaluates the set of possible solutions and discards those which would violate one or more constraints. The results of this constraints violation avoidance process are stored in a database and used as arguments to be presented to the user.

Non-relaxable	**Relaxable**	**How**
-Rules of engagement	-Availability of supporting resources	-*Free resources*
-Availability of ammunition	-Damage status	-*Repair*
-Lethality	-Assignment status	-*Re-assign*
-Appropriateness of resource choice	-Coverage limitations (*Envelope, Blind Zone, Obstruction*)	-*Wait, move*
	- Predicted Performance (e.g. PK)	-*Wait*

Table 2. Examples of constraints considered during engageability assessment for a given resource against a given object, at time instant t.

The argumentation module can display its dialectical skills using both proactive and reactive interaction modes. The *response coordinator* selects and coordinates dynamically the two modes. The difference between them lies in the fact that the dialectical cycle is initiated by the argumentation module in the pro-active mode, while it is initiated by the user in the reactive mode. An argument is called response when provided reactively (in response to an objection). The numbers in Figure 4 show the chronology of the events for each mode. The role of the *response coordinator* is twofold: i) receiving the user's objections, and ii) coordinating the deployment of the interaction mode.

Having prepared itself for all possible cases of disagreement, the coordinator will first activate the proactive mode and proceed by presenting its best arguments. These are those arguments that are the most persuasive responses to what it considers to be the most likely objections. It will then shift to a reactive mode and provide justification only upon user's further objections. This will be the case if the operator formulates more specific objections or if more detailed or low-level information is needed.

Naturally the operational context described here, where time is a serious issue, does not allow for a genuine dialogue between the system and the operator and therefore models such as that of the deliberation dialogue [4] cannot be applied.

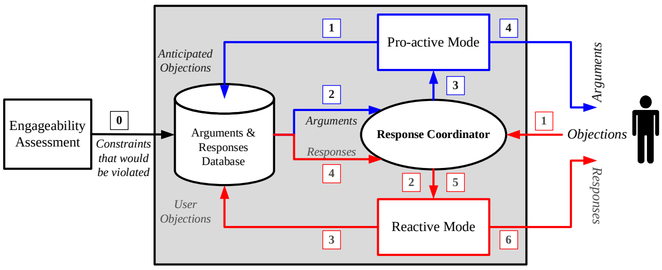


Figure 4. Argumentation Module Architecture

In the above-described process of argumentation, the nature of the constraints plays a major role in the weight of the justification (*i.e.*, its persuasive power). Logically, avoiding the violation of non-relaxable constraints will have a higher justificatory power than avoiding the

[7] This is a common practice in modern navies, where capability limitations are only considered at the later stage of response planning process, with possibility of plan revision in case of an empty feasibility space.

violation of relaxable ones. From an argumentation perspective, it is assumed that the former constitutes a sufficient condition for the conclusion to obtain, while the latter does not. It is also expected that the user will object to the arguments based on relaxable constraints by asking the system to modify them so that they can be satisfied. Examples of such possible objections are given in the column "How" of Table 2.

For the TEWA problem, the engageability assessment module will have to verify a set of N_R relaxable constraints and a set of N_{NR} non-relaxable constraints, for a total of $N_R + N_{NR}$ constraints. The set of non-satisfied constraints will be used to constitute dynamically the system's arguments/responses database (see Figure 4). Based on the content of this database, the system provides pro-actively a maximum of N arguments to the user. Given their higher justificatory power, priority is given to arguments related to the non-relaxable constraints. The presence of at least one non-relaxable constraint that could be violated eliminates the need to consider arguments related to relaxable constraints. If there is no such non-relaxable constraint, the system will present the N arguments related to relaxable constraints that are deemed most likely to be mentioned by the user. The remaining set of constraints that may not be satisfied will be provided reactively on a one-by-one basis, should the user continue to object to the system's recommendations.

To illustrate the idea, let us take the same example as previously where two objects (O_i, O_j) have been detected within VOI and assessed hostile to own-force. Object O_j has been assessed more threatening than O_i. Engageability for both objects has been evaluated. As a result and based on the different constraints, engagement of O_j is deemed non-feasible (*i.e.*, $E_s(O_j) = 0$) and only O_i is engageable and will be engaged ($E_s(O_i) \neq 0$).

Situation 1 (Sufficient Arguments) – this corresponds to the case where one or more non-relaxable constraints would not be satisfied. For example, if ROEs prevent own-force from engaging O_j, any solution that includes engagement action on O_j will not satisfy this hard non-relaxable constraint. This information can be used as a sufficient argument that cannot be objected to by the user, and no further arguments will be required. This argument is presented pro-actively, and there is no need to consider arguments related to relaxable constraints.

Situation 2 (Non-sufficient Arguments) – this corresponds to the case where all non-relaxable constraints are satisfied and one or more relaxable constraints are not satisfied. Based on the set of constraints that would be violated by engagement action on O_j, the DSS decides to present pro-actively the two ($N = 2$) following arguments, regarding the recommendation of not engaging O_j. These arguments are: i) O_j lies within the blind zone of the only available Fire Control Radar (Coverage limitation constraint), and ii) the other Fire Control Radar is assigned to another target (Assignment status constraint). The other constraints that would be violated, if any, will be used by the reactive mode.

Given the relaxable nature of the constraints they are related to, these arguments are not sufficient. As a consequence, it is expected that the operator will object, asking why the constraints are not relaxed so that the feasibility space can be extended (*i.e.*, the engageability score $E(O_j)$). Examples of objections/responses that may be used in the reactive mode of the system following the first argument, are given below (see Table 2).

1. **Objection 1 (*Wait*)–** meaning: wait until the object O_j gets out of the Fire Control Radar blind zone and provide engagement solution. Example of a possible response to this objection is: *object will get out of the weapon range as well*.
2. **Objection 2 (*Move*)–** meaning: move the ship to clear blind zone. Examples of possible responses to this objection are: *Physical obstacle prevents from moving; Not enough time to move; Jeopardizes other engagements that are in progress; Increases ship's Radar Cross Section (visibility by the enemy sensors); Puts more threatening objects within blind zones*.

The above list gives examples of potential reasons that may render the decision of moving the ship (one of user's anticipated objections) not feasible.

The examples discussed above show how the system can exploit knowledge of the domain and knowledge of the user to justify a recommendation that does not meet the initial beliefs of the operator. They also show how the system can display a strategic behaviour by planning its argumentation.

5 CONCLUSION

The organization of the system's knowledge into argument structures provides insight into the system's states, procedures and goals, and shows the extent of its domain knowledge and capacities. A better understanding of these features will hopefully result in a more efficient use of the system proposed. The argumentation capability described above, not only outlines the system's reasoning process, but it also engages a dialectical exchange by anticipating possible objections and by organizing its responses to them according to their degree of justification. The two-phase approach, proactive and reactive argumentation, can be very effective for handling decision making issues in a time-constrained context such as TEWA. The same analysis as the one described for engageability assessment is being performed for threat evaluation and the whole system is under design for implementation for the Canadian Navy.

REFERENCES

[1] S. Das, 'Symbolic argumentation for decision making under uncertainty', in *Proceedings of Fusion 2005*, Philadelphia, PA, USA, (July 2005).

[2] D. Hitchcock, 'The significance of informal logic for philosophy', *Informal Logic*, **20**(2), (2000).

[3] D. Hitchcock, P. McBurney, and S. Parsons, 'A framework for deliberation dialogues', in *Proceedings of the Ontario Society for the Study of Argumentation*, Windsor, Ontario, Canada, (2001).

[4] R. Johnson, *Manifest Rationality*, Lawrence Erlbaum, 2000.

[5] G.M. Kasper, 'A theory of decision support system design for user calibration', *Information Systems Research*, **7**(2), (1996).

[6] S. Paradis, A. Benaskeur, M. Oxenham, and P. Cutler, 'Threat evaluation and weapons allocation in network-centric warfare', in *Proceedings of Fusion 2005*, Philadelphia, PA, USA., (July 2005).

[7] S.E. Toulmin, *The Uses of Argument*, Cambridge University Press, 1964.

[8] D. N. Walton and C.A. Reed, 'Argumentation schemes and defeasible inferences', in *Proceedings of the 2nd Workshop on Computational Models of Natural Argument*, Lyon, France, (2002).

[9] L.R. Ye and P.E. Johnson, 'The impact of explanation facilities on user acceptance of expert system advices', *MIS Quarterly: Management Information Systems*, **June**, (1995).

[10] M. Zanella and G. Lamperti, 'Justification dimensions for complex computer systems', in *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics*, pp. 317–324, Orlando, Florida, USA, (1999).

Dialectical text planning

Rodger Kibble
Dept of Computing, Goldsmiths College
University of London, UK
`r.kibble@goldac.uk`

Abstract. A key requirement for the automatic generation of argumentative or explanatory text is to present the constituent propositions in an order that readers will find coherent and natural, to increase the likelihood that they will understand and accept the author's claims. Natural language generation systems have standardly employed a repertoire of coherence relations such as those defined by Mann and Thompson's Rhetorical Structure Theory. This paper models the generation of persuasive monologue as the outcome of an "inner dialogue", where the author attempts to anticipate potential challenges or clarification requests. It is argued that certain RST relations such as Motivate, Evidence and Concession can be seen to emerge from various pre-empting strategies.

1 Introduction

A key requirement for the automatic generation of argumentative or explanatory text is to present the constituent propositions in an order that readers will find coherent and natural, to increase the likelihood that they will understand and accept the author's claims. Ideally, any objections or clarification requests that an audience might raise will already have been countered by elements of the author's argument. In fact this paper models the generation of persuasive monologue as the outcome of an "inner dialogue", where the author attempts to anticipate potential challenges or clarification requests. It will be argued that certain coherence relations can be seen to emerge from various strategies for pre-empting or "obviating" challenges or clarification requests.

This paper assumes a model of dialogue as updating participants' information states (IS), where an IS consists of a record of each interlocutor's propositional and practical commitments (cf [7, 2, 17]) rather than "mental states" such as belief and intention (cf [3]). This approach is motivated at greater length and contrasted with other commitment-based approaches such as [12] in [8, 9]; the key assumptions for the purposes of this paper are:

1. Each agent in a dialogue keeps a score of social commitments for all participants, including itself. Commitments can be classified into *practical* (commitments to act, corresponding to *intentions* in mentalistic accounts) and propositional or *doxastic* (commitments to justify an assertion, corresponding to *beliefs*).
2. Agents play one of three dynamically assigned roles at any given point in a dialogue: Speaker (**Sp**), Addressee (**Ad**), or Hearer (**He**) who is not directly addressed.
3. For an agent α to assert ϕ is to acknowledge commitment to ϕ; other agents may also attribute consequential commitments to α.
4. Additionally, a dialogue act constitutes an attempt to commit Addressee(s) to a proposition or a course of action, as detailed in the following section.
5. Addressee's options include accepting the proffered commitment, challenging it or requesting clarification.

This paper will focus on modelling persuasive monologue, or extended dialogue turns, as emerging from a process of internal argumentation, with the virtual agents Planner (**Pl**) in place of **Sp** and Critic **(Cr)** substituted for **Ad**. I will aim to show how a variety of Mann and Thompson's RST relations such as Motivate, Justify, Evidence, Concession and Elaboration can be seen to emerge from different text planning strategies [11, 16] . It might be argued that this is an essentially trivial exercise in shifting information from a pre-defined set of coherence relations to a pre-defined set of dialogue acts and moves. However, there are independent motivations for developing models for dialogue and argumentation, and the argument in this paper is that a (possibly partial) account of coherence relations in monologue emerges as a side-effect of these models. The paper will conclude by addressing some apparent differences between dialogue and monologue as discussed by [14] and [6].

2 Argumentation and discourse relations

The full framework will include specifications for the proto-speech acts listed below. Note that I use upper-case Greek letters such as Φ to represent speech acts themselves and lower-case letters such as ϕ for the propositional content of the speech acts.

assert(Sp, ϕ, Ad, He) undertake commitment to justify a propositional claim; attempt to bestow same commitment on **Ad**.

instruct(Sp, ϕ, Ad, He) attempt to bestow a practical commitment on Addressee.

endorse(Sp, ϕ, Ad, He) Speaker adopts a commitment specified by Addressee

challenge(Sp, Φ, Ψ, Ad, He): require agent to justify or retract a commitment offer Φ, with Ψ as an optional counter-commitment. Note that the challenge may be directed at the propositional content ϕ, or at the appropriateness of the speech act itself.

respond(Sp, challenge(Ad, Φ, Ψ, Sp, He), Ξ, Ad, He)
respond to a challenge with a dialogue act Ξ which may be:

- asserting ξ as evidence for ϕ, or as justification for uttering Φ;
- retracting commitment to ϕ, the propositional content of Φ;
- withdrawing a claim to justification for the speech act Φ;
- challenging Ψ;

- requesting clarification of Ψ;
- ϵ - the null act. How this is interpreted will depend on the particular conventions currently in force: it may be understood at different times as implicit endorsement, implicit denial or non-committal.

retract(Sp, ϕ, Ad, He) withdraw a commitment to ϕ.
query(Sp, Φ, Ad, He) request clarification of Φ
respond(Sp, query(Ad, Φ, Sp, He), Ψ, Ad, He)
respond to request for clarification of Φ by uttering the speech act Ψ.

2.1 Examples of dialogue and monologue

The following examples consist of a short dialogue followed by two variants of a monologue expressing roughly the same content and exemplifying particular rhetorical structures.

Example (a)

A: You should take an umbrella.
B: Why?
A: It's going to rain.
B: It doesn't look like rain to me. It's sunny
A: Michael Fish predicted it.
B: Who's he?
A: He's a weather forecaster on the BBC.
B: OK.

In terms of the speech acts defined above, this exchange can be represented (somewhat simplified) as follows:

A: instruct(A, *take-umbrella*, B, _);
B: challenge(B, *take-umbrella*, _, A, _);
A: respond(A, challenge(B, *take-umbrella*, _, A, _), assert(A, *rain-later*, B, _), B, _)
B: challenge(B, *rain-later*, *sunny-now*, A, _);
A: respond(A, challenge(B, *rain-later*, *sunny-now*, A, _), assert(A,*fish*,B, _), B, _)
B: query(B, *fish, A,* _)
A: respond(A, query(B, *fish, A,* _), assert(A, *BBC-forecaster*, B, _), B, _)
B: endorse(B, {*BBC-forecaster*; *fish*; *rain-later*; *take-umbrella*}, A, _)

Example (b)

A: You should take an umbrella. It's going to rain. I heard it on the BBC.

A possible RST analysis of this example is:

Motivate
Nucleus *take-umbrella*
Satellite: Evidence

Nucleus *rain-later*

Satellite *BBC-forecast*

Example (b′)

A: You should take an umbrella. It's going to rain, even though it looks sunny right now. I heard it on Michael Fish's slot. He's a weather forecaster at the BBC.

Proposed RST analysis:

Motivate
Nucleus *take-umbrella*
Satellite

Evidence

Nucleus

Concession

Nucleus *rain-later*

Satellite *sunny-now*

Satellite

Background

Nucleus *fish*

Satellite *BBC-forecaster*

Example (c)

A: I listened to the weather forecast on the BBC. It's going to rain. You should take an umbrella.

Proposed RST analysis: same rhetorical structure as (b) but realised in a satellite-first sequence:

Motivate
Satellite: Evidence

Satellite *BBC-forecast*

Nucleus *rain-later*

Nucleus *take-umbrella*

2.2 Speaker strategies

In the above scenario, suppose A has the goal that B undertake a practical commitment to carry an umbrella. Examples (a - c) illustrate three different strategies:

(i) Issue a bare instruction; offer justification only if challenged.
(ii) Issue an instruction, followed by an assertion that **pre-empts** a potential challenge, and recursively pre-empt challenges to assertions.
(iii) **Obviate** the challenge by uttering the justification **before** the instruction, and recursively obviate potential challenges to assertions.

(The terms **pre-empt** and **obviate** are used with these particular meanings in this paper, which may not be inherent in their ordinary usage.) Note that examples (a) and (b′) exhibit the same sequence of propositions, which is consistent with the assumption that (b′) results from a process of internal argumentation with a virtual agent that raises **Ad**'s potential objections. The following section will sketch a formulation of strategies (i - iii) in terms of the Text Planning task of natural language generation.

3 Dialectical text planning

I will assume some familiarity with terms such as “text planning” and “sentence planning”. These are among the distinct tasks identified in Reiter’s “consensus architecture” for Natural Language Generation [15]; see also [1]:

Text Planning/Content Determination - deciding the content of a message, and organising the component propositions into a text structure (typically a tree). I will make a distinction between the **discourse plan** where propositions in the initial message are linked by coherence relations, and the **text plan** where constituents may be re-ordered or pruned from the plan.

Sentence Planning - aggregating propositions into clausal units and choosing lexical items corresponding to concepts in the knowledge base; this is the level at which the order of arguments and choice of referring expressions will be determined.

Linguistic realisation - surface details such as agreement, orthography etc.

3.1 Discourse planning

Text planning is modelled in what follows as the outcome of an inner dialogue between two virtual agents, the Planner (**Pl**) and the Critic (**Cr**). The Critic is a user model representing either a known interlocutor or a “typical” reader or hearer. A’s options (i -iii) in Section 2.2 above can be seen to correspond to three different strategies which I will call *one-shot*, *incremental* and *global*. These strategies are presented in rather simplified pseudo-code below, in particular I only consider the assert action and selected responses to it.

One-shot planning

Speaker produces one utterance per dialogue turn which may be:

- a bare assertion ϕ;
- response to a challenge or clarification request from Addressee;
- challenge to Address’s most recent or salient assertion, or request for clarification;
- ϵ

The message is passed directly to the text planner without being checked by the Critic. This strategy is appropriate when no user model is available.

Incremental Planning

Speaker generates the “nuclear” utterance and then calculates whether a challenge is likely, and recursively generates a response to the challenge if possible. This is the strategy of **pre-empting** challenges referred to in section 2.2. The response is immediately committed to the right frontier of Speaker’s text plan.

```
procedure inc-tp($\Phi$)
  where $\Phi$ is some speech act with propositional content $\phi$;
send $\Phi$ to text planner;
assert(Pl, $\phi$, Cr, _);
if challenge(Cr, $\phi$, $\psi$, Pl, _)
then do inc-tp(respond(Pl, challenge(Cr, $\phi$, $\psi$, Pl, _), $\Xi$, Cr,
  _);
else quit.
```

This strategy is appropriate when a suitable user model is available but resource limits or time-criticality make it desirable to interleave discourse planning, text planning and sentence generation.

Goal-directed Planning

The sequence is globally planned in order to rebut potential challenges by generating responses to them ahead of the nuclear proposition. This is the strategy I have dubbed **obviating** challenges in section 2.2.

```
procedure gd-tp($\Phi$)
  where $\Phi$ is some speech act with propositional content $\phi$;
initialise stack = [ ];
call gd-tp-stack($\Phi$);
do until stack = [ ]:
   pop $\Psi$ from stack;
   add $\Psi$ to text plan;
end gd-tp()


procedure gd-tp-stack($\Phi$)
stack = [$\Phi$ | stack];
assert(Pl, $\phi$, Cr, _);
if challenge(Cr, $\phi$, $\psi$, Pl, _)
then do gd-tp-stack(respond(Pl, challenge(Cr, $\phi$, $\psi$, Pl, _),
  $\Xi$, Cr, _);
else quit gd-tp-stack
end gd-tp-stack()
```

This strategy is appropriate for applications where resources allow for the full discourse plan to be generated in advance of text planning so that constituents may subsequently be reordered or pruned to produce a possibly more “natural” and readable text.

3.2 Text planning and plan pruning

If we consider the examples in section 2.1: (b), (b′) are typical products of incremental planning and (c) of goal-directed planning. The former will result in **nucleus-first** structures, while the default ordering resulting from the latter will realise satellites **before** nuclei. Two refinements are discussed in this section: **plan pruning** and **re-ordering** of the text plan.

The differences between (b) and (b′) demonstrate that the text planner has a choice over whether to realise only the Planner’s contributions or those of the Critic as well. The latter option, retaining the proposition *sunny-now*, results in instances of RST’s Concession relation. This is a special case of **plan pruning** as described by [6], where a constituent may be removed if it is inessential to the speaker’s purpose: for instance it may be inferrable from other material in the plan. Green and Carberry motivate this with the aid of the following example (their (13a-e)), illustrating how a question-answering system might decide how much unrequested information to include in an indirect answer to a yes-no question.

Example (d)

(i) Q: Can you tell me my account balance?
(ii) R: [No.]
(iii) [I cannot access your account records on our computer system.]
(iv) The line to our computer system is down.
(v) You can use the ATM machine in the corner to check your account.

Items (ii - iii), shown in square brackets, can be suppressed since (iii) is inferrable from (iv) and in turn implies (ii). This assumes that the user is aware, or can accommodate the fact that their account balance is kept on the computer system. This example is compared with an "imaginary dialogue" where each statement responds to a specific question from the user.

As stated above, the planning strategies outlined in section 3 produce texts that are uniformly either satellite-first or nucleus-first by default. There is a need to generalise the strategies so that the planner can dynamically switch from one to the other, in order to produce texts such as:

Example (e)

It's going to rain. I heard it on the BBC. You should take an umbrella.
RST analysis:

Motivate

Satellite: Evidence

Nucleus *rain-later*
Satellite *BBC-forecast*

Nucleus *take-umbrella*

By distinguishing between the **discourse plan** and **text plan** we allow for re-ordering of constituents at the level of the text plan, within the partial ordering defined by the discourse plan. For instance, a different ordering of propositions might improve the referential coherence of a text according to Centering Theory [10].

3.3 Summary

In contrast to approaches to text generation that carry out top-down planning using pre-defined coherence relations I have argued that certain RST relations can be seen to emerge from sequences of internalised dialogue moves that aim to pre-empt or obviate potential challenges or clarification requests, as follows:

instruct-challenge-respond underlies Motivation or Justify depending on the content of the challenge and response;
assert-challenge-respond underlies Evidence if the propositional content is challenged, or Justify if the appropriateness of the assert act itself is at issue.
<any-speech-act>-challenge-respond underlies Concession if the content of the challenge is realised in the text.
<any-speech-act>-query-respond underlies Background.

It remains to be seen if further RST relations can be modelled using the "dialectical" method.

4 Discussion and future work

4.1 Objections to "implicit dialogue"

Reed [14] argues against identifying a persuasive monologue with an implicit dialogue and emphasises the importance of distinguishing the *process* of creating a monologue from the *product*, the monologue itself. Now, it is not argued here that a monologue is nothing more than a trace of the dialogical process of constructing an argument. The "goal-directed" strategy allows for a phase of pruning and re-ordering the text plan (not described in detail here) although the default is for propositions to be realised in the sequence in which they are added to the discourse plan.

Reed puts forward an important argument: that a crucial difference is the fact that unlike a dialogue, a "pure" monologue must not contain a *retraction* in the sense of asserting a proposition and its negation. This has implications for the discussion of text planning strategies in section 3 above, since there is the possibility of a contradiction occurring in a sequence of responses to recursive challenges. On the one hand, goal-directed planning could be extended with a backtracking facility and consistency checking such that indefensible claims or even the nuclear proposition itself could be withdrawn before proceeding to sentence generation, if a challenge generated by the Critic shows up a contradiction in the existing plan. However, the essence of incremental planning is intended to be that each proposition is committed to the text plan, to be passed on to the sentence planner, *before* considering potential challenges. The algorithm as adumbrated above certainly allows the possibility that contradictory propositions will be added to the plan, as a consequence of limitations on speakers' memory and reasoning capabilities.

The proscription of overt retraction would certainly be a reasonable design feature for a computer system generating argumentative text. However, this paper is also concerned with modelling the ways in which human speakers might construct an argument, and so this comes down to an *empirical* question as to whether speakers delivering an extempore monologue will ever realise part-way through that there are insuperable objections to their initial claim (or a subordinate claim), and end up withdrawing it. For instance, the medium of communication might be an electronic "chat" forum such that all keystrokes are instantly and irrevocably transmitted to other logged-on users. It is not obvious that this possibility should be ruled out in principle, or even that it can be ruled out in a resource-limited system following "incremental planning" as defined here.

4.2 Future work

The following issues will be addressed in future research:

Coherence, user modelling and reasoning. It is assumed that for a text to be *coherent* as perceived by the intended audience means that there is an increased likelihood that they will endorse the proffered (practical or doxastic) commitments *and* that this will require less cognitive effort on the audience's part, by comparison with less coherent texts. The success of a dialectical, user-model oriented text planning regime will clearly depend crucially on the reliability of the user models and the validity of the reasoning processes by which the planner calculates potential challenges and suitable responses. Some important topics are:

- modelling *specific* users to whom a message is directed, versus *typical* readers of a text which is not directed at any particular individual;
- modelling information states of the virtual agents **Pl** and **Cr**, in view of arguments that speakers and hearers have asymmetric context models in dialogue [4].

Complexity. Goal-directed planning requires more computational resources on the part of the Speaker but arguably results in (satellite-initial or mixed) texts that are easier for Hearers to process. The question arises whether speakers optimise their utterances for the audience or follow a path of least effort. This is a topic of debate amongst researchers in psycholinguistics, as evidenced by the claims put forward by [13] and the various responses collected together in the same journal issue.

Preempting clarification requests. This paper has modelled the Background relation as resulting from preemption of a clarification request (CR).) Studies including [5] have shown that CRs can be directed at various levels of linguistic representation or content. In the following example (constructed for this paper), the elliptical query *Maclean?* could have any of the responses shown:

Example (f)

(i)A: Maclean's defected to the USSR.
(ii) B: Maclean?
(iii) A: Yes, Maclean of all people.
(iv) A: Donald Maclean, head of the American desk at the FO.
(v) A: That's M - a - c - l - e - a - n.

This raises architectural issues since it has been assumed in this paper that preemptions are generated at the discourse planning stage, where details of linguistic realisation such as how to spell a proper name may not be available. Future work will address the question of whether and how clarifications at distinct levels of representation can be integrated into the dialectical planning model.

REFERENCES

[1] John Bateman and Michael Zock, 'Natural language generation', in *The Oxford Handbook of Computational Linguistics*, ed., Ruslan Mitkov, 284 – 304, Oxford University Press, Oxford, (2003).

[2] Robert Brandom, *Making it Explicit*, Harvard University Press, Cambridge, Massachusetts and London, 1994.

[3] Philip Cohen and Hector Levesque, 'Persistence, intention and commitment', in *Intentions in Communication*, eds., Philip Cohen, Jerry Morgan, and Martha Pollack, 33 – 69, MIT Press, Cambridge, Massachusetts and London, (1990).

[4] Jonathan Ginzburg, 'On some semantic consequences of turn taking', in *Proceedings of the 11th Amsterdam Colloquium*, (1997).

[5] Jonathan Ginzburg and Robin Cooper, 'Clarification ellipsis and the nature of contextual updates in dialogue', *Linguistics and Philosophy*, **27(3)**, 297 – 365, (2004).

[6] Nancy Green and Sandra Carberry, 'A computational model for taking initiative in the generation of indirect answers', *User Modeling and User-Adapted Interaction*, **9(1/2)**, 93–132, (1999). Reprinted in Computational Models of Mixed-Initiative Interaction, Susan Haller, Alfred Kobsa, and Susan McRoy, eds., Dordrecht, the Netherlands, 277-316.

[7] Charles Hamblin, *Fallacies*, Methuen, London, 1970.

[8] Rodger Kibble, 'Elements of a social semantics for argumentative dialogue', in *Proceedings of the Fourth Workshop on Computational Modelling of Natural Argumentation*, Valencia, Spain, (2004).

[9] Rodger Kibble, 'Reasoning about propositional commitments in dialogue', (2006). To appear in *Research on Language and Computation*.

[10] Rodger Kibble and Richard Power, 'Optimizing referential coherence in text generation', *Computational Linguistics*, **30 (4)**, 401–416, (2004).

[11] William C. Mann and Sandra A. Thompson, 'Rhetorical structure theory: A theory of text organization', Technical report, Marina del Rey, CA: Information Sciences Institute, (1987).

[12] Colin Matheson, Massimo Poesio, and David Traum, 'Modelling grounding and discourse obligations using update rules', in *Proceedings of NAACL 2000*, (2000).

[13] Martin Pickering and Simon Garrod, 'Toward a mechanistic psychology of dialogue', *Behavioral and Brain Sciences*, **27**, 169–225, (2004).

[14] Chris Reed, 'Is it a monologue, a dialogue or a turn in a dialogue?', in *Proceedings of the 4th International Conference on Argumentation (ISSA98)*, Amsterdam, (1998). Foris.

[15] Ehud Reiter, 'Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible?', in *Proceedings of 7th International Natural Language Generation Workshop*, pp. 163–170, (1994).

[16] Maite Tabaoda and William Mann, 'Rhetorical Structure Theory: Looking back and moving ahead', *Discourse Studies*, **8(3)**, (2006). To appear.

[17] Douglas Walton and Eric Krabbe, *Commitment in dialogue*, State University of New York Press, Albany, 1995.

Argumentation in Negotiation Dialogues: Analysis of the Estonian Dialogue Corpus

Mare Koit[1]

Abstract. Estonian spoken dialogues have been analysed with the purpose to model natural argumentation. Calls from an educational company to different institutions are studied where a salesclerk argues for taking a training course by a customer. The study demonstrates that salesclerks try to persuade customers stressing the usefulness of a course in most cases. Our further goal is to model natural dialogue where the computer as a dialogue participant (a salesclerk) follows norms and rules of human-human communication.

1 INTRODUCTION

How do people argue? To answer this question, one has to study corpora that include human-human conversations. Argumentation is used in the dialogues that deal with cooperative problem solving. Let us list some of the most important corpora [6].

The HCRC Maptask Corpus consists of 128 dialogues where participants are marking a route on a map. The TRAINS corpus includes 98 problem solving dialogues where one participant plays the role of a user and has a certain task to accomplish, and another plays the role of the system by acting as a planning assistant. The Monroe corpus contains 20 human-human mixed-initiative, task-oriented dialogues about disaster-handling tasks. The COCONUT corpus includes computer-mediated human-human dialogues in which two subjects cooperate on buying furniture for a house. The Linköping Dialogue Corpus consists of 60 dialogues collected in Wizard of Oz-experiments using two scenarios: car repair and travel. The VERBMOBIL corpus includes bilingual situational dialogues recorded with a role-playing manner (schedule arrangement, hotel, sight seeing). Switchboard is a collection of about 2430 spontaneous conversations between 543 speakers in which the subjects were allowed to converse freely about a given topic.

Dialogue acts and some other phenomena are annotated in the corpora. Different coding schemes are used for various purposes: for annotation and analysis of units of dialogue, to support the design of a dialogue system, to support machine learning of dialogue acts and sequences, theoretical analysis of the pragmatic meanings of utterances. DAMSL (Dialogue Act Markup in Several Layers) is a well-known system for annotating dialogues [3]. A more elaborate version of the SWBD-DAMSL (Switchboard Shallow-Discourse Function Annotation), has been used to code the Switchboard corpus [3]. The Maptask coding scheme is used to annotate transactions, dialogue games and moves in dialogues [1]. The VERBMOBIL corpus uses 18 dialogue acts for annotation of topics.

Our current research is done on the Estonian Dialogue Corpus (EDiC) which contains dialogues of two kinds [2]. The main part of EDiC is made up of spoken human-human dialogues – 715 calls and 116 face-to-face conversations. The remaining part of EDiC – 21 written dialogues – is collected in the Wizard of Oz experiments [7]. We have two purposes collecting the corpus – (1) to study human-human conversations and human-computer interactions, and (2) to develop a DS which interacts with a user in Estonian.

Dialogue acts are annotated in EDiC using a DAMSL-like typology which is based on the conversation analysis approach [2]. According our typology, dialogue acts are divided into two groups: (1) acts that form so-called adjacency pairs (AP) like proposal – agreement (A: *Call me later.* – B: *OK*), and (2) non-AP acts like acknowledgement. The number of the dialogue acts is about 120.

In this paper, we will investigate the conversations where the goal of one partner, A, is to get another partner, B, to carry out a certain action D. Such communication process can be considered as exchange of arguments (and counter-arguments) pro and con of doing D. This type of dialogue forms one kind of so-called agreement negotiation dialogues [8].

Because of this, we have modelled the reasoning processes that people supposedly go through when working out a decision whether to do an action or not. Our model is implemented as an experimental dialogue system and can be used, among other applications, as a "communication trainer".

In our previous paper, calls to a travel agency have been analysed with the aim to find out strategies implemented by a travel agent in order to influence the reasoning processes of a customer to book a trip [4]. It turned out that customers wanted only to get information in most of the analysed calls, and argumentation has been used only in a limited number of cases.

In this paper, we consider the dialogues where a salesclerk of an educational company calls another institution (a manager or another responsible person) and offers courses of his/her company. Both the participants are official persons. We may expect that a salesclerk tries to influence the partner in such a way that (s)he decides to book a course for the employees of his/her institution. Our further goal is to model a salesclerk in a DS.

[1] University of Tartu, Estonia email: mare.koit@ut.ee

The paper is organised as follows. Section 2 gives an overview of our model of conversation agent which includes a reasoning model. In section 3, a corpus analysis is carried out. Section 4 represents an argumentation model which can be used by a conversation agent, and some conclusions are made in section 5.

2 MODELLING COMMUNICATION

In our model, a conversation agent is a program that consists of 6 (interacting) modules (cf. [5]):

(PL, PS, DM, INT, GEN, LP),

where PL – planner, PS – problem solver, DM – dialogue manager, INT – interpreter, GEN – generator, LP – linguistic processor. PL directs the work of both DM and PS, where DM controls communication process and PS solves domain-related tasks. The task of INT is to make semantic analysis of partner's utterances and that of GEN is to generate semantic representations of agent's own contributions. LP carries out linguistic analysis and generation. Conversation agent uses goal base GB and knowledge base KB in its work. A necessary precondition of interaction is existence of shared (mutual) knowledge of agents.

2.1 Reasoning Model

We try to model a "naïve" theory of reasoning, a "theory" that people themselves use when they are interacting with other people and trying to predict and influence their decisions.

The reasoning model consists of two functionally linked parts: 1) a model of human motivational sphere; 2) reasoning schemes. In the motivational sphere three basic factors that regulate reasoning of a subject concerning an action D are differentiated. First, subject may wish to do D, if pleasant aspects of D for him/her overweight unpleasant ones; second, subject may find reasonable to do D, if D is needed to reach some higher goal, and useful aspects of D overweight harmful ones; and third, subject can be in a situation where (s)he must (is obliged) to do D – if not doing D will lead to some kind of punishment. We call these factors WISH-, NEEDED- and MUST-factors, respectively.

The values of the dimension obligatory/prohibited are in a sense absolute: something is obligatory or not, prohibited or not. On the other hand, the dimensions pleasant/unpleasant, useful/harmful have a scalar character: something is pleasant or useful, unpleasant or harmful to a certain degree. For simplicity's sake, it is supposed that these aspects have numerical values and that in the process of reasoning (weighing the pro- and counter-factors) these values can be summed up.

We have represented the model of motivational sphere of a subject by the following vector of weights:

w = (*w(resources), w(pleasant), w(unpleasant), w(useful), w(harmful), w(obligatory), w(prohibited), w(punishment-D), w(punishment-not-D)*).

In the description, *w(pleasant), w(unpleasant), w(useful), w(harmful)* mean weight of pleasant, unpleasant, useful, and harmful aspects of D, *w(punishment-D)* – weight of punishment for doing D if it is prohibited and *w(punishment-not-D)* – weight of punishment for not doing D if it is obligatory. Here *w(resources)* = 1, if subject has resources necessary to do D (otherwise 0); *w(obligatory)* = 1, if D is obligatory for the reasoning subject (otherwise 0); *w(prohibited)* = 1, if D is prohibited (otherwise 0). The values of other weights are non-negative natural numbers.

The second part of the reasoning model consists of reasoning schemes, that supposedly regulate human action-oriented reasoning. A reasoning scheme represents steps that the agent goes through in his/her reasoning process; these consist in computing and comparing the weights of different aspects of D; and the result is the decision to do or not to do D. There are three reasoning procedures in our model which depend on the factor that triggers the reasoning (WISH, NEEDED or MUST).

The reasoning model is connected with the general model of conversation agent in the following way. First, the planner PL makes use of reasoning schemes in order to predict the user's decision and second, the KB contains the vector w^A (A's subjective evaluations of all possible actions) as well as vectors w^{AB} (A's beliefs concerning B's evaluations, where B denotes agent(s) A may communicate with). The vectors w^{AB} are used as partner models.

For the DS, its partner (user) is similarly a conversation agent.

2.2 Communicative Strategies and Tactics

A communicative strategy is an algorithm used by a participant for achieving his/her goal in interaction.

The participant A, having a goal that B will decide to do D, can realize his/her communicative strategy in different ways (using different arguments for): stress pleasant aspects of D (i.e. entice B), stress usefulness of D for B (i.e. persuade B), stress punishment for not doing D if it is obligatory (threaten B). We call communicative tactics these concrete ways of realization of a communicative strategy. Communicative tactics are ways of argumentation. The participant A, trying to direct B's reasoning to the positive decision (to do D), proposes various arguments for doing D while B, when opposing, proposes counter-arguments.

There are three tactics for A in our model which are connected with the three reasoning procedures (WISH, NEEDED, MUST). By tactics of *enticing* the reasoning procedure WISH, by tactics of *persuading* the procedure NEEDED and by tactics of *threatening* the procedure MUST will be tried to trigger in the partner.

In case of institutional communication, both of enticing and threatening can be excluded because a clerk is an official person and (s)he is obligated to communicate cooperatively, impersonally, friendly, peacefully (i.e. to stay in a fixed point of the communicative space). (S)he only can persuade a customer. The general idea underlying the tactic of persuading is that A proposes arguments for usefulness of D trying to keep the weight of usefulness for B high enough and the possible negative values of other aspects brought out by B low enough so that the sum of positive and negative aspects of D would bring B to the decision to do D [5].

3 CORPUS ANALYSIS

For this paper, a closed part of the EDiC has been used, consisting of 44 calls where a salesclerk of an education agency offers different courses of his/her agency (language, book-keeping, secretary treaning etc.) to customers. The dialogues have been put into a secret list on the ethical reasons, according to an agreement with the company.

14 dialogues out of 44 are excluded from the current study because they do not include argumentation at all (the needed person is not present, the number the clerk is calling is wrong, the recording breaks off). The remaing 30 dialogues can be divided into two groups: 1) the salesclerk (A) and the manager or personell administrator (B) of another organization are communicating for the first time (6 dialogues), 2) they have been in the contact previously (24 dialogues). The action D is ’to book the offered course’.

A call consists of three parts: (1) a ritual beginning, (2) the main part which starts with A’s proposal and ends with B’s agreement or rejection, (3) a ritual ending.

3.1 The first contact

Let us start with considering the dialogues where the participants are communicating for the first time. The average length of these dialogues is 88 turns (min 54 and max 113 turns). In two dialogues, the salesclerk starting a conversation points another person from the same institution who has recommended just that person.

A typical dialogue starts with A’s introduction and a question whether B does know the education company. Then a short overview of the company is given (e.g. *we are an international company, we are acting six years in Estonia, we are dealing with sale, service, management, marketing*). All the statements can be considered as arguments for taking a training course. Then a proposal is made by A to take some courses. A points the activities of B’s organisation which demonstrates that (s)he has previous knowledge about the institution (e.g. *your firm is dealing with retail and whole sale, therefore you could be interested in our courses*, Ex[2] 1). If B does not make a decision then A asks B to tell more about B’s institution in order to get more arguments for necessity of the courses for B, and offers them again.

(1)

```
A:     ja no Ti- Tiritamm pakub just nüd ka
sellist sellist koolitust et kuidas kuidas neid
(0.5) mm kliente nüd
and Tiritamm offers just such such a training
how how [to find] customers
(1.8)
leida eks=ole, oma turgu
to find, yes, [to increase] your own market
(1.5)
e suurendada. ja (0.8) ja (0.5) ja samas ka see
et=et kuidas neid püsikliente ´hoida (1.0) kas
e (…) suhtlemist et. kuidas teiele tundub kas
ned teemad võiksid teile huvi pakkuda?
to increase, and how to keep regular customers.
how do you think – are you interested in that
themes?
```

All the dialogues end with an agreement to keep the contact (A promises to send information materials to B, call B later), B does not decide to accept nor reject a course but postpones the decision. Still, that can be considered as a good result for A, it shows that his/her arguments were reasonable. B needs some time for reasoning, weighing positive and negative aspects of D.

3.2 Continuining communication

Most of the analysed dialogues represent the situations where A and B have been in contact before. B has had the time to evaluate the information about the courses in order to make a decision). The average length of such dialogues is 94 turns (min 12, max 264 turns). Therefore, these dialogues are in general longer than the first conversations. B agrees to take a course only in one conversation, (s)he agrees with reservations in two dialogues, and does not agree in one dialogue. In the remaining dialogues, A and B come to the agreement to keep the contact like in the case of the first communication. So, B postpones the decision. A always starts the conversation with pointing to a previous contact (*we communicated in November*, *I sent catalogues to you – did you receive them*, *which decision did your direction make*, Ex 2).

(2)

```
A:     ´kevadel rääkisime natuke ´pikemalt sin
(.) ´viimati. (.) et e (.) kudas teil ´läheb
ka? (.)
we talked in the spring quite long the last
time, how do you do?
```

It is significant that the introductory part is quite long in the dialogues. A behaves very politely, friendly and sometimes familiarly (this holds especially for male clerks), Ex 3.

(3)

```
A:     mt (.) kuidas on elu ´vahepeal läinud,
kõik kenad ´reisid on ´seljataha jäänud.
how did you do meanwhile, all the nice trips
are remained behind?
```

In this way, A prepares the background for his/her proposal and herewith makes a refusal more difficult for B, Ex 4.

(4)

```
B:     [jah väga meeldiv.] tähendab ä nüd on
nimodi=et selleks ´suureks ´koolituseks me .hh
(0.8) otsustasime: ühe ´teise firma kasuks. .hh
küll aga ma sooviksin regist´reerida sis sinna
´juhtide avalikule m esinemis´kursusele nüd ühe
´inimese.
yes, very nice. it means that it is so that we
decided for another firm for the long training
but I’d like to register one person to the
public performance training course
```

In the main part of a dialogue, A gives various arguments for the usability of the courses for B’s institution and meanwhile collects new information by asking questions in order to learn more about it and have new arguments for doing D (Ex 5,6).

(5)

```
A:  ee küsiks nüd ´seda et=et ta on (.) noh
põhimõtselt mõeldud ütleme mt (.) e ´juhtidele
ja ´spetsialistidele ütleme kes ´vastutavad
´rahvusvaheliste kontaktide ´arendamise eest.
I’d like to ask that, it is designed for
managers in general and for the specialists who
are responsible for development of
international contacts
B:  mhmh.
hem
A:  a kas teil on ´rahvusvahelisi ´suhteid,
but do you have international relations?
B:  mm=
hem
```

(6)

```
A: e on nad selealast ´koolitust ka ´saanud,
```

[2] Transcription of conversation analysis is used in the examples.

did they obtain a (language) training too?
B: ee üldselt ´mitte (.) @ täendap ´mina ei ole inglise keelt ´kunagi ´kusagil ´õppinud. @
no, in general, it means, I have never learned English
A: ahaa
aha!

4 MODELLING ARGUMENTATION

The tactic of persuasion based on the reasoning procedure NEEDED (cf. above) is implemented in our model of conversation agent (Fig. 1). When persuasing B, A tries to indicate useful aspects of D in such a way that the usefulness of D would go greater than its harmfulness and B therefore would trigger the reasoning procedure NEEDED [5].

```
WHILE B is not agreeing AND A is not giving up DO
  CASE B’s answer of
    no resources :
present a counter-argument in order to point at the possibility to gain the resources, at the same time showing that the cost of gaining these resources is lower than the weight of the usefulness of D
    much harm :
present a counter-argument to decrease the value of harmfulness in comparison with the weight of usefulness
    much unpleasant :
present a counter-argument in order to downgrade the unpleasant aspects of D as compared to the useful aspects of D
    D is prohibited and the punishment is great :
present a counter-argument in order to downgrade the weight of punishment as compared to the usefulness of D
  END CASE
Present an argument to stress the usefulness of D.
```

Fig. 1. Persuasion (author – A, addressee – B).

If B when opposing indicates other aspects of D then A reacts them but in addition tries to direct B’s reasoning to the relationship of usefulness and harmfulness of D. For example, if B indicates that the resources for doing D are missing then A answers with an argument which explains how to gain them and that it does not cost much (Ex 7).

(7)
B: .hh meil ei ole ´praegu eriti: ´ruumi vel põhimõtteliselt meie ainukene ´õppe ´klass on tehtud ´arvutiklassiks
/---/
we have no room at the moment, our single classroom has been changed to a computer room
/---/
A: [jajaa] a´haa /--/ et noh oleks võimalik võtta ka ütme ´tulla (.) ´meile seda tegema et=see ühe ruumi üür ei ole eriti=eriti ´soolane
yes, yes, aha, it would be possible to take, let me say to come to us to make it, the room rent is not very salty
B: [((yawns))]

In institutional negotiation dialogues, persuasion (mainly) operates with usefulness, harmfulness and resources of doing D. There are no examples in our corpus where B would indicate that D is unpleasant or prohibited.

An experimental dialogue system has being implemented which can play the role of both A or B in interaction with a user. At the moment the computer operates with semantic representations of linguistic input/output only, the surface linguistic part of interaction is provided in the form of a list of ready-made utterances (sentences in Estonian) which are used both by the computer and user. Our implementation represents just a prototype realisation of our theoretical ideas and we are working on refining it.

5 CONCLUSION

We investigated the conversations where the goal of one partner, A, is to get another partner, B, to carry out a certain action D. Because of this, we have modelled the reasoning processes that people supposedly go through when working out a decision whether to do an action or not.

The goal of this paper was to verify our argumentation model on Estonian spoken human-human dialogues. Calls of salesclerks of an educational company were analysed in order to find out how clerks try to bring customers to a decision to take a training course. Various arguments are used by the clerks to stress usefulness of courses for customers. Still, customers seldom agree to take a course. In most cases, a decision will be postponed.

Our next aim is to investigate these dialogues from the point of view of customers. We will try to find out the ways of argumentation which are used by customers who avoid making a final decision.

ACKNOWLEDGEMENTS

This work is supported by the Estonian Science Foundation (grant number 5685).

REFERENCES

[1] J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon and A. Anderson, *The reliability of a dialogue structure coding scheme.* Computational Linguistics, 23(1):13–31, (1997).

[2] O. Gerassimenko, T. Hennoste, M. Koit, A. Rääbis, K. Strandson, M. Valdisoo and E. Vutt, *Annotated dialogue corpus as a language resource: an experience of building the Estonian dialogue corpus.* The first Baltic conference “Human language technologies. The Baltic perspective”. Commission of the official language at the chancellery of the president of Latvia, 150–155, Riga, (2004).

[3] D. Jurafsky and J.H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics.* Prentice-Hall, (2000).

[4] M. Koit, *Argumentation in Institutional Dialogues: Corpus Analysis.* IJCAI-05 Workshop on Computational Models of Natural Argument. Working Notes. Ed. Chris Reed. Edinburgh, Scotland, 80-83, (2005).

[5] M. Koit and H. Õim, *Dialogue management in the agreement negotiation process: a model that involves natural reasoning* . The 1st SIGdial Workshop on Discourse and Dialogue. Ed. L. Dybkjaer, K. Hasida, D. Traum. HongKong, 102-111, (2000).

[6] M.F. McTear. *Spoken Dialogue Technology. Toward the Conversational User Interface.* Springer, (2004).

[7] M. Valdisoo, E. Vutt and M. Koit, *On a method for designing a dialogue system and the experience of its application.* Journal of Computer and Systems Sciences International, 42(3), 456–464, (2003).

[8] T. Yuan, D. Moore and Alec Grierson, *Human-Computer Debate: a Computational Dialectics Approach.* Proc. of CMNA-02. http://www.csc.liv.ac.uk/~floriana/CMNA/YuanMooreGrierson.pdf

Agent-based Argumentation for Ontology Alignments

Loredana Laera and **Valentina Tamma** and **T.J.M. Bench-Capon**[1] and **Jérôme Euzenat** [2]

Abstract. When agents communicate they do not necessarily use the same vocabulary or ontology. For them to interact successfully they must find correspondences between the terms used in their ontologies. While many proposals for matching two agent ontologies have been presented in the literature, the resulting alignment may not be satisfactory to both agents and can become the object of further negotiation between them.

This paper describes our work constructing a formal framework for reaching agents' consensus on the terminology they use to communicate. In order to accomplish this, we adapt argument-based negotiation used in multi-agent systems to deal specifically with arguments that support or oppose candidate correspondences between ontologies. Each agent can decide according to its interests whether to accept or refuse the candidate correspondence. The proposed framework considers arguments and propositions that are specific to the matching task and related to the ontology semantics. This argumentation framework relies on a formal argument manipulation schema and on an encoding of the agents preferences between particular kinds of arguments. The former does not vary between agents, whereas the latter depends on the interests of each agent. Therefore, this work distinguishes clearly between the alignment rationales valid for all agents and those specific to a particular agent.

1 Introduction

When agents transfer information, they need a conceptualisation of the domain of interest and a shared vocabulary to communicate facts with respect to this domain. The conceptualisation can be expressed in a so-called *ontology*. An ontology abstracts the essence of the domain of interest and helps to catalogue and distinguish various types of objects in the domain, their properties and relationships (see, e.g. [14]). An agent can use such a vocabulary to express its beliefs and actions, and so communicate about them. Ontologies thus contribute to semantic interoperability when agents are embedded in open, dynamic environments, such as the Web, and its proposed extension the Semantic Web [7]. It has long been argued that in this type of environment there cannot be a single universal shared ontology, that is agreed upon by all the parties involved, as it would result in imposing a standard communication vocabulary. Interoperability therefore relies on the ability to reconcile different existing ontologies that may be heterogeneous in format and partially overlapping [22]. This reconciliation usually exists in the form of correspondences (or mapping) between agent ontologies and to use them in order to interpret or translate messages exchanged by agents. The underlying problem is usually termed an *ontology alignment* problem [13].

There are many matching algorithms able to produce such alignments [17]. In general, alignments can be be generated by trustable alignment services that can be invoked in order to obtain an alignment between two or more ontologies, and use it for translating messages [12]. Alternatively, they can be retrieved from libraries of alignments. However, the alignments provided by such services may not suit the needs of all agents. Indeed agents should be able to accept or refuse a proposed correspondence according to their own interests. In order to address this problem, we develop a formal framework for reaching agents consensus on the terminology they need to use in order to communicate. The framework allows agents to express their preferred choices over candidate correspondence. This is achieved adapting argument-based negotiation used in multi-agent systems to deal specifically with arguments that support or oppose the proposed correspondences between ontologies. The set of potential arguments are clearly identified and grounded on the underlying ontology languages, and the kind of mapping that can be supported by any one argument is clearly specified.

In order to compute preferred alignments for each agent, we use a value-based argumentation framework [5] allowing each agent to express its preferences between the categories of arguments that are clearly identified in the context of ontology alignment.

Our approach is able to give a formal motivation for the selection of any correspondence, and enables consideration of an agents interests and preferences that may influencethe selection of a correspondence.

Therefore, this work provides a concrete instantiation of the meaning negotiation process that we would like agents to achieve. Moreover, in contrast to current ontology matching procedures, the choice of an alignment is based on two clearly identified elements: (i) the argumentation framework, which is common to all agents, and (ii) the preference relations which are private to each agent.

The remainder of this paper is structured as follows. Section 2 defines the problem of reaching agreement over ontology alignments among agents. In section 3 we present in detail the argumentation framework and how it can be used. Section 4 defines the notion of agreeable alignments for two agents, and proposes a procedure to find these agreeable alignments. Next, in section 5, an example is provided to illustrate the idea. Section 6 points out some related work. Finally, section 7 draws some concluding remarks and identifies directions for further exploration.

2 Reaching agreement over ontology alignments

Before describing the framework, we first need to delimit the problem of reaching agreement over ontology alignments and state the assumptions upon which we build the theoretical framework.

In this paper, we concentrate on agents situated in a system, that need to display *social ability* and communicate in order to carry out some task. Each agent has a name, a role and a knowledge base. In some agent models, the basic knowledge base of an agent may be

[1] University of Liverpool,email: lori,valli,tbc@csc.liv.ac.uk
[2] INRIA Rhône-Alpes, email:Jerome.Euzenat@inrialpes.fr

consist of a set of beliefs, a set of desires and a set of intentions. However, for the purpose of this paper, we do not need to distinguish between beliefs, desire and intentions, and we will simply assume that an agent has a knowledge base where it stores facts about the domain it knows (which correspond to an ontology). Moreover, we do not make explicit use of the agent role.

Ontology can be defined as a tuple [11] $\langle C, H_C, R_C, H_R, I, R_I, A^O \rangle$, where the concepts C are arranged in a subsumption hierarchy H_C. Relations R_C is a set of relation between single concepts. Relations (or properties) can also be arranged in a hierarchy H_R. Instances I of a specific concept are interconnected by property instances R_I . Axioms A^O can be used to infer knowledge from that already existing. We further assume that ontologies are encoded in the same language, the standard OWL[3], removing us from the problem of integrating the ontology languages.

In order for agents to communicate, they need to establish alignments between their ontologies. We assume that such an alignment is generated by an alignment service agent and consists of a set of correspondences. A correspondence (or a mapping) can be described as a tuple: $\langle e, e', R \rangle$, where e and e' are the entities (concepts, relations or individuals) between which a relation is asserted by the correspondence; and R is the relation (e.g., equivalence, more general, etc.), holding between e and e', asserted by the correspondence [17]. For example a equivalence correspondence will stand between the concept 'car' in an ontology O and the concept 'automobile' in an ontology O' . A correspondence delivered by such an algorithm and not yet agreed by the agents will be called a *candidate mapping*. Note that we assume that an alignment service agent is able to generate an alignment using an independently defined decision-making process. We make no assumptions about how the agents achieve such decisions, as this is an internal agent process separate from the argumentation framework we present here.

Therefore, let two autonomous agents be committed to two ontologies O and O'. The *reaching agreement* problem is defined as follows:

Definition *"Find an agreement on the correspondences between the vocabularies they use, expressed as an ontology alignment."*.

Figure 1 illustrates the situation. Note that the definition consider two agents that want to communicate, but it can easily be extended to multi-agent systems. It is noteworthy that the process of reaching agreement should be as automatic as possible and should not require any feedback from human users. Indeed, essential to our approach, is that ontological discrepancies are treated at the level of agents themselves, without the aid of an external observer. The framework accounts for the detection and handling of ontological discrepancies by the agents themselves, on the basis of their own subjective view on the world. Agents should work towards agreement on the basis of their interest and preference states. We believe that this approach is both theoretically and practically important for agent systems.

In the next section, we show how this can be achieved using argumentation. Note that the framework requires that agents are able to justify why they have selected a particular mapping when challenged, since they will exchange arguments supplying the reasons for such a choice.

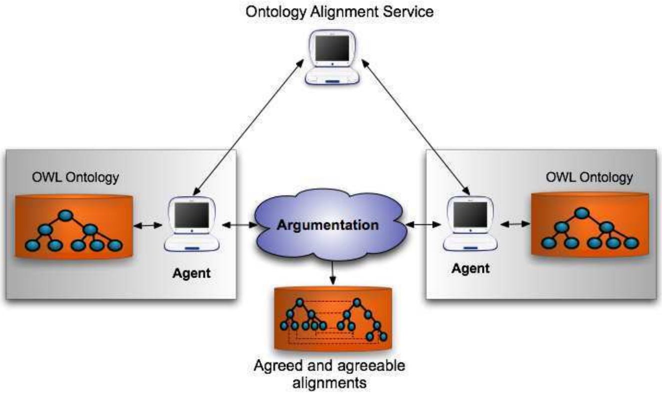


Figure 1. Reaching agreement over ontology alignments

3 Argumentation Framework

In order for the agents to consider potential mappings and the reasons for and against accepting them we use an argumentation framework. Our framework is based on the Value-based Argument Frameworks (VAFs) [5], a development of the classical argument systems of Dung [9]. We start with the presentation of Dung's framework, upon which the Value-based Argument Frameworks (VAFs) rely.

3.1 Classical argumentation framework

Definition An Argumentation Framework (AF) is a pair $AF = \langle AR, A \rangle$, where AR is a set of arguments and $A \subset AR \times AR$ is the *attack* relationship for AF. A comprises a set of ordered pairs of distinct arguments in AR. A pair $\langle x, y \rangle$ is referred to as "x *attacks* y". We also say that a set of arguments S attacks an argument y if y is attacked by an argument in S.

An argumentation framework can be simply represented as a directed graph whose vertices are the arguments and edges correspond to the elements of R. In Dung's work arguments are atomic and cannot be analysed further. In this paper, however, we are concerned only with arguments advocating mappings. We can therefore define arguments as follows:

Definition An argument $x \in AF$ is a triple $x = \langle G, m, \sigma \rangle$ where:

- m is a correspondence $\langle e, e', R \rangle$
- G is the grounds justifying a prima facie belief that the mapping does, or does not hold;
- σ is one of $\{+, -\}$ depending on whether the argument is that m does or does not hold.

When the set of such arguments and counter arguments have been produced, it is necessary to consider which of them should be accepted. Given an argument framework we can use definitions from [9] to define acceptability of an argument.

Definition Let $\langle AR, A \rangle$ be an argumentation framework. For R and S, subsets of AR, we say that:

- An argument $s \in S$ is attacked by R if there is some $r \in R$ such that $\langle r, s \rangle \in A$.
- An argument $x \in AR$ is *acceptable* with respect to S if for every $y \in AR$ that attacks x there is some $z \in S$ that attacks y.

[3] http://www.w3.org/OWL/

- S is *conflict free* if no argument in S is attacked by any other argument in S.
- A conflictfree set S is *admissible* if every argument in S is acceptable with respect to S.
- S is a *preferred extension* if it is a maximal (with respect to set inclusion) admissible subset of AR.

An argument x is *credulously accepted* if there is *some* preferred extension containing it; x is *sceptically accepted* if it is a member of *every* preferred extension.

The key notion here is the *preferred extension* which represents a consistent position within AF, which is defensible against all attacks and which cannot be further extended without becoming inconsistent or open to attack.

In Dung's framework, attacks always succeed. This is reasonable when dealing with deductive arguments, but in many domains, including the one under consideration, arguments lack this coercive force: they provide reasons which may be more or less persuasive. Moreover, their persuasiveness may vary according to their audience. To handle such defeasible reason giving arguments we need to be able to distinguish attacks from successful attacks, those which do defeat the attacked argument. One approach, taken in [1], is to rank arguments individually: an alternative, which we follow here, is to use a Value Based Argumentation framework (VAF) [5] which describes different strengths to arguments on the basis of the values they promote, and the ranking given to these values by the audience for the argument. This allows us to systematically relate strengths of arguments to their motivations, and to accommodate different audiences with different interests and preferences. VAFs are described in the next sub-section.

3.2 Value-based argumentation framework

We use the *Value-Based Argumentation Frameworks* (VAF) of Bench-Capon [5], to determine which mappings are acceptable, with respect to the different *audiences* represented by the different agents:

Definition A *Value-Based Argumentation Framework* (VAF) is defined as $\langle AR, A, \mathcal{V}, \eta \rangle$, where (AR, A) is an argumentation framework, $\mathcal{V}$ is a set of k *values* which represent the types of arguments and η: $AR \rightarrow \mathcal{V}$ is a mapping that associates a value $\eta(x) \in \mathcal{V}$ with each argument $x \in AR$

Definition An *audience* for a VAF is a binary relation $\mathcal{R} \subset \mathcal{V} \times \mathcal{V}$ whose (irreflexive) transitive closure, $\mathcal{R}^*$, is asymmetric, i.e. at most one of (v, v'), (v', v) are members of $\mathcal{R}^*$ for any distinct v, $v' \in \mathcal{V}$. We say that v_i *is preferred to* v_j *in the audience* $\mathcal{R}$, denoted $v_i \succ_{\mathcal{R}} v_j$, if $(v_i, v_j) \in \mathcal{R}^*$.

Let $\mathcal{R}$ be an audience, α is a *specific audience*(compatible with $\mathcal{R}$) if α is a *total ordering* of $\mathcal{V}$ *and* $\forall\ v,\ v' \in \mathcal{V}\ (v, v') \in \alpha \Rightarrow (v', v) \notin \mathcal{R}^*$

In this way, we take into account that different audiences (different agents) can have different perspectives on the same candidate mapping. [5] defines acceptability of an argument in the following way. Note that all these notions are now relative to some audience.

Definition Let $\langle AR, A, \mathcal{V}, \eta \rangle$ be a VAF and $\mathcal{R}$ an audience.

a. For arguments x, y in AR, x is a *successful attack* on y (or x *defeats* y) *with respect to the audience* $\mathcal{R}$ if: $(x, y) \in \mathcal{A}$ *and* it is *not* the case that $\eta(y) \succ_{\mathcal{R}} \eta(x)$.

b. An argument x is *acceptable to the subset* S with respect to an audience $\mathcal{R}$ if: for every $y \in AR$ that *successfully attacks* x with respect to $\mathcal{R}$, there is some $z \in S$ that successfully attacks y with respect to $\mathcal{R}$.

c. A subset S of AR is *conflict-free with respect to the audience* $\mathcal{R}$ if: for each $(x, y) \in S \times S$, either $(x, y) \notin \mathcal{A}$ or $\eta(y) \succ_{\mathcal{R}} \eta(x)$.

d. A subset S of AR is *admissible* with respect to the audience $\mathcal{R}$ if: S is conflictfree with respect to $\mathcal{R}$ and every $x \in S$ is acceptable to S with respect to $\mathcal{R}$.

e. A subset S is a *preferred extension* for the audience $\mathcal{R}$ if it is a maximal admissible set with respect to $\mathcal{R}$.

f. A subset S is a *stable extension* for the audience $\mathcal{R}$ if S is admissible with respect to $\mathcal{R}$ and for all $y \notin S$ there is some $x \in S$ which successfully attacks y with respect to $\mathcal{R}$.

In order to determine whether the dispute is resoluble, and if it is, to determine the preferred extension with respect to a value ordering promoted by distinct audiences, [5] introduce the notion of objective and subjective acceptance as follows.

Definition *Subjective Acceptance.* Given an VAF, $\langle AR, A, \mathcal{V}, \eta \rangle$, an argument $x \in AR$ is subjectively acceptable if and only if, x appears in the preferred extension for some specific audiences but not all.

Definition *Objective Acceptance.* Given an VAF, $\langle AR, A, \mathcal{V}, \eta \rangle$, an argument $x \in AR$ is objectively acceptable if and only if, x appears in the preferred extension for *every* specific audience.

An argument which is neither objectively nor subjectively acceptable is said to be *indefensible*. These definitions are particularly of interest in the case of the universal audience: subjective acceptability indicating that there is *at least one* specific audience (total ordering of values) under which x is accepted; objective acceptability that x must be accepted irrespective of the value ordering described by a specific audience; and, in contrast, x being indefensible indicating that no specific audience can ever accept$x$.

4 Arguing about correspondences

Our goal is to take advantage of value based argumentation so that agents can find the most mutually acceptable alignment. Section 4.1 defines the various categories of arguments that can support or attack mappings . Section 4.2 defines the notion of agreed and agreeable alignments for agents. Finally, in section 4.3 we demonstrate how the argumentation frameworks are constructed, in order to find such agreed and agreeable alignments.

4.1 Categories of arguments for correspondences

As we mentioned in Section 1, potential arguments are clearly identified and grounded on the underlying ontology languages, and the language of choice is the *de-facto* standard, OWL. Therefore, the grounds justifying correspondences can be extracted from the knowledge in ontologies. This knowledge includes both the extensional and intensional OWL ontology definitions. Our classification of the grounds justifying correspondences is the following:

semantic (M): the sets of models of some expressions do or do not compare;

internal structural (IS): the two entities share more or less internal structure (e.g., the value range or cardinality of their attributes);

external structural (ES): the set of relations of two entities with other entities do or do not compare;

terminological (T): the names of entities share more or less lexical features;

extensional (E): the known extension of entities do or do not compare.

These categories correspond to the type of categorizations underlying matching algorithms [22].

In our framework, we will use the types of arguments mentioned above as types for the value-based argumentation; hence $\mathcal{V} = \{M, IS, ES, T, E\}$. Therefore, for example, an audience may specify that terminological arguments are preferred to semantic arguments, or vice versa. Note that this may vary according to the nature of the ontologies being aligned. Semantic arguments will be given more weight in a fully axiomatised ontology rather than in a lightweight ontology where there is very little reliable semantic information on which to base such arguments.

The reader may find it interesting to refer to the table 2, which summarises a number of reasons capable of justifying candidate OWL ontological alignments. Therefore, the table represents an (extensible) set of argument schemes, instantiations of which will comprise AR. Attacks between these arguments will arise when we have arguments for the same mapping but with different signs, thus yielding attacks that can be considered symmetric. Moreover the relations in the mappings can also give rise to attacks: if relations are not deemed exclusive, an argument against inclusion is a fortiori an argument against equivalence (which is more general).

Example Consider a candidate mapping $m = \langle c, c', \_, \equiv\rangle$ between two OWL ontologies O_1 and O_2, with concepts c and c' respectively. A list of arguments for or against accepting the mapping m, may be:

- The labels of the concept c and c' are synonymous.
  $\langle label(c) \approx label(c'), m, +\rangle$ (Terminological)
- Some of their instances are similar.
  $\langle E(c) \cap E(c') \neq \emptyset, m, +\rangle$ (Extensional)
- Some of their properties are similar.
  $\langle properties(c) \cap properties(c') \neq \emptyset, m, +\rangle$ (Internal Structural)
- Some of the super-classes of c and c' are dissimilar
  $\langle S(c) \cap S(c') = \emptyset, m, -\rangle$. (External Structural)

Similar arguments can be made for and against cases in which we consider properties or instances.

Therefore, in VAF arguments against or in favour of a candidate mapping, are seen as grounded on their type. In this way, we are able to motivate the choice between preferred extensions by reference to the type ordering of the audience concerned.

4.2 Agreed and agreeable alignments

Although in $VAFs$ there is always a unique non-empty preferred extension with respect to a specific audience, provided the AF does not contain any cycles in a single argument type, an agent may have multiple preferred extensions either because no preference between two values in a cycle has been expressed, or because a cycle in a single value exists. The first may be eliminated by committing to a more specific audience, but the second cannot be eliminated in this way. In our domain, where many attacks are symmetric, two cycles will be frequent and in general an audience may have multiple preferred extensions.

Thus given a set of arguments justifying mappings organised into an argumentation framework, an agent will be able to determine which mappings are acceptable by computing the preferred extensions with respect to its preferences. If there are multiple preferred extensions, the agent must commit to the arguments present in all preferred extensions, but has some freedom of choice with respect to those in some but not all of them. This will partition arguments into three sets: *desired arguments*, present in all preferred extensions, *optional arguments*, present in some but not all, and *rejected arguments*, present in none. If we have two agents belonging to different audiences, these sets may differ. [8] describes a means by which agents may negotiate a joint preferred extension on the basis of their partitioned arguments so to maximise the number of desired arguments included while identifying which optional arguments need to be included to support them.

Based on these above considerations, we thus define an *agreed alignment* as the set of correspondences supported [4] by those arguments which are in every preferred extension of every agent, and an *agreeable alignment* extends the agreed alignment with the correspondences supported by arguments which are in some preferred extension of every agent. The next section shows how the argumentation frameworks are constructed.

4.3 Constructing argumentation frameworks

Given a single agent, we could construct an argumentation framework by considering the repertoire of argument schemes available to the agent, and constructing a set of arguments by instantiating these schemes with respect to the interests of the agent. Having established the set of arguments, we then determine the attacks between them by considering their mappings and signs, and the other factors discussed above.

If we have multiple agents, we can simply merge their individual frameworks by forming the union of their individual argument sets and individual attack relations, and then extend the attack relation by computing attacks between the arguments present in the framework of one, but not both, agents. We employ the algorithm in [4] for computing the preferred extensions of a value-based argumentation framework given a value ordering. The global view is considered by taking the union of these preferred extensions for each audience. Then, we consider which arguments are in every preferred extension of every audience. The mappings that have only arguments for will be included in the agreed alignments, and the mappings that have only arguments against will be rejected. For those mappings where we cannot establish their acceptability, we extend our search space to consider those arguments which are in some preferred extension of every audience. The mappings supported by those arguments are part of the set of agreeable alignments. Algorithm 1 shows how to find such agreed and agreeable alignments.

The dialogue between agents can thus consist simply of the exchange of individual argumentation frameworks, from which they can individually compute acceptable mappings. If necessary and desirable, these can then be reconciled into a mutually acceptable position through a process of negotiation, as suggested in [8] which defines a dialogue process for evaluating the status of arguments in a VAF, and shows how this process can be used to identify mutually acceptable arguments. In the course of constructing a position, an ordering of values best able to satisfy the joint interests of the agents concerned is determined.

[4] Note that a correspondence m is *supported* by an argument x if x is $\langle G, m, +\rangle$

Algorithm 1 Find agreed and agreeable alignments

```
Require: a set of $VAFs$ $\langle AR, A, \mathcal{V}, \eta \rangle$, a set of audiences $\mathcal{R}_i$, a set
    of candidate mappings $M$
Ensure: Agreed alignments $AG$ and agreeable alignments $AG_{ext}$
$AG$:=$\emptyset$
$AG_{ext}$:=$\emptyset$
for all audience $\mathcal{R}_i$ do
    for all $VAF$ do
        compute the preferred extensions for $\mathcal{R}_i$,
        $P_j(\langle AR, A, \mathcal{V}, \eta \rangle, \mathcal{R}_i)$, $j \geq 1$
    end for
    $P_k(\mathcal{R}_i)$:=$\bigcup_j P_j(\langle AR, A, \mathcal{V}, \eta \rangle, \mathcal{R}_i)$, $k \geq 1$
end for
$AGArg$:=$x \in \bigcap_{k,i} P_k(\mathcal{R}_i)$, $\forall k \geq 1$, $\forall i \geq 0$
for all $x \in AGArg$ do
    if $x$ is $\langle G, m, + \rangle$ then
        $AG := AG \cup \{m\}$
    else
        reject mapping $m$
    end if
end for
if $\exists$ $m \in M$ such that $m$ is neither in $AG$ and rejected then
    $AGArg_{ext}$:=$x \in \bigcap_i P_k(\mathcal{R}_i)$, $\forall i \geq 0$,$k \geq 1$
    for all $x \in AGArg_{ext}$ do
        if $x$ is $\langle G, m, + \rangle$ then
            $AG_{ext} := AG_{ext} \cup \{m\}$
        end if
    end for
end if
```

The above technique considers sets of mappings and complete argumentation frameworks. If instead the problem is to determine the acceptability of a single mapping it may be more efficient to proceed by means of a dialectical exchange, in which a mapping is proposed, challenged and defended. Argument protocols have been proposed in e.g. [15]. Particular dialogue games have been proposed based on Dung's Argumentation Frameworks (e.g. [10]), and on VAFs [6].

5 A walk through example

Having described the framework, we will go through an practical example.

Let us assume that some agents need to interact with each others using two independent but overlapping ontologies. One ontology is the bibliographic ontology[5] from the University of Canada, based on the bibTeX record. The other is the General University Ontology[6] from the French company Mondeca[7]. For space reasons, we will only consider a subset of these ontologies, shown in figure 2 and figure 3, where the first and second ontologies are represented byO_1 and O_2 respectively.

We will reason about the following candidate mappings:
m_1=$\langle O_1$:$Press, O_2$:$Periodical, \_, = \rangle$,
m_2=$\langle O_1$:$publication, O_2$:$Publication, \_, = \rangle$,
m_3=$\langle O_1$:$hasPublisher, O_2$:$publishedBy, \_, = \rangle$,

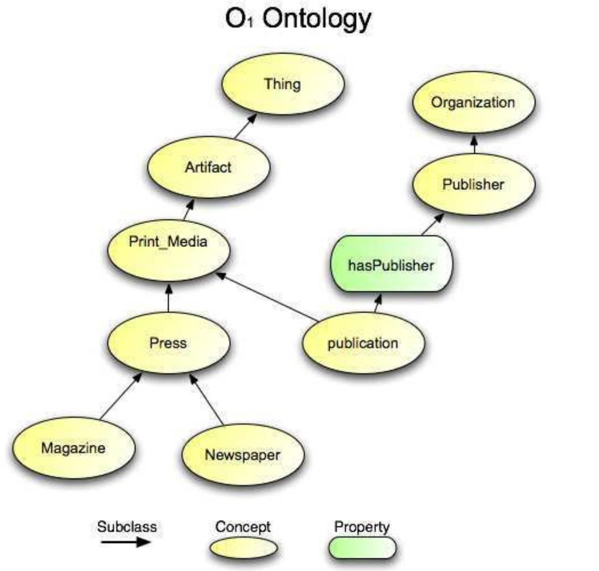


Figure 2. Ontology O_1

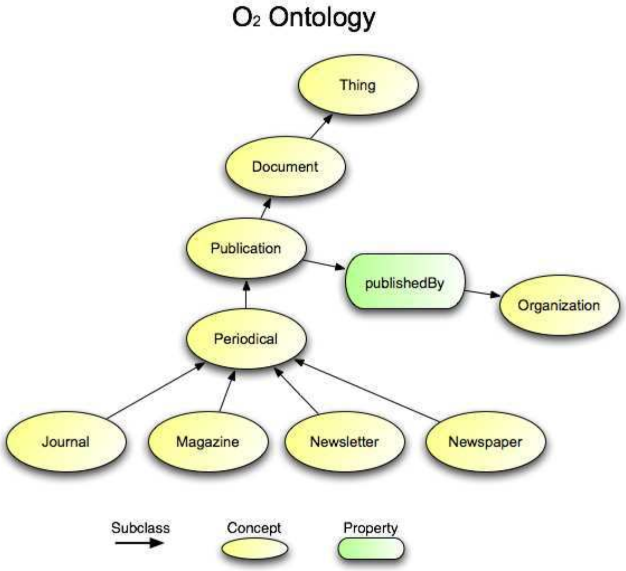


Figure 3. Ontology O_2

The following mappings are taken to be already accepted:

m_4=$\langle O_1$:$Magazine, O_2$:$Magazine, \_, = \rangle$,
m_5=$\langle O_1$:$Newspaper, O_2$:$Newspaper, \_, = \rangle$
m_6=$\langle O_1$:$Organization, O_2$:$Organization, \_, = \rangle$,

We begin by identifying a set of arguments and the attacks between them. This is achieved by instantiating the argumentation schemes, discussed previously, with respect to the interests of the agent. Table 1 shows each argument, labeled with an identifier , its type, and the attacks that can be made on it by opposing agents.

[5] http://www.cs.toronto.edu/semanticweb/maponto/ontologies/BibTex.owl
[6] http://www.mondeca.com/owl/moses/univ.owl
[7] Note that ontology O_2 has been slightly modified for the purposes of this example.

Based upon these arguments and the attacks, we can construct the argumentation frameworks which bring the arguments together so that they can be evaluated. These are shown in Figure 4, where nodes represent arguments, with the respective type value, and arcs represent the attacks. Now we can look in more detail at each argumentation framework.

In the argumentation framework (a), we have two arguments against m_1, and one for it. A is against the correspondence m_1, since none of the super-concepts of the $O_1\!:Press$ are similar to any super-concept of $O_2\!:Periodical$. B argues for m_1 because two sub-concepts of $O_1\!:Press$, $O_1\!:Magazine$ and $O_1\!:Newspaper$, are similar to two sub-concepts of $O_2\!:Periodical$, $O_1\!:Magazine$ and $O_1\!:Newspaper$, as established by m_4 and m_5. C pleads against m_1, because $Press$ and $Periodical$ do not have any lexical similarity.

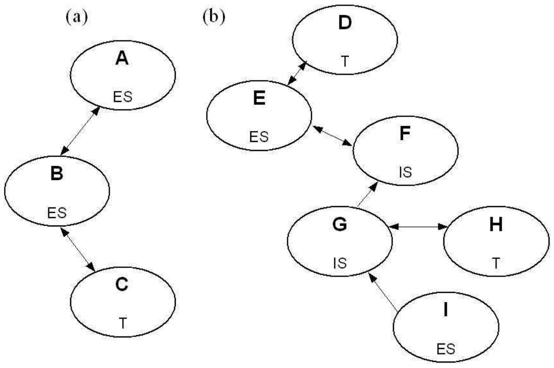


Figure 4. Value-Based Argumentation Frameworks

In the second argumentation framework (b) we relate the following arguments: D justifies the mapping m_2, since the labels of $O_1\!:publication$ and $O_2\!:Publication$ are lexically similar. Their super-concepts, however, are not similar (E). Argument F is based on the fact that $O_1\!:publication$ and $O_2\!:Publication$ have similar properties, $O_1\!:hasPublisher$ and $O_1\!:publishedBy$, as defined inm_3. F is then attacked by G, which states that the range of these properties, $O_1\!:Publisher$ and $O_2\!:Organization$, are not similar. This is in turn counter-attacked by the arguments H and I. The argument H states the mapping m_3 is correct, since $O_1\!:hasPublisher$ and $O_1\!:publishedBy$ are lexically similar. The argument I attacks the justification on$G$ stating that the ranges of these properties are similar, since a super-concept of $O_1\!:Publisher$, $O_1\!:Organization$, is already mapped to $O_2\!:Organization$.

The above analysis gives different, but sometimes overlapping reasons to argue for and against several candidate mappings. Assume now that there are two possible audiences, $\mathcal{R}_1$, which prefers terminology to external structure, ($T \succ_{\mathcal{R}_1} ES$), and $\mathcal{R}_2$, which prefers external structure to terminology ($ES \succ_{\mathcal{R}_2} T$). For $\mathcal{R}_1$, we get two preferred extensions for the union of the argumentation frameworks $\{A, C, D, F, I, H\}$, and $\{A, C, D, E, I, H\}$, since E and F form a two cycle between types about which no preference has been expressed. For $\mathcal{R}_2$, however, the preferred extensions are $\{A, C, D, F, I, H\}$, $\{B, D, F, I, H\}$, $\{A, C, E, I, H\}$ and $\{B, E, I, H\}$, as there is a two cycle in ES which is no longer broken by C and no preference has been expressed between ES and IS. Therefore, the arguments that are accepted by both audiences are only $\{I, H\}$. Arguments A, C, D, E, and F are, however, all potentially acceptable, since both audiences can choose to accept them, as they appear in some preferred extension for each audience. This means that the mapping m_1 will be rejected (since B is unacceptable to $\mathcal{R}_1$), while the mapping m_2 will be accepted (it is accepted by $\mathcal{R}_1$ and acceptable to $\mathcal{R}_2$). m_3 will be accepted because H is agreed acceptable for these audiences. The *agreeable alignment* is then m_2 and m_3. Interestingly, in this scenario, should an agent wish to reject the mappings m_2 and m_3, it can achieve this by considering a new audience $\mathcal{R}_3$, in which internal structure is valued more then external structure, which is valued more than terminology ($IS \succ_{\mathcal{R}_3} ES \succ_{\mathcal{R}_3} T$). In this case, the preferred extension from framework (b) is $\{E, G, I\}$, since the new preference allows G to defeat H and resist I. G will also defeat F leaving E available to defeat D. This clearly shows how the acceptability of an argument crucially depends on the audience to which it is addressed.

6 Related work

There are few approaches in the literature which have tackled the problem of agents negotiating about ontology alignments. An ontology mapping negotiation [19] has been proposed to establish a consensus between different agents which use the MAFRA alignment framework [20]. The approach is based on the utility and meta-utility functions used by the agents to establish if a mapping is accepted, rejected or negotiated. However, the approach is highly dependent on the use of the MAFRA framework and cannot be flexibly applied in other environments. [21] present an approach for agreeing on a common grounding ontology, in a decentralised way. Rather than being the goal of any one agent, the ontology mapping is a common goal for every agent in the system. [3] present an ontology negotiation protocol which enables agents to exchange parts of their ontology, by a process of successive interpretations, clarifications, and explanations. However, the end result of this process is that each agent will have the same ontology made of some sort of union of all the terms and their relations. In our context, agents keep their own ontologies, that they have been designed to reason with, while keeping track of the mappings with other agent's ontologies.

Unlike other approaches cited above, our work takes into consideration agents interests and preferences that may influencethe selection of a given correspondence.

Contrastingly, significant research exists in the area of argumentation-based negotiation [18][16] in multi-agent systems. However, it has fundamentally remained at the level of a theoretical approach, and the few existing applications are concerned with legal cases and recently, in political decision-making [2].

7 Summary and Outlook

In this paper we have outlined a framework that provides a novel way for agents, who use different ontologies, to come to agreement on an alignment. This is achieved using an argumentation process in which candidate correspondences are accepted or rejected, based on the ontological knowledge and the agent's preferences. Argumentation is based on the exchange of arguments, against or in favour of a correspondence, that interact with each other using an *attack* relation. Each argument instantiates an argumentation schema, and utilises domain knowledge, extracted from extensional and intensional ontology definitions. When the full set of arguments and counter-arguments has been produced, the agents consider which of

Table 1. Arguments for and against the correspondences m_1, m_2 and m_3

SupC = super-classes, SubC = sub-classes, Pr = properties, Lb = label, Rg = Range, Sb = sibling-classes

Id	**Argument**	A	$\mathcal{V}$
A	$\langle SupC(Press) \cap SupC(Periodical) = \emptyset, m_1, - \rangle$	B	ES
B	$\langle SubC(Press) \cap SubC(Periodical) = \not\emptyset, m_1, + \rangle$	A,C	ES
C	$\langle Lb(Press) \not\approx Lb(Periodical), m_1, - \rangle$	B	T
D	$\langle Lb(publication) \approx Lb(Publication = \emptyset), m_2, + \rangle$	E	T
E	$\langle SupC(publication) \cap SupC(Publication), m_2, - \rangle$	D,F	ES
F	$\langle Pr(publication) \cap (Publication) \neq \emptyset, m_2, + \rangle$	E	IS
G	$\langle Rg(hasPublisher) \not\approx Rg(publishedBy), m_3, - \rangle$	F,H	IS
H	$\langle Lb(hasPublisher) \approx Lb(publishedBy), m_3, + \rangle$	G	T
I	$\langle SupC(Publisher) \cap (Organization \neq \emptyset), m_4, + \rangle$	G	ES

them should be accepted. As we have seen, the acceptability of an argument depends on the ranking - represented by a particular preference ordering on the type of arguments. Our approach is able to give a formal motivation for the selection of any correspondence, and enables consideration of an agent's interests and preferences that may influencethe selection of a correspondence. An implementation of the framework is under development. Thus the effective results of an empirical evaluation are expected in the near future. Moreover, in future work we intend to investigate use of a negotiation process to enable agents to reach an agreement on a mapping when they differ in their ordering of argument types.

ACKNOWLEDGEMENTS

The research presented in this paper is partially supported under Knowledge Web European network of excellence (FP6-IST 2004-507482) and PIPS (FP6-IST 2004-507019). Special thanks to Floriana Grasso for her comments and suggestions.

REFERENCES

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

[2] K. Atkinson, T. J .M. Bench-Capon, and P. McBurney, 'Persuasive political argument', in *Proceedings of the Fifth International Workshop on Computational Models of Natural Argument (CMNA 2005)*, eds., C. Reed F. Grasso and R. Kibble, pp. 44–51, (2005).

[3] S. C. Bailin and W. Truszkowski, 'Ontology Negotiation: How Agents Can Really Get to Know Each Other', in *Proceedings of the First International Workshop on Radical Agent Concepts (WRAC 2002)*, (2002).

[4] T. Bench-Capon, 'Value based argumentation frameworks', in *Proceedings of Non Monotonic Reasoning*, pp. 444–453, (2002).

[5] T. Bench-Capon, 'Persuasion in Practical Argument Using Value-Based Argumentation Frameworks.', in *Journal of Logic and Computation*, volume 13, pp. 429–448, (2003).

[6] T. J .M. Bench-Capon, 'Agreeing to Differ: Modelling Persuasive Dialogue Between Parties Without a Consensus About Values', in *Informal Logic*, volume 22, pp. 231–245, (2002).

[7] T. Berners-Lee, J. Hendler, and O. Lassila, 'The Semantic Web', *Scientific American*, **284**(5), 34–43, (2001).

[8] S. Doutre, T.J.M. Bench-Capon, and P. E. Dunne, 'Determining Preferences through Argumentation', in *Proceedings of AI*IA'05*, pp. 98–109, (2005).

[9] P.H. Dung, 'On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games', in *Artificial Intelligence*, volume 77, pp. 321–358, (1995).

[10] P. Dunne and T. J .M. Bench-Capon, 'Two Party Immediate Response Disputes: Properties and Efficienc y', in*Artificial Intelligence*, volume 149, pp. 221–250, (2003).

[11] M. Ehrig and S. Staab, 'QOM - Quick Ontology Mapping', in *Proceedings of the International Semantic Web Conference*, (2004).

[12] J. Euzenat, 'Alignment infrastructure for ontology mediation and other applications', in *Proceedings of the First International workshop on Mediation in semantic web services*, eds., Martin Hepp, Axel Polleres, Frank van Harmelen, and Michael Genesereth, pp. 81–95, (2005).

[13] J. Euzenat and P. Valtchev, 'Similarity-based ontology alignment in OWL-Lite', in *Proceedings of European Conference on Artificial Intelligence (ECAI04)*, (2004).

[14] T. R. Gruber, 'A Translation Approach to Portable Ontology Specifications', *Knowledge Acquisition*, **5**(2), 199–220, (1993).

[15] P. McBurney and S. Parsons, 'Locutions for Argumentation in Agent Interaction Protocols', in *Proceedings of International Workshop on Agent Communication, New-York (NY US)*, pp. 209–225, (2004).

[16] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg, 'Argumentation-based negotiation', in *The Knowledge Engineering Review*, volume 18, pp. 343–375, (2003).

[17] P. Shvaiko and J. Euzenat, 'A survey of schema-based matching approaches', *Journal on data semantics*, **4**, 146–171, (2005).

[18] C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons, 'A Framework for Argumentation-Based Negotiation', in *Proceedings of the 4th International Workshop on Intelligent Agents IV, Agent Theories, Architectures, and Languages*, (1997).

[19] N. Silva, P. Maio, and J. Rocha, 'An Approach to Ontology Mapping Negotiation', in *Proceedings of the Workshop on Integrating Ontologies*, (2005).

[20] N. Silva and J. Rocha, 'MAFRA Semantic Web Ontology MApping FRAmework', in *Proceedings of the Seventh Multi-Conference on Systemics, Cybernetics and Informatics*, (2003).

[21] J. van Diggelen, R. Beun, F. Dignum, R. van Eijk, and J.-J. Meyer, 'A decentralized approach for establishing a shared communication vocabulary', in *Proceedings of the AMKN*, (2005).

[22] P.R.S. Visser, D.M. Jones, T.J.M. Bench-Capon, and M.J.R. Shave, 'Assessing Heterogeneity by Classifying Ontology Mismatches', in *Proceedings of the FOIS'98*, ed., N. Guarino, (1998).

Table 2. Argument scheme for OWL ontological alignments

Mapping	σ	**Grounds**	**Comment**
$\langle e, e', \sqsubseteq \rangle$	+	$S(e) \subseteq S(e')$	(some or all) neighbours (e.g., super-entities, sibling-entities, etc.) of e are similar in those of e'
$\langle e, e', \sqsubseteq \rangle$	-	$S(e') \subseteq S(e)$	no neighbours of e are similar in those of e'
$\langle e, e', \sqsubseteq \rangle$	-	$S(e') \subseteq S(e)$	(some or all)neighbours of e' are similar in those of e
$\langle e, e', \equiv \rangle$	+	$S(e) \cap S(e') \neq \emptyset$	Entities have similar neighbours (e.g., super-entities, sibling-entities, etc.)
$\langle e, e', \equiv \rangle$	-	$S(e) \cap S(e') = \emptyset$	Entities does not have similar neighbours
$\langle c, c', \sqsubseteq \rangle$	+	$properties(c) \subseteq properties(c')$	(some or all) properties of c are similar in those of c'
$\langle c, c', \sqsubseteq \rangle$	-	$properties(c') \not\subseteq properties(c)$	no properties of c are similar in those of c'
$\langle c, c', \sqsubseteq \rangle$ $\langle c, c', \equiv \rangle$	- +	$properties(c') \subseteq properties(c)$ $properties(c) \cap properties(c') \neq \emptyset$	(some or all) properties of c' are included in those of c the concepts c and c' have common properties
$\langle c, c', \equiv \rangle$	-	$properties(c) \cap properties(c') = \emptyset$	no properties in c and c' are similar
$\langle p, p', \equiv \rangle$ $\langle p, p', \sqsubseteq \rangle$	+	$I(p) \approx I(p')$	Properties have similar structure (e.g., range, domain or cardinality)
$\langle p, p', \equiv \rangle$ $\langle p, p', \sqsubseteq \rangle$	-	$I(p) \not\approx I(p')$	Properties do not have similar structure
$\langle i, i', \equiv \rangle$ $\langle i, i', \sqsubseteq \rangle$	+	$properties(i, i'') \approx properties(i', i'')$	Each individual i and i' referees to a third instance i" via similar properties
$\langle p, p', \equiv \rangle$ $\langle p, p', \sqsubseteq \rangle$	-	$properties(i, i'') \not\approx properties(i', i'')$	The properties that link each individual i and i' to a third instance i"are dissimilar
$\langle e, e', \sqsubseteq \rangle$	+	$E(e) \subseteq E(e')$	(some or all) instances of e are similar in those of e'
$\langle e, e', \sqsubseteq \rangle$	-	$E(e) \not\subseteq E(e')$	no instances of e are similar in those of e'
$\langle e, e', \sqsubseteq \rangle$	-	$E(e') \subseteq E(e)$	(some or all) instances of e' are similar in those of e
$\langle e, e', \equiv \rangle$	+	$E(e) \cap E(e') \neq \emptyset$	e instances are similar in those of e' and/or vice versa.
$\langle e, e', \equiv \rangle$	-	$E(e) \cap E(e') = \emptyset$	Entities e and e' does not have common instances
$\langle e, e', \equiv \rangle$ $\langle e, e', \sqsubseteq \rangle$	+	$label(e) \approx label(e')$	Entities's labels are similar (e.g., synonyms and lexical variants)
$\langle e, e', \equiv \rangle$ $\langle e, e', \sqsubseteq \rangle$	-	$label(e) \not\approx label(e')$	Entities' labels are dissimilar (e.g., homonyms)
$\langle e, e', \equiv \rangle$ $\langle e, e', \sqsubseteq \rangle$	+	$URI(e) \approx URI(e')$	Entities' URIs are similar
$\langle e, e', \equiv \rangle$ $\langle e, e', \sqsubseteq \rangle$	-	$URI(e) \not\approx URI(e')$	Entities' URIs are dissimilar

Argumentative Reasoning Patterns

Fabrizio Macagno[1] and **Doug Walton**[2]

This paper is aimed at presenting a preliminary study on argument schemes. Argumentation theory has provided several sets of forms such as deductive, inductive and presumptive patterns of reasoning. The earliest accounts of argument schemes were advanced in Arthur Hastings' Ph.D. thesis at Northwestern University (1963), and in Perelman and Obrechts-Tyteca's work on the classification of *loci* in 1969. Other scheme sets have been developed by Toulmin, Rieke, Janik (1984), Schellens (1985),van Eemeren and Kruiger (1987), Kienpointner (1992) and Grennan (1997). Each scheme set put forward by these authors presupposes a particular theory of argument. Each theory, in turn, implies a particular perspective regarding the relation between logic and pragmatic aspects of argumentation, and notions of plausibility and defeasibility. The history of argument schemes begins with the concepts of *topos* and *locus*.

1 Loci and argumentation schemes

In the field of argumentation there are conflicting views about what an argument is and what must be present for something to be regarded as an argument. Arguments may be thought of as complex speech acts or as propositional complexes (the result of speech acts, namely a speech act's propositional product). These two perspectives follow from two different approaches to argument schemes. Both perspectives, though, have in common a fundamental feature; namely, they both identify recurrent patterns or argument schemes from arguments. This common feature distinguishes the modern theories on argumentation from traditional dialectical and rhetorical studies. In the ancient tradition, the focus of the studies was limited to the *locus*. The locus of an argument is the proposition upon which the argument is based and is the proposition that is accepted by everyone (*maxima proposition*). Modern theories, in their study on argument schemes, comprehend not only what was traditionally thought of as *topoi* or *loci*, but also the use of *topoi* or *loci* in actual argumentation.

1.1 Aristotelian *Topoi*

The whole occidental tradition on dialectics stems from Aristotle's *Topics*. The first translation of the Topics by Cicero was later commented and conceptually reorganised by Boethius in *De Differentiis Topicis*. This later treatise was the primary source for most of medieval commentaries and dialectical works on what is nowadays called argumentation. In Aristotle, *topoi* have the twofold function of proof and invention, that is, they are regarded as points of view under which a conclusion can be proved true or false, and as places where arguments can be found (De Pater, 1965, p. 116). Their logical structure has been studied by (Kienpointner 1987, p. 281).

1.2 Loci in the Ancient Tradition

In the middle ages, the Aristotelian topics were completely reinterpreted and their function and role substantially changed. Two main developments in the treatment of the topics can be recognized (Stump, 1989, p. 287). First, all syllogisms were regarded as dependent upon topics and, secondly, later on, all topical arguments were considered necessary. In order to understand these two developments, it is useful to analyse Boethius' *De Differentiis Topics* and their interpretation in Abelard and in the following theories in the 12th and 13th century, until the works Burley in the 14th century. The roots of medieval dialectics can be found in Boethius' work *De differentiis topicis*. Some of the *topoi* (Boethius, 1185C, 1185D) are necessary connections, while others (for instance, from the more and the less) represent only frequent connections. Dialectical *loci* are distinct from rhetorical *loci* because, the former are relative to abstract concepts (the things, such as robbery), the latter stem from things having the qualities (the concrete cases, such as a particular case of robbery) (1215C)[3]. During the middle ages, the focal point of the study of argument was the connection between dialectics and demonstration. Beginning with the XIth century, Garlandus Compotista conceived all the topics under the logical forms of topics from antecedent and consequent, whose differentiae (the genera of *maximae propositiones*) are the syllogistic rules (Stump, 1982, p. 277). In the XIIth century, Abelard in his *Dialectica* examined for the first time[4] the structure of dialectical consequence in its components. In this work, the *maxima proposition*, expressing a necessary truth, is structurally connected to the *endoxon*. The relation between contingent and necessary truth is considered to be an assumption. Burley and Ockham organised the consequences into classes, according to the type of medium, which can be extrinsic (such as the rule of conversion) or intrinsic (for instance, the topic from genus), formal (holding by means of an extrinsic topics) or material (supported by an intrinsic topic, dependent on the meaning of the terms) (Boh, 1984, p. 310). The doctrine of loci was then taken over in the Renaissance by Rudulphus Agricola. Topics were deemed to be the means by which arguments are discovered and knowledge is obtained. In this treatise, the difference between dialectical and rhetorical *loci*, a distinction maintained throughout the whole Middle Age is blurred. While Logic is related to the abstract, i.e. formal relationships between concepts, the topics pertain to the discussion and to the matter treated in the dialogue (Agricola, 1976, p.12-13). In the Port Royal logic, in 17th Century, topics were regarded as part of the *inventio*

[1] Department of Linguistics, Catholic University of Milan
[2] Department of Philosophy, University of Winnipeg

[3] Rhetorical loci do not proceed from relations between concepts, but from stereotypes and are relative to what is implied or presupposed by a particular fact. For instance, given a murder and a person accused of homicide, the rhetorical reasoning can proceed from the place and time of the plaintiff (he was seen close to the scene of the murder, therefore he may have committed the murder). See Boethius 1215b.

[4] M. Kienpointer, 1987, p. 283.

and were classified according to criteria that differed from that of Aristotle and that were maintained throughout the Middle Age. The focus of this work is on the different kinds of argument and the division is based on the fields of human knowledge the premises of the argument belong to (Arnauld, 1964, p. 237).

1.3 *Topoi* and their development into argumentation schemes

The ancient dialectical tradition of topics is the predecessor to and the origin of the modern theories of argument schemes. In this section, the most important and relevant approaches of modern theories of argument schemes are outlined.

1.3.1 Hastings

Hastings described nine modes of reasoning, grouped into three classes: verbal and semantic procedure (argument from example, from verbal classification and from definition), causal connections (arguments from sign, from cause and from circumstantial evidence) and arguments supporting either verbal or causal conclusions (arguments from comparison, analogy and testimony). In his work, Hastings analysed the necessary conditions for the correct use of each scheme. The critical questions matching a scheme provide criteria for evaluation of the type of argument (Hastings 1963, p. 55).

1.3.2 Perelman

In Perleman and Olbrecht-Tyteca's theory, *loci* are seen as general strategies or rathercatalogs of the habits of mind endemic to a given culture[5]. About 100 argument patterns are described in their work and are classified into two main categories: arguments by association[6] and arguments by dissociation[7]. Arguments from association are divided into three main classes: Quasi-logical Arguments, Relations Establishing the Structure of Reality and Arguments based on the Structure of Reality. In arguments from dissociation, concepts conceived as a whole are separated into two new concepts, introducing polisemy.

1.3.3 Schellens

Schellens' argument schemes (Schellens 1985) are primarily drawn from Hastings' and are classified into four classes according to their pragmatic function (Kienpointner, 1992, pp. 201-215). The first group is comprised of pragmatic arguments and is normative and descriptive. The second group is comprised of unbound arguments and is either normative or descriptive. Every scheme is associated to a set of evaluation questions, similar to Hastings' critical questions.

1.3.4 Kienpointner

In Alltagslogik, Kienpointner classifies roughly 60 context-independent argument schemes in three main groups according to their relation with the rule or generalization (*endoxon*). Argument schemes may be based on rules taken for granted, establish them by means of induction, or illustrate or confirm them. Argument schemes, in turn, may have descriptive or normative variants and different logical forms (*Modus Ponens, Modus Tollens*, Disjunctive Syllogism, etc.).

1.3.5 Grennan

In Grennan's (1997, p. 163-165) typology all the structurally valid inductive inference patterns are classified according to 8 warrant types (effect to cause, cause to effect, sign, sample to population, parallel case, analogy, population to sample, authority, ends-means), combined with the types of claims the warrant connects (utterance-types expressing the minor premise and the conclusion of an argument, such as obligation). In this perspective, both the abstract form of the inference and the pragmatic role of the utterances expressing the sentences are taken into consideration

The main patterns of reasoning found in modern argumentation theories primarily stem from the Aristotelian and medieval dialectical *topoi*. Many arguments can be traced back to these patterns. The theory presented in the following section is focused on the treatment of real arguments and is aimed at individuating the possible patterns of reasoning they are based on.

2 Argumentation schemes in a pragmatic approach

The innovation that Walton's approach brings to this topic is the adoption of a more descriptive perspective. From this perspective, argument schemes are analysed in relation to fallacies. Many sophisms are patterns of inference that can be valid in certain contexts of argumentation. Hamblin (1970) first pointed out the necessary connection between fallacies and inferences. He attacked the standard treatment of fallacies for its lack of an explanatory theory regarding the inferences underlying the sophisms. In Walton's approach, most of the traditional fallacies are regarded as kinds of errors or failure in particular argumentation schemes, infractions of the necessary conditions required for the correct deployment of a *topos* in a type of dialogue.

2.1 Walton's pragmatic approach: Structure of an argument scheme

In Walton's perspective, arguments are analysed in a specific conversational context. The propositional content of the argument is considered in relation to its use in a type of dialogue and arguments are evaluated also by means of the rules of the dialogue game the interlocutors are involved in. Arguments usually considered as fallacious, for instance the *ad hominem* argument, can be acceptable if certain dialogical conditions are respected. Each argument scheme provides not only the general structure of the propositions constituting the argument, but also the necessary conditions by which its acceptability is determined. Argument schemes are presumptive and defeasible. Since each argument scheme is not only regarded to be an abstract propositional form but also a pattern instantiated in real dialogues, it cannot be said to be always valid in a discussion. It is subject to defeasibility when new information is added and either contradicts the argument's premises or conclusion, or weakens its force by making it irrelevant to support the position. For this reason, arguments can be presumptively accepted by the other party, but their relevance and role in the dialogue depend upon the fulfilment of the critical questions. Examples are argument from expert opinion (Walton 2002, pp. 49-50) and *argumentum ad hominem* (Walton 1998, pp. 199-215)

[5] Warnick, 2000, p. 111.

[6] For example, two different concepts might be associated into a unity, such as in the example: I have accused; you have condemned, is the famous reply of Domitius Afer. (Perelman, Olbrechts-Tyteca, 1969, p. 223)

[7] For example, the concept of religion is divided into *apparent religion* vs. *true religion*: What religion do I profess? None of all those that you mention. And why none? For religion's sake! (Perelman, Olbrechts-Tyteca ,1969, p. 442)

2.2 Types of argument schemes

Argumentation schemes include many patterns of reasoning in dialogue. Arguments can have deductive, inductive or abductive logical forms. They can proceed from causal connections between things, from the meaning of terms, from the relationship between the interlocutors, or from the status of the speaker. The premises can be rules, dialogical norms, or accepted opinions. A distinct classification is difficult to find, but, at the same time, is necessary in order to organize analytical tools reconstructing arguments. In the diagram below, the first scheme has a constructive aim, while the second can be used only to rebut the first. The refutation scheme stems from the third critical question of the constructive argument (Walton, 1996, p. 92).

Argument from established rule	*Argument from exceptional case*
M.p: If carrying out types of actions including the state of affairs A is the established rule for x, then (unless the case is an exception), x must carry out A. **m.p.:** Carrying out types of actions including state of affairs A is the established rule for a **Concl.:** Therefore a must carry out A. CQ_1 **:** Does the rule require carrying out types of actions that include A as an instance? CQ_2 **:** Are there other established rules that might conflict with, or override this one? CQ_3 **:** Is this case an exceptional one, that is, could there be extenuating circumstances or an excuse for noncompliance ?	**M.p.:** Generally, according to the established rule, if x has property F, then x also has property G. **m.p.:** In this legitimate case, a has F but does not have G. **Concl.:** Therefore an exception to the rule must be recognized, and the rule appropriately modified or qualified.

Along with this distinction in levels of dialogue, argument schemes can be classified according to the components of the argumentative process. In addition to patterns aimed at the subject of the discussion, schemes can also involve the emotions of the interlocutor, or the ethos of the speaker, or the common ground between the interlocutors. An example can be given of the three classes of scheme in the patterns below, respectively argument from distress (Walton 1997, p. 105), argument from popularity (Walton 1999, p. 223) and Ethotic Argument (Walton 1995, p. 152):

Hearer	**Common Ground**	**Speaker**
Argument from Distress	*Argument from Popularity*	*Ethotic Argument*
M.p.: Individual x is in distress (is suffering). **m.p.:** If y brings about A, it will relieve or help to relieve this distress. **Concl:** Therefore, y ought to bring about A.	**P.:** Everybody (in a particular reference group, G) accepts A **Concl:** Therefore, A is true (or you should accept A).	**M.P:** If x is a person of good (bad) moral character, then what x says should be accepted as more plausible (rejected as less plausible). **m.p.:** a is a person of good (bad) moral character. **Concl.:** Therefore what x says should be accepted as more plausible (rejected as less plausible).

Almost all the arguments taken into consideration in most of the theories are related to the topic of the discussion itself and they can be divided according to both their content and their logical form.

2.3 Argument schemes and missing premises: the reconstruction of real arguments

Argument schemes are an extremely useful tool for argument reconstruction. Arguments in real conversational situations almost always proceed from premises that are taken for granted. This is the case because these premises are shared by the community of speakers or presumed to be commonly accepted. When a difference occurs between those premises which are actually granted by the interlocutor and those assumptions upon which the argument is based, a fallacy often results. For instance, the speaker may take for granted a premise that the hearer does not accept, or a proposition is assumed as necessary or highly plausible while the interlocutor consider it only slightly possible. The argument scheme is fundamental for the reconstruction of the implicit premises because the missing logical step can be found by considering the structure of the inference.

3 Conclusions

The aim of the paper has been to offer a prolegomenon to the project of constructing a typology of argument schemes. Since many argument schemes found in contemporary theories stem from the ancient tradition, we took into consideration classical and medieval dialectical studies and their relation with argumentation theory. This overview on the main works on topics and schemes provides a basis for approaching main principles of classification.

REFERENCES

[1] Abelard, P.(1970). Dialectica. In L.M. de Rijk (ed.), *Petrus Abelardus: Dialectica*. Assen: Van Gorcum.

[2] Agricola (1976). *De Inventione Dialectica libri tres*. New York: George Olms Verlag.

[3] Aristotle (1939). *Topica*. Translated by E. S. Forster, Loeb Classical Library. Cambridge, Mass.: Harvard University Press.

[4] Arnauld, A. (1964). *The Art of Thinking*. Indianapolis, In.: The Bobbs-Merril Company.

[5] Boh, I. (1981). Consequences. In N. Kretzmann, et al. (eds.). *The Cambridge History of Later Medieval Philosophy*, pp. 302 -313. Cambridge: Cambridge University Press.

[6] Boh, I. (1984). Epistemic and alethic iteration in later medieval logic. *Philosophia Naturalis* 21: 492-506.

[7] Green-Pedersen, N.J. (1973). On the interpretation of Aristotle's Topics in the late 13th Century. *Cahiers de l'institut du Moyen-Âge grec et latin*, 9.

[8] Grennan, W. (1997). *Informal Logic*. London: McGill-Queen's University Press.

[9] Hamblin, C.L. (1970). *Fallacies*. London: Methuen.

[10] Hurley, P.J. (2000). *A Concise Introduction to Logic*. Belmont: Wadsworth.

[11] Kienpointner, M. (1987). Towards a typology of argumentativeschemes. In F. H. V. Eemeren, R. Grootnedorst, J. A. Blair,& C. A. Willard (eds.). *Argumentation: Across the lines ofdiscipline. Proceedings of the conference on argumentation 1986* (pp. 275-287). Providence, USA: Foris Publications.

[12] Kienpointner, M. (1992). *Alltagslogik. Struktur und Funktion von Argumentationsmustern*. Stuttgart- Bad Cannstatt: Frommann-Holzboog.

[13] Perelman, C. & Olbrechts-Tyteca, L. (1969). *The New Rhetoric: A Treatise on Argumentation.* Translated by Wilkinson, J. & Weaver, P. Notre Dame, Ind.: University of Notre Dame Press.

[14] Reed, C. & Rowe, G. (2001). *Araucaria: Software for Puzzles in Argument Diagramming and XML.* Department of Applied Computing, University of Dundee Technical Report.

[15] Rigotti, E. (2006). *Elementi di Topica*. To appear.

[16] Stump, E. (1982). Topics: Their Development and Absorption into Consequences. In N. Kretzmann, et al. (eds.). *The Cambridge History of Later Medieval Philosophy*, pp. 273-99. Cambridge: Cambridge University Press.

[17] Stump, E. (1989). *Dialectic and its place in the development of medieval logic.* Imprint Ithaca, N.Y.: Cornell University Press.

[18] Stump, E. (trans.) (1978). *Boethius' "De topicis differentiis"*, Ithaca, NY: Cornell University Press.

[19] Toulmin, S., Rieke, R. & Janik, A. (1984). *An introduction to reasoning*. New York: Macmillan Publishing co.

[20] Van Eemeren, F. H. & Kruiger, T. (1987). Identifying Argumentation Schemes. In *Argumentation: Perspectives and Approaches*, edited by Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair, and Charles A. Willard. Dordrecht and Providence: Foris Publications.

[21] Walton, D. (1995). *A Pragmatic Theory of Fallacy*. Tuscaloosa: The University of Alabama Press.

[22] Walton, D. (1997). *Appeal to Expert Opinion*. University Park: The Pennsylvania State University Press.

[23] Walton, D. (1998). *Ad Hominem Arguments*. Tuscaloosa: University of Alabama Press.

[24] Walton, D. (2002). *Legal Argumentation and Evidence*. University Park: The Pennsylvania State University Press.

[25] Warnick, B. (2000). Two Systems of Invention: The Topics in the Rhetoric and The New Rhetoric. Alan G. Gross and Arthur E. Walzer (eds.), *Rereading Aristotle's "Rhetoric,"* Carbondale: Southern Illinois UP, 107-29.

A formal framework for inter-agents dialogue to reach an agreement about a representation[1]

Maxime Morge and **Yann Secq** and **Jean-Christophe Routier** [2]

Abstract. We propose in this paper a framework for inter-agents dialogue to reach an agreement, which formalize a debate in which the divergent representations are discussed. For this purpose, we propose an argumentation-based representation framework which manages the conflicts between claims with different relevances for different audiences to compute their acceptance. Moreover, we propose a model for the reasoning of agents where they justify the claims to which they commit and take into account the claims of their interlocutors. This framework bounds a dialectics system in which agents play a dialogue to reach an agreement about a conflict of representation.

1 Introduction

A fundamental communication problem in open multiagent systems is caused by the heterogeneity of agents knowledge, in particular the discrepancy of the underlying ontologies. The approaches, such as standardization [6] and ontology alignment [4], are not suited due to the system openness. Since standardization requires that all parties involved reach a consensus on the ontology to use, it seems very unlikely that it will ever happen. On the other hand, ontology alignment is a technique that enables agents to keep their individual ontologies by making use of mappings. However, we do not know *a priori* which ontologies should be mapped within an open multiagent system. Conflicts of representation should not be avoid but resolved [1]. Contrary to [3], our work is not restricted to a protocol but also provide a model of reasoning and a model of agents.

Argumentation is a promising approach for reasonning with inconsistency information. In [14], Dung formalizes the argumentation reasonning with a framework made of abstract arguments with a contradiction relation to determine their acceptances. Classicaly, the extensions of this framework are built upon a background logic language [13, 7]. Therefore, arguments are not abstract entities but relations of consequence between a premise and a conclusion. Moreover, are introduced argumentative frameworks which assign strength to the arguments according to one (in [13]) or many priority relationships (in [12, 7]).

In this paper, we aim at using argumentative technics in order to provide a dialogical mechanism for the agents to reach an agreement on their representations. For this purpose, we extend DIAL [7], a formal framework for inter-agents dialogue based upon the argumentative technics. We propose here an argumentation-based representation framework, offering a way to compare definition with contradiction relation and to compute their acceptance. We propose a model of agent reasonning to put forward some definitions and take into account the definitions of their interlocutors. Finally, we bound here a dialectic system in which a protocol enables two agents to reach an agreement about their representations.

Paper overview. Section 2 introduces the example of dialogue that will illustrate our framework throught this paper. In section 3, we provide the syntax and the semantic of the description logic which is adopted in this paper. Section 4 presents the argumentation framework that manages interaction between conflicting representations. In accordance with this background, we describe in section 5 our agent model. In section 6, we define the formal area for agents debate. The section 7 presents the protocol used to reach an agreement.

2 Natural language

A dialogue is a coherent sequence of moves from an initial situation to reach the goal of participants [9]. For instance, the goal of dialogues consists in resolving a conflict about a representation. In the initial situation, two participants do not share the same definition of a concept, either because one participant ignore such a definition, or their own definitions are contradictory. Such cases appear quite often in dialogues and may cause serious communication problems. At the end of the dialogue, the participants must reach an agreement about the definition of this concept.

Before we start to formalize such dialogues, let us first discuss the following natural language dialogue example between a visitor and a guide in the Foire de Paris:

- visitor : Which kind of transport service can I use to go the Foire de Paris ?
- guide : The subway is a suitable transport service.
- visitor : Why the subway is a suitable transport service ?
- guide : The subway can transport you in the Hall C at the level 2.
- visitor : To my opinion, the service must transport me anywhere in Paris.
- guide : To my opinion, the service does not need to transport you anywhere in Paris but a taxi can.

In this dialogue, two participants share the concept “suitable transport service”. However, this dialogue reveals a conflict in the divergent definitions of this concept and resolve it. The guide considers that the definition of the visitor make authority and adjust her own representation to adopt this definition. Below we will assume the guide gives priority to the visitor's concepts.

[1] This work is supported by the CPER TAC of the region Nord-Pas de Calais and the european fund FEDER.

[2] Laboratoire d'Informatique Fondamentale de Lille, F-59655 VILLENEUVE D'ASCQ Cedex FRANCE, email: {morge,secq,routier}@lifl.fr

3 Ontology and Description Logic

In this section, we provide the syntax and the semantics for the well-known $\mathcal{ALC}$ [8] which is adopted in the rest of the paper.

The data model of a knowledge base (KBase, for short) can be expressed by means of the Description Logic (DL, for short) which has a precise semantic and effective inference mechanisms. Moreover, most ontologies markup langagues (e.g. OWL) are partly founded on DL. Although, it can be assumed that annotations and conceptual models are expressed using the XML-based languages mentioned above. The syntax of the representation adopted here is taken from standard constructors proposed in the DL literature. This representation language is sufficiently expressive to support most of the principal constructors of any ontology markup language.

In $\mathcal{ALC}$, primitive concepts, denoted $C, D, \dots$ are interpreted as unary predicates and primitive roles, denoted $R, S, \dots$, as binary predicates. We call description a complex concepts which can be built using constructors. The syntax of $\mathcal{ALC}$ is defined by the following BNF definition: $C \rightarrow \top|\bot|C|\neg C|C \sqcup D|C \sqcap D|\exists R.C|\forall R.C$

The semantics is defined by an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is the non-empty domain of the interpretation and $\cdot^{\mathcal{I}}$ stands for the interpretation function. The semantics of the constructors are summarized in the figure 1.

Figure 1. Semantics of the $\mathcal{ALC}$ constructors

Name	Syntax	Semantics
top concept	$\top$	$\Delta^{\mathcal{I}}$
bottom concept	$\bot$	$\emptyset$
concept	C	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
concept negation	$\neg C$	$\Delta^{\mathcal{I}} - C^{\mathcal{I}}$
concept conjunction	$C_1 \sqcap C_2$	$C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$
concept disjunction	$C_1 \sqcup C_2$	$C_1^{\mathcal{I}} \cup C_2^{\mathcal{I}}$
existencial restriction	$\exists R.C$	$\{x \in \Delta^{\mathcal{I}}; \exists y \in \Delta^{\mathcal{I}}((x,y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}})\}$
universal restriction	$\forall R.C$	$\{x \in \Delta^{\mathcal{I}}; \forall y \in \Delta^{\mathcal{I}}((x,y) \in R^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}})\}$

A KBase $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ contains a T-box $\mathcal{T}$ and a A-box $\mathcal{A}$. The T-box includes a set of concept definition ($C \equiv D$) where C is the concept name and D is a description given in terms of the language constructors. The A-box contains extensional assertions on concepts and roles. For example, a (resp. (a, b)) is an instance of the concept C (resp. the role R) iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ (resp. $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$). We call **claims**, the set of concept definitions and assertions contained in the KBase. A notion of subsumption between concepts is given in terms of the interpretations.

Definition 1 (Subsumption)**.** *Let C and D be two concepts. C* ***subsumes*** *D (denoted $C \sqsupseteq D$) iff for every interpretation $\mathcal{I}$ its holds that $C^{\mathcal{I}} \supseteq D^{\mathcal{I}}$.*

Indeed, $C \equiv D$ amounts to $C \sqsupseteq D$ and $D \sqsupseteq C$. We allow that the KBase contains partial definitions, *i.e.* axioms based on subsumption ($C \sqsupseteq D$). Below we will use $\mathcal{ALC}$ in our argumentation-based representation framework.

4 Argumentation KBase

At first, we consider that agents share a common KBase. In order to manage the interactions between conflicting claims with different revelances, we introduce an argumentation KBase.

We present in this section a value-based argumentation KBase, *i.e.* an argumentation framework built around the underlying logic language $\mathcal{ALC}$, where the revelance of claims (concept definitions and assertions) depends on the audience. The KBase is a set of sentences in a common language, denoted $\mathcal{ALC}$, associated with a classical inference, denoted $\vdash$. In order to take into account of the variability of particular situations, we are concerned by a set of audiences (denoted $\mho_A = \{\mathrm{a}_1, \dots, \mathrm{a}_n\}$), which adhere to different claims with a variable intensity.

The audiences share an argumentation KBase, *i.e.* a set of claims promoting values:

Definition 2. *Let $\mho_A = \{a_1, \dots, a_n\}$ be a set of audiences. The* ***value-based argumentation KBase*** *$AK = \langle \mathcal{K}, V, promote \rangle$ is defined by a triple where:*

- *$\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ is a KBase, i.e. a finite set of claims in $\mathcal{ALC}$;*
- *V is a non-empty finite set of values $\{v_1, \dots, v_t\}$;*
- *promote : $\mathcal{K} \rightarrow V$ maps from the claims to the values.*

We say that the claim ϕ relates to the value v if ϕ promotes v. For every $\phi \in \mathcal{K}$, promote$(\phi) \in V$.

To distinguish different audiences, values, both concrete and abstract, constitute starting points [10]. Values are arranged in hierarchies. For example, an audience will value both justice and utility but an argument may require a determination of strict preference between the two. Since audiences are individuated by their hierarchies of values, the values have different priorities for different audiences. Each audience a_i is associated with a **value-based argumentation KBase** which is a 4-tuple $\mathrm{AK}_i = \langle \mathcal{K}, V, \mathrm{promote}, \ll_i \rangle$ where:

- $\mathrm{AK} = \langle \mathcal{K}, V, \mathrm{promote} \rangle$ is a value-based argumentation KBase as previously defined;
- $\ll_i$ is the priority relation of the audience a_i, *i.e.* a strict complete ordering relation on V.

A priority relation is a transitive, irreflexive, asymmetric, and complete relation on V. It stratifies the KBase into finite non-overlapping sets. The priority level of a non-empty KBase $K \subseteq \mathcal{K}$ (written $\mathrm{level}_i(K)$) is the least important value promoted by one element in K. On one hand, a priority relation captures the value hierarchy of a particular audience. On the other hand, the KBase gathers claims (concept definitions and assertions) that are shared by audiences. Definitions, that are consequence relations between a premise and a conclusion, are built on this common KBase.

Definition 3. *Let K be a KBase in $\mathcal{ALC}$. A* ***definition*** *is couple $A = \langle \Phi, \phi \rangle$ where ϕ is a claim and $\Phi \subseteq K$ is a non-empty set of claims such as : Φ is consistent and minimal (for set inclusion); $\Phi \vdash \phi$. Φ is the premise of A, written Φ = premise(A). ϕ is the conclusion of A, denoted ϕ = conc(A).*

In other words, the premise is a set of claims from which the conclusion can be inferred. The definition A' is a **sub-definition** of A if the premise of A' is included in the premise of A. A' is a **trivial definition** if the premise of A' is a singleton. Since the KBase $\mathcal{K}$ can be inconsistent, the set of definitions (denoted $\mathcal{A}(\mathcal{K})$) will conflict.

Definition 4. *Let K be a KBase in $\mathcal{ALC}_{\mho}$ and $A = \langle \Phi, \phi \rangle, B = \langle \Psi, \psi \rangle \in \mathcal{A}(K)$ two definitions. A* ***attacks*** *B iff : $\exists \Phi_1 \subseteq \Phi, \Psi_2 \subseteq \Psi$ such as $\Phi_1 \vdash \chi$ and $\Psi_2 \vdash \neg \chi$.*

Because each audience is associated with a particular priority relation, audiences individually evaluate the revelance of definitions.

Definition 5. *Let $AK_i = \langle \mathcal{K}, V, promote, \ll_i \rangle$ be the value-based argumentation KBase of the audience a_i and $A = \langle \Phi, \phi \rangle \in$*

*$\mathcal{A}(\mathcal{K})$ a definition. According to AK_i, the **revelance of** A (written revelance$_i(A)$) is the least important value promoted by one claim in the premise.*

In other words, definitions revelance depends on the priority relation. Since audiences individually evaluate definitions revelance, an audience can ignore that a definition attacks another. According to an audience, a definition defeats another definition if they attack each other and the second definition is not more revelant than the first one:

Definition 6. *Let $AK_i = \langle \mathcal{K}, V, \mathit{promote}, \ll_i \rangle$ be the value-based argumentation KBase of the audience a_i and $A = \langle \Phi, \phi \rangle$, $B = \langle \Psi, \psi \rangle \in \mathcal{A}(\mathcal{K})$ two definitions. A **defeats** B **for the audience $\mathbf{a}_i$** (written defeats$_i(A, B)$) iff $\exists \Phi_1 \subseteq \Phi, \Psi_2 \subseteq \Psi$ such as : i) $\Phi_1 \vdash \chi$ and $\Psi_2 \vdash \neg\chi$; ii) $\neg(\mathit{level}_i(\Phi_1) \ll_i \mathit{level}_i(\Psi_2))$. Similarly, we say that a set S of definitions defeats B if B is defeated by a definition in S.*

Considering each audience own viewpoint, we define the subjective acceptance notion:

Definition 7. *Let $AK_i = \langle \mathcal{K}, V, \mathit{promote}, \ll_i \rangle$ be the value-based argumentation KBase of the audience a_i. Let $A \in \mathcal{A}(\mathcal{K})$ be a definition and $S \subseteq \mathcal{A}(\mathcal{K})$ a set of definitions. A **is subjectively acceptable by the audience $\mathbf{a}_i$ with respect to S** iff $\forall B \in \mathcal{A}(\mathcal{K})$ defeats$_i(B, A) \Rightarrow$ defeats$_i(S, B)$.*

The following example illustrates our argumentation-based representation framework.

Example 1. *Let us consider two participants coming to the "Foire de Paris" and arguing about suitable transport service. Without loosing generality, we restrict the KBase to the T-box in this example. The value-based argumentation KBase of the audience a_1 (resp. a_2) is represented in the figure 2 (resp. figure 3). The audience is associated with a KBase, i.e. a set of claims. The different claims $\phi_{11}, \ldots, \phi_7$ relate to the different values $v_1, \ldots, v_7$. According to an audience, a value above another one in a table has priority over it. The five following definitions conflict:*
$A_1 = (\{\phi_{11}, \phi_3, \phi_{22}\}, \mathit{Taxi}(x))$;
$A_2 = (\{\phi_{11}, \phi_5, \phi_6\}, \mathit{Taxi}(x))$;
$B = (\{\phi_{11}, \phi_4, \phi_7, \phi_{12}\}, \neg\mathit{Taxi}(x))$;
$B' = (\{\phi_{11}, \phi_4, \phi_7\}, \mathit{Subway}(x))$.
B' is a sub-definition of B.
If we consider the value-based argumentation KBase of the audience a_1, A_1 relevance is v_3 and B' is v_4. Therefore, B defeats A_1 but A_1 does not defeat B. If we consider thevalue-based argumentation KBase of the audience a_2, A_1 revelance is v_3 and B' is v_7. Therefore, A_1 defeats B but B does not defeat A_1. Whatever the audience is, the set $\{A_1 A_2\}$ is subjectively acceptable wrt $\mathcal{A}(\mathcal{K})$.

Figure 2. The value-based argumentation KBase of the first participant

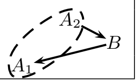
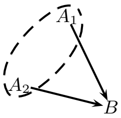
$\ll_1$	V	$\mathcal{K}$	
↑	$\mathbf{v_1}$	$\phi_{\mathbf{11}}$: Trans($\mathbf{x}$) $\phi_{\mathbf{21}}$: Trans($\mathbf{x}$) $\sqsupseteq$ Subway($\mathbf{x}$) $\sqcup$ Taxi($\mathbf{x}$)	
	$\mathbf{v_2}$	$\phi_{\mathbf{12}}$: Taxi($\mathbf{x}$) $\sqcap$ Subway($\mathbf{x}$) $\equiv \perp$ $\phi_{\mathbf{22}}$: Trans($\mathbf{x}$) $\sqsupseteq$ Dest($\mathbf{x}$, inParis)	
	v_7	ϕ_7 : Trans(x) $\sqsupseteq$ Dest(x, level2hallc)	
	v_6	ϕ_6 : Trans(x) $\sqsupseteq$ Dest(x, versailles)	
	v_5	ϕ_5 : Dest(x, versailles) $\sqsupseteq$ Taxi(x)	
	v_4	ϕ_4 : Dest(x, level2hallc) $\sqsupseteq$ Subway(x)	
	v_3	ϕ_3 : Dest(x, inParis) $\sqsupseteq$ Taxi(x)	

Figure 3. The value-based argumentation KBase of the second participant

$\ll_2$	V	$\mathcal{K}$	
↑	$\mathbf{v_1}$	$\phi_{\mathbf{11}}$: Trans($\mathbf{x}$) $\phi_{\mathbf{21}}$: Trans($\mathbf{x}$) $\sqsupseteq$ Taxi($\mathbf{x}$) $\sqcup$ Subway($\mathbf{x}$)	
	$\mathbf{v_2}$	$\phi_{\mathbf{12}}$: Taxi($\mathbf{x}$) $\sqcap$ Subway($\mathbf{x}$) $\equiv \perp$ $\phi_{\mathbf{22}}$: Trans($\mathbf{x}$) $\sqsupseteq$ Dest($\mathbf{x}$, inParis)	
	v_3	ϕ_3 : Dest(x, inParis) $\sqsupseteq$ Taxi(x)	
	v_4	ϕ_4 : Dest(x, level2hallc) $\sqsupseteq$ Subway(x)	
	v_5	ϕ_5 : Dest(x, versailles) $\sqsupseteq$ Taxi(x)	
	v_6	ϕ_6 : Trans(x) $\sqsupseteq$ Dest(x, versailles)	
	v_7	ϕ_7 : Trans(x) $\sqsupseteq$ Dest(x, level2hallc)	

We have defined here the mechanism to manage interactions between conflicting claims. In the next section, we present a model of agents which put forward claims and take into account other claims coming from their interlocutors.

5 Model of agents

In multi-agent setting it is natural to assume that all the agents do not use exactly the same ontology. Since agents representations (set of claims and priorities) can be common, complementary or contradictory, agents have to exchange hypotheses and argue. Our agents individually valuate the perceived commitments with respect to the estimated reputation of the agents from whom the information is obtained.

The agents, which have their own private representations, record their interlocutors commitments [5]. Moreover, agents individually valuate their interlocutors reputation. Therefore, an agent is in conformance with the following definition:

Definition 8. *The **agent** $a_i \in \mho_A$ is defined by a 6-tuple*
$a_i = \langle \mathcal{K}_i, V_i, \ll_i, \mathit{promote}_i, \cup_{j \neq i} CS^i_j, \prec_i \rangle$ where:

- *$\mathcal{K}_i$ is a personal KBase, i.e. a set of personal claims in $\mathcal{ALC}$;*
- *V_i is a set of personal values;*
- *promote$_i$: $\mathcal{K}_i \rightarrow V_i$ maps from the personal claims to the personal values;*
- *$\ll_i$ is the priority relation, i.e. a strict complete ordering relation on V_i;*
- *CS^i_j is a commitment store, i.e. a set of claims in $\mathcal{ALC}_{\mho}$. $CS^i_j(t)$ contains propositional commitments taken before or at time t, where agent a_j is the debtor and agent a_i the creditor;*
- *$\prec_i$ is the reputation relation, i.e. a strict complete ordering relation on $\mho_A$.*

The personal KBase are not necessarily disjoint. We call **common KBase** the set of claims explicitly shared by the agents: $\mathcal{K}_{\Omega_A} \subseteq \cap_{\mathrm{a}_i \in \mho_A} \mathcal{K}_i$. Similarly, we call **common values** the values explicitly shared by the agents: $V_{\Omega_A} \subseteq \cap_{\mathrm{a}_i \in \mho_A} V_i$. The common claims relate to the common values. For every $\phi \in \mathcal{K}_{\Omega_A}$, $\mathrm{promote}_{\Omega_A}(\phi) = v \in V_{\Omega_A}$. The personal KBase can be complementary or contradictory. We call **joint KBase** the set of claims distributed in the system: $\mathcal{K}_{\mho_A} = \cup_{\mathrm{a}_i \in \mho_A} \mathcal{K}_i$. The agent own claims relate to the agent own values. For every $\phi \in \mathcal{K}_i - \mathcal{K}_{\Omega_A}$, $\mathrm{promote}_i(\phi) = v \in V_i - V_{\Omega_A}$.

We can distinguish two ways for an agent to valuate her interlocutors commitments: either in accordance with a global social order [11], or in accordance with a local perception of the interlocutor, called reputation. Obviously, this way is more flexible. Reputation is a social concept that links an agent to her interlocutors. It is also a leveled relation [2]. The individuated reputation relations, which are transitive, irreflexive, asymmetric, and complete relations on $\mho_A$, preserve these properties. $\mathrm{a}_j \prec_i \mathrm{a}_k$ denotes that an agent a_i trusts an agent a_k more than another agent a_j.In order to take into account the claims notified in the commitment stores, each agent is associated with the following extended KBase:

Definition 9. *The **extended KBase of the agent $\mathbf{a}_i$** is the value-based argumentation KBase $AK_i^* = \langle \mathcal{K}_i^*, V_i^*, promote_i^*, \ll_i^* \rangle$ where:*

- *$\mathcal{K}_i^* = \mathcal{K}_i \cup [\bigcup_{j \neq i} CS_j^i]$ is the agent extended personal KBase composed of its personal KBase and the set of perceived commitments;*
- *$V_i^* = V_i \cup [\bigcup_{j \neq i} \{v_j^i\}]$ is the agent extended set of personal values composed of the set of personal values and the reputation values associated with her interlocutors;*
- *$promote_i^* : \mathcal{K}_i^* \rightarrow V_i^*$ is the extension of the function $promote_i$ which maps claims in the extended personal KBase to the extended set of personal values. On the one hand, personal claims relate to personal values. On the other hand, claims in the commitment store CS_j^i relate to the reputation value v_j^i;*
- *$\ll_i^*$ is the agent extended priority relation, i.e. an ordered relation on V_i^*.*

Since the debate is a collaborative social process, agents share common claims of prime importance. To reach the global goal of the multi-agent system, the common values have priority over the other values.

Let us consider a debate between two agents, a visitor and a guide in the "Foire de Paris". The guide considers that visitor's claims make authority and adjust her own representation to adopt these claims. By opposite, we will assume the visitor gives priority to the guide's claims. Therefore, there is an authority relation between the visitor and the guide. On one hand, a guide should consider that visitor's claims are more revelant than her own. Therefore, her interlocutor reputation values have priority over her personal values. If a_j is a visitor, the guide extended priority relation a_i is constrained as follows : $\forall v_\omega \in V_{\Omega_A} \forall v \in V_i - V_{\Omega_A}$ $(v \ll_i^* v_j^i \ll_i^* v_\omega)$. On the other hand, a visitor should consider that her own claims are more revelant than the guide ones. If a_j is a guide, the visitor extended priority relation a_i is constrained as follows: $\forall v_\omega \in V_{\Omega_A} \forall v \in V_i - V_{\Omega_A}$ $(v_j^i \ll_i^* v \ll_i^* v_\omega)$.

We can easily demonstrate that the extended priority relation is a strict complete ordering relation. The one-agent notion of conviction is then defined as follows:

Definition 10. *Let $a_i \in \mho_A$ be an agent associated with the extended KBase $AK_i^* = \langle \mathcal{K}_i^*, V_i^*, promote_i^*, \ll_i^* \rangle$ and $\phi \in \mathcal{ALC}$ a claim. The **agent $\mathbf{a}_i$ is convinced by the claim ϕ** iff ϕ is the conclusion of an acceptable definition for the audience a_i with respect to $\mathcal{A}(\mathcal{K}_i^*)$.*

Agents utter messages to exchange their representations. The syntax of messages is in conformance with the common **communication language**, $\mathcal{CL}$. A message $M_k = \langle S_k, H_k, A_k \rangle \in \mathcal{CL}$ has an identifier M_k. It is uttered by a speaker ($S_k = \mathrm{speaker}(M_k)$) and addressed to an hearer ($H_k = \mathrm{hearer}(M_k)$). $A_k = \mathrm{act}(M_k)$ is the message speech act. It is composed of a locution and a content. The locution is one of the following: `question, propose, unknow, concede, counter-propose, challenge, withdraw`. The content, also called **hypothesis**, is a claim or a set of claims in $\mathcal{ALC}$.

Speech acts have an argumentative semantic, because commitments enrich the extended KBase of the creditors, and a public semantic, because commitments are justified by the extended KBase of the debtor.

For example, an agent can propose a hypothesis if he has a definition for it. The corresponding commitments stores are updated. More formaly, an agent a_i can propose to the agent a_j a hypothesis h at time t if a_i has a definition for it. The corresponding commitments stores are updated: for any agent a_k ($\neq \mathrm{a}_i$) $\mathrm{CS}_i^k(t) = \mathrm{CS}_i^k(t-1) \cup \{h\}$.

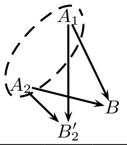
The argumentative and social semantic of the speech act `counter-propose` is equivalent with the proposition one. The rational condition for the proposition and the rational condition for the concession of the same hypothesis by the same agent distinguish themselves. Agents can propose hypotheses whether they are supported by a trivial definition or not. By contrast, an agent does not concede all the hypotheses he hears in spite of they are all supported by a trivial definition which are in the commitment stores.

The others speech acts (question(h), challenge(h), unknow(h), and withdraw(h)) are used to manage the sequence of moves (cf section 7). They have no particular effects on commitments stores, neither particular rational conditions of utterance. Since withdraw(h) speech act has no effect on commitments stores, we consider that commitments stores are cumulative [9].

The hypotheses which are received must be valuated. For this purpose, commitments will be individually considered in accordance with the speaker estimated reputation. The following example illustrates this principle.

Example 2. *If the agent a_1 utters the following message: $M_1 = \langle a_1, a_2, propose(Subway(x)) \rangle$, then the extended KBase of the agent a_2 is as represented in the table 4.*

Figure 4. The extended KBase of the agent a_2

$\ll_2^*$	V_2^*	$\mathcal{K}_2^*$	
	$\mathbf{v_1}$	$\phi_{\mathbf{11}}$: Trans$(\mathbf{x})$ $\phi_{\mathbf{21}}$: Trans $\sqsupseteq$ Taxi$(\mathbf{x}) \sqcup$ Subway$(\mathbf{x})$	
	$\mathbf{v_2}$	$\phi_{\mathbf{12}}$: Taxi $\sqcap$ Subway $\equiv \bot$ $\phi_{\mathbf{22}}$: Trans$(\mathbf{x}) \sqsupseteq$ Dest$(\mathbf{x}$, inParis$)$	
	v_3	ϕ_3 : Dest$(x$, inParis$) \sqsupseteq$ Taxi(x)	
	v_4	ϕ_4 : Dest$(x$, level2hallc$) \sqsupseteq$ Subway(x)	
	v_5	ϕ_5 : Dest$(x$, versailles$) \sqsupseteq$ Taxi(x)	
	v_6	ϕ_6 : Trans$(x) \sqsupseteq$ Dest$(x$, versailles$)$	
	v_7	ϕ_7 : Trans$(x) \sqsupseteq$ Dest$(x$, level2hallc$)$	
	v_1^2	$\{\mathrm{Subway(x)}\} = \mathrm{CS}_1^2$	

We have presented here a model of agents who exchange hypotheses and argue. In the next section, we bound a formal area where debates take place.

6 Dialectic system

When a set of social and autonomous agents argue, they reply to each other in order to reach the interaction goal, *i.e.* an agreement about a claim. We bound a formal area, called dialectic system, which is inspired by [7] and adapted to this paper context.

During exchanges, speech acts are not isolated but they respond to each other. Moves syntax is in conformance with the common **moves language** : $\mathcal{ML}$. A move $\mathrm{move}_k = \langle M_k, R_k, P_k \rangle \in \mathcal{ML}$ has an identifier move_k. It contains a message M_k as defined before. Moves are messages with some attributes to control the sequence. $R_k = \mathrm{reply}(\mathrm{move}_k)$ is the move identifier to which move_k responds.

A move (move_k) is either an initial move ($\text{reply}(\text{move}_k) = \text{nil}$) or a replying move ($\text{reply}(\text{move}_k) \neq \text{nil}$). $\text{P}_k = \text{protocol}(\text{move}_k)$ is the protocol name which is used.

A dialectic system is composed of two agents. In this formal area, two agents play moves to check an initial hypothesis, *i.e.* the topic.

Definition 11. *Let $AK_{\Omega_A} = \langle \mathcal{K}_{\Omega_A}, V_{\Omega_A}, promote_{\Omega_A} \rangle$ be a common value-based argumentation KBase and ϕ_0 a claim in $\mathcal{ALC}$. The* ***dialectics system*** *on the topic ϕ_0 is a quintuple $DS_{\Omega_M}(\phi_0, AK_{\Omega_A}) = \langle N, H, T, protocol, Z, \rangle$ where :*

- *$N = \{init, part\} \subset \mho_A$ is a set of two agents called players: the initiator and the partner;*
- *$\Omega_M \subseteq \mathcal{ML}$ is a set of well-formed moves;*
- *H is the set of histories, i.e. the sequences of well-formed moves s.t. the speaker of a move is determined at each stage by a turn-taking function and the moves agree with a protocol;*
- *$T : H \rightarrow N$ is the turn-taking function determining the speaker of a move. If $|h| = 2n$ then $T(h) = init$ else $T(h) = part$;*
- *protocol $: H \rightarrow \Omega_M$ is the function determining the moves which are allowed or not to expand an history;*
- *Z is the set of dialogue, i.e. terminal histories.*

In order to be well-formed, the initial move is a question about the topic from the initiator to the partner and a replying move from a player always references an earlier move uttered by the other player. In this way, backtracks are allowed. We call dialogue line the sub-sequence of moves where all backtracks are ignored. In order to avoid loops, hypothesis redundancy is forbidden within propositions belonging to the same dialogue line. Obviously, all moves should contain the same parameter protocol value.

We have bound here the area in which dialogues take place. We formalize in the next section a particular protocol to reach a representation agreement.

7 Protocol

When two agents have a dialogue, they collaborate to confront their representations. For this purpose, we propose in this section a protocol.

To be efficient, the protocol is a unique-response one where players can reply just once to the other player's moves. The protocol is a set of sequence rules (cf figure 5). Each rule specifies authorized replying moves. In this figure, speech acts resist or surrender to the previous one. For example, the "Propose/Counter-Propose" rule (written $\text{sr}_{P/C}$) specifies authorized moves replying to the previous propositions ($\text{propose}(\Phi)$). Contrary to resisting acts, surrendering acts close the debate. A concession ($\text{concede}(\Phi)$) surrenders to the previous proposition. A challenge ($\text{challenge}(\phi)$) and a counter-proposition ($\text{counter-propose}(\phi)$) resist to the previous proposition.

The figure 6 shows a debate in the extensive form game representation where nodes are game situations and edges are moves. For example, 2.3^{init} denotes a game situation where the exponent indicates that the initiator is the next move speaker. The exponent of game-over situations are boxes (*e.g.* $2.1^{\Box}$, $3.2^{\Box}$, and $4.2^{\Box}$). For evident clarity reasons, the games that follows situations 2.2^{init}, 4.4^{init}, and 6.3^{init} are not represented. In order to confront her representation with a partner, an initiator begins a dialogue. If the partner has no representation of the topic, he pleads ignorance and closes the dialogue (cf game situation $2.1^{\Box}$). If players have the same representation, the dialogue closes (cf game situation $3.2^{\Box}$). Otherwise, the goal of the dialogue is to reach an agreement by verbal means. The following example illustrates such a dialogue.

Example 3. *Let us consider a dialogue between a visitor and a guide in the "Foire de Paris". In the initial situation, the value-based argumentation KBase of the visitor (resp. the guide) is represented in the figure 7 (resp. figure 8). Commitments stores are the results of moves sequence (cf figure 9).*

Figure 7. Extended argumentation KBase of the visitor

$\ll_1^*$	V_1^*	$\mathcal{K}_1^*$	
↑	$\mathbf{v_1}$	$\phi_{\mathbf{11}}$: Trans$(\mathbf{x})$ $\phi_{\mathbf{21}}$: Trans $\sqsupseteq$ Taxi$(\mathbf{x}) \sqcup$ Subway$(\mathbf{x})$	
	$\mathbf{v_2}$	$\phi_{\mathbf{12}}$: Taxi $\sqcap$ Subway $\equiv \bot$ $\phi_{\mathbf{22}}$: Trans$(\mathbf{x}) \sqsupseteq$ Dest$(\mathbf{x}$, inParis$)$	
	v_3	ϕ_3 : Dest$(x$, inParis$) \sqsupseteq$ Taxi(x)	$\mathcal{A}_1$
	v_2^1	$\emptyset = \text{CS}_2^1$	

Figure 8. Extended argumentation KBase of the guide

$\ll_2^*$	V_2^*	$\mathcal{K}_2^*$	
↑	$\mathbf{v_1}$	$\phi_{\mathbf{11}}$: Trans$(\mathbf{x})$ $\phi_{\mathbf{21}}$: Trans $\sqsupseteq$ Taxi$(\mathbf{x}) \sqcup$ Subway$(\mathbf{x})$	
	$\mathbf{v_2}$	$\phi_{\mathbf{12}}$: Taxi $\sqcap$ Subway $\equiv \bot$ $\phi_{\mathbf{22}}$: Trans$(\mathbf{x}) \sqsupseteq$ Dest$(\mathbf{x}$, inParis$)$	
	v_1^2	$\emptyset = \text{CS}_1^2$	
	v_4	ϕ_4 : Dest$(x$, level2hallc$) \sqsupseteq$ Subway(x)	
	v_7	ϕ_7 : Trans$(x) \sqsupseteq$ Dest$(x$, level2hallc$)$	$\mathcal{B}$

Figure 9. Dialogue to reach an agreement

<table>
<tr><th colspan="2" rowspan="2">$\mathcal{K}_1^* - \mathcal{K}_{\Omega_A}$</th><th>$\mathcal{K}_{\Omega_A}$</th><th colspan="2" rowspan="2">$\mathcal{K}_2^* - \mathcal{K}_{\Omega_A}$</th></tr>
<tr><th>$\phi_{11}, \phi_{21}, \phi_{12}, \phi_{22}$</th></tr>
<tr><td>$\mathcal{K}_1$</td><td>CS_2^1</td><td>Game situation</td><td>CS_1^2</td><td>$\mathcal{K}_2$</td></tr>
<tr><td>ϕ_3</td><td>$\emptyset$</td><td>0</td><td>$\emptyset$</td><td>ϕ_4, ϕ_7</td></tr>
<tr><td colspan="5">$\rightarrow$ question(Trans(x)) $\rightarrow$</td></tr>
<tr><td>ϕ_3</td><td>$\emptyset$</td><td>1</td><td>$\emptyset$</td><td>ϕ_4, ϕ_7</td></tr>
<tr><td colspan="5">$\leftarrow$ propose(Subway(x)) $\leftarrow$</td></tr>
<tr><td>ϕ_3</td><td>Subway(x)</td><td>2</td><td>$\emptyset$</td><td>ϕ_4, ϕ_7</td></tr>
<tr><td colspan="5">$\rightarrow$ challenge(Subway(x)) $\rightarrow$</td></tr>
<tr><td>ϕ_3</td><td>Subway(x)</td><td>3</td><td>$\emptyset$</td><td>ϕ_4, ϕ_7</td></tr>
<tr><td colspan="5">$\leftarrow$ propose$(\phi_4, \phi_7, \phi_{11})$ $\leftarrow$</td></tr>
<tr><td>ϕ_3</td><td>Subway$(x), \phi_4, \phi_7$</td><td>4</td><td>$\emptyset$</td><td>ϕ_4, ϕ_7</td></tr>
<tr><td colspan="5">$\rightarrow$ counter-propose$(\phi_{11}, \phi_3, \phi_{22})$ $\rightarrow$</td></tr>
<tr><td>ϕ_3</td><td>Subway$(x), \phi_4, \phi_7$</td><td>5</td><td>ϕ_3</td><td>ϕ_4, ϕ_7</td></tr>
<tr><td colspan="5">$\leftarrow$ concede(Taxi(x)) $\leftarrow$</td></tr>
</table>

8 Conclusion

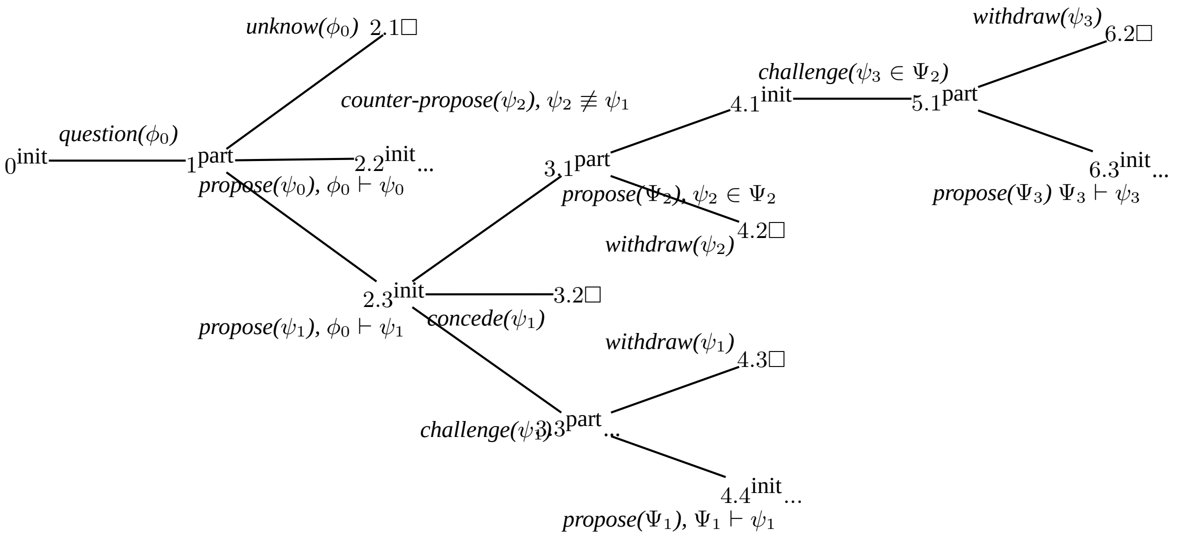
We have proposed in this paper a framework for inter-agents dialogue to reach an agreement, which formalize a debate in which the divergent representations are discussed. For this purpose, we have proposed an argumentation-based representation framework which manages the conflicts between claims with different relevances for different audiences to compute their acceptance. Moreover, we have proposed a model for the reasoning of agents where they justify the claims to which they commit and take into account the claims of their interlocutors. This framework bounds a dialectics system in which agents play a dialogue to reach an agreement about a conflict of representation.

Future works will investigate the applications of such dialogue for the services composition. For this purpose, we have to shift from our notion of propositional commitment to the notion of commitment in actions.

Figure 5. Set of speech acts and their potential answers.

Sequences rules	Speech acts	Resisting replies	Surrendering replies
sr$_{Q/A}$	question(ϕ)	propose(ϕ'), $\phi \vdash \phi'$	unknow(ϕ)
sr$_{P/C}$	propose(Φ)	challenge(ϕ), $\phi \in \Phi$ counter-propose(ϕ), $\phi \not\in \Phi$	concede(ϕ), $\Phi \vdash \phi$
sr$_{C/P}$	challenge(ϕ)	propose(Φ), $\Phi \vdash \phi$	withdraw(ϕ)
sr$_{Rec/P}$	counter-propose(Φ)	propose(Φ'), $\Phi \subseteq \Phi'$	withdraw(Φ)
sr$_T$	unknow(Φ)	$\emptyset$	$\emptyset$
	concede(Φ)	$\emptyset$	$\emptyset$
	withdraw(Φ)	$\emptyset$	$\emptyset$

Figure 6. Debate in an extensive form game representation



REFERENCES

[1] S. Bailin and W. Truszkowski, 'Ontology negotiation between intelligent information agents', *Knowledge Engineering Review*, **17**(1), (March 2002).

[2] C. Castelfranchi and R. Falcone, 'Principles of trust in mas: Cognitive anatomy, social importance, and quantification', in *Proceedings of ICMAS'98*, pp. 72–79, (1998).

[3] Van digglen Jurriaan, Beun Robbert-Jan, Dignum Frank, Van Eijk Rogier, and Meyer John-Jules, 'Anemone: An effective minimal ontology negotiation environment', in *Proc. of AAMAS*, (2006). To appear.

[4] Jerome Euzenat et al., 'State of the art on ontology alignment', Technical report, IST Knowledge web NoE, (2004). deliverable 2.2.3.

[5] Nicolleta Fornara and Marco Colombetti, 'Operational specification of a commitment-based agent communication language', in *Proc. of the first international joint conf. on autonomous agent and multiagent systems*, eds., Cristano Castelfranchi and W. Lewis Johnson, volume part 2, pp. 535–542. ACM press, (2002).

[6] Thomas R. Gruber, 'Toward principles for the design of ontologies used for knowledge sharing', *International Journal of Human-Computer Studies, special issue on Formal Ontology in Conceptual Analysis and Knowledge Representation*, **43**(5-6), 907–928, (1995).

[7] Maxime Morge, 'Collective decision making process to compose divergent interests and perspectives', *Artificial Intelligence and Law*, (2006). to appear.

[8] Manfred Schmidt-Schauß and Gert Smolka, 'Attributive concept descriptions with complements,', *Artificial Intelligence*, **48**(1), 1–26, (1991).

[9] D. Walton and E. Krabbe, *Commitment in Dialogue*, SUNY Press, 1995.

[10] Perelman C. and Olbrechts-Tyteca L. *Traité de l'Argumentation - La Nouvelle Rhétorique*. Presses Universitaires de France, 1958.

[11] Leila Amgoud, Nicolas Maudet, and Simon Parsons. An argumentation-based semantics for agent communication languages. In *Proc. of the 15th European Conference on Artificial Intelligence*, pages 38–42. IOS Press, Amsterdan, 2002.

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

[13] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Maths and AI*, 34(1-3):197–215, 2002.

[14] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–357, 1995.

A Utility and Information Based Heuristic for Argumentation[1]

Nir Oren and **Timothy J. Norman** and **Alun Preece**[2]

Abstract. While researchers have looked at many aspects of argumentation, an area often neglected is that of argumentation strategies. That is, given multiple possible arguments that an agent can put forth, which should be selected in what circumstances. In this paper, we propose a heuristic that implements one such strategy. The heuristic assigns a utility cost to revealing information, as well as a utility to winning, drawing and losing an argument. An agent participating in a dialogue then attempts to maximise its utility. After informally presenting the heuristic, we discuss some of its novel features, after which some avenues for future work are examined.

1 Introduction

Argumentation has emerged as a powerful reasoning mechanism in many domains. One common dialogue goal is to persuade, where one or more participants attempt to convince the others of their point of view. This type of dialogue can be found in many areas including distributed planning and conflict resolution, education and in models of legal argument. At the same time that the breadth of applications of argumentation has expanded, so has the sophistication of formal models designed to capture the characteristics of the domain. While many researchers have focused on the question of “what are the properties of an argument”, fewer have looked at “how does one argue well”.

In this paper, we propose a decision heuristic for an agent allowing it to decide which argument to advance. The basis for our idea is simple; the agent treats some parts of its knowledge as more valuable than other parts, and, while attempting to win the argument, attempts to minimise the amount of valuable information it reveals. This heuristic often emerges in negotiation dialogues, as well as persuasion dialogues in hostile setting (such as takeover talks or in some legal cases). Utilising this heuristic in arguments between computer agents can also be useful; revealing confidential information in an ongoing dialogue may damage an agent’s chances of winning a future argument.

In the remainder of this paper, we will briefly describe the framework, provide an example as to its functioning, and then examine its features in more detail and look at possible extensions to our approach. First however, we will examine a number of existing approaches to strategy selection.

2 Background and related research

Argumentation researchers have recognised the need for argument selection strategies for a long time. However, the field has only recently started receiving more attention. Moore, in his work with the DC dialectical system [8], suggested that an agent’s argumentation strategy should take three things into account:

- Maintaining the focus of the dispute.
- Building its point of view or attacking the opponent’s one.
- Selecting an argument that fulfils the previous two objectives.

In most cases, there is no need for a strategy to maintain the focus of a dispute; many argumentation protocols are designed so as to fore this focus to occur. Nevertheless, this item should be taken into consideration when designing a general purpose strategy. The first two items correspond to the military concept of a strategy, i.e. a high level direction and goals for the argumentation process. The third item corresponds to an agent’s tactics. Tactics allow an agent to select a concrete action that fulfils its higher level goals. While Moore’s work focused on natural language argument, these requirements formed the basis of most other research into agent argumentation strategies.

In 2002, Amgoud and Maudet [1] proposed a computational system which would capture some of the heuristics for argumentation suggested by Moore. Given a preference ordering over arguments, the created agents which could follow a “build” or “destroy” strategy, either defending their own arguments or attacking an opponent’s.

Using some ideas from Amgoud’s work, Kakas et al. [7] proposed a three layer system for agent strategies in argumentation. The first layer contains “default” rules, of the form $utterance \leftarrow condition$, while the two higher layers provide preference orderings over the rules (effectively acting as meta-rules to guide dialogue). Assuming certain restrictions on the rules, they show that only one utterance will be selected using their system, a trait they refer to as determinism. While their approach is able to represent strategies proposed by a number of other techniques, it does require hand crafting of the rules. No suggestions are made regarding what a “good” set of rules would be.

In [2], Amgoud and Prade examined negotiation dialogues in a possibilistic logic setting. An agent has a set of goals it attempts to pursue, a knowledge base representing its knowledge about the environment, and another knowledge base which is used to keep track of what it believes the other agent’s goals are. The authors then present a framework in which these agents interact which incorporates heuristics for suggesting the form and contents of an utterance, a dialogue game allowing agents to undertake argumentation, and a decision procedure to determine the status of the dialogue. They then suggest and formalise a number of strategies that an agent can follow.

Other notable mentions and formalisations of argumentation strategies can be found in [4, 10, 3]. In the latter, Bench-Capon identifies a number of stages in the dialogue in which an agent might be

[1] A more detailed version of this paper was presented at ECAI-2006

[2] Department of Computing Science, University of Aberdeen, AB24 3UE, Scotland, email: noren,tnorman,apreece@csd.abdn.ac.uk

faced with a choice, and provides some heuristics as to what argument should be advanced in each of these cases.

Apart from guiding strategy, heuristics have seen other uses in dialogue games. Recent work by Chesñevar et al. [5] has seen heuristics being used to minimise the search space when analysing argument trees. Argument schemes [13] are well used tools in argumentation research, and can be viewed as a form of heuristic that guides the reasoning procedure.

3 Confidentiality Based Argumentation

In many realms of argument, auxiliary considerations (apart from simply winning or losing the argument) come into play. In many scenarios, one such consideration involves hiding certain information from an opponent. In this section, we describe a utility based heuristic to guide an agent taking part in a dialogue while being careful about what information it reveals. When faced with a number of possible arguments that it can advance, we claim it should put forth the one that minimises the exposure of information that it would like to keep private. The limitations of our current approach, as well as extensions and refinements to it are discussed in Section 5.

This work emerged while investigating the properties of other formal argument systems (such as [6, 12, 11, 15]). It is thus based on our own formal argumentation system. We believe, and plan to show in future work, how our heuristic can be implemented in other, more widely accepted argumentation frameworks.

Our system can be divided into two parts; at the lower level lies the logical machinery used to reason about arguments, while at the higher level we have a dialogue game, definitions of agents and the environment, and the heuristic itself. In this section, we will informally discuss our framework. A formal definition of the system can be found in [9].

3.1 The Argumentation Framework

The framework underpinning our heuristic is very simple, but still allows for argumentation to takes place. Argumentation takes place over a language containing propositional literals and their negation. Arguments consist of conjunctions of premises leading to a single propositional conclusion. A conclusion a which requires no premises can be represented by the argument $(\{\top\}, a)$.

We are interested in the status of literals (given a set of arguments), rather than the status of the arguments themselves. We can classify a literal into one of three sets: *proven*, *in conflict*, and *unknown*. A literal is in conflict if we can derive both it and its negation from a set of arguments. It is (un)proven, if it can (not) be derived and it is not in conflict, and unknown if neither it, nor its negation can be derived.

Our derivation procedure is based on the forward chaining of arguments. We begin by looking at what can be derived requiring no premises. By using these literals as premises, we compute what new literals can be generated, and continue with this procedure until no further literals can be computed. At each step of the process, we check for conflicts in the derived literals. When a conflict occurs, the literal (and its negation) are removed from the derived set and placed into a conflict set. Arguments depending on these literals are also removed from the derivation procedure. At the end of the derivation procedure, we can thus compute all three classes of literals[3].

3.2 Agents, the Dialogue Game and the Heuristic

Agents engage in a dialogue using the argumentation framework described above in an attempt to persuade each other of certain facts. In our system, an agent is an entity containing a private knowledge base of arguments, a function allowing it to compute the cost of revealing literals, and a set of utilities specifying how much it would gain for winning, drawing or losing the argument. The dialogue takes place within an environment, that, apart from containing agents, contains a public knowledge base which holds all arguments uttered by the agents.

Our dialogue game proceeds by having agents take turns to make utterances[4]. An utterance consists of a set of logically linked individual arguments. Alternatively, an agent may pass, and the game ends when no new arguments have been introduced into the public knowledge by any of the participants during their turn (which means that a dialogue is guaranteed to end given assuming a finite number of arguments). Once this occurs, it is possible to determine the status of each agent's goal, allowing one to determine the net utility gain (or loss) of all the agents in the system.

An agent wins an argument if its goal literal is in the proven set, while it draws an argument if the goal literal is in the conflict set or unknown. Otherwise, an agent is deemed to lose the argument. The net utility for an agent is determined by subtracting the utility cost for all literals appearing in the conflict and knowledge set from the utility gained for winning/drawing/losing the game.

To determine what argument it should advance, an agent computes what the public knowledge base would look like after each of its possible utterances. Using the derivation procedure described previously, it determines whether making the utterance will allow it to win/draw/lose the dialogue, and, by combining this information with the utility cost for exposed literals, it computes the utility gain for every possible utterance. It then selects the utterance which will maximise its utility. If multiple such utterances exist, another strategy (such as the one described in [10]) can be used.

It should be noted that it is possible to remove literals from the conflict set by attacking the premises of the arguments that inserted them into the set (thus reinstating other arguments). The lack of a preference relation over arguments means that attack in our framework is symmetric. While limiting, we are still able to model a useful subclass of arguments.

Before discussing the properties of the system, we show how a dialogue might look when this heuristic is used.

4 Example

The argument consists of a hypothetical dialogue between a government and some other agent regarding the case for, or against, weapons of mass destruction (WMDs) existing at some location.

Assume that $Agent_0$ would like to show the existence of WMDs. Proving this gains it 100 utility, while showing that WMDs don't exist means no utility is gained. Uncertainty (i.e. a draw) yields a utility gain of 50. Furthermore, assume the agent begins with the following arguments in its knowledge base:

$$(\{\top\}, spysat), (\{\top\}, chemicals), (\{\top\}, news), (\{\top\}, factories)$$

$$(\{\top\}, smuggling), (\{smuggling\}, \neg medicine), (\{news\}, WMD)$$

$$(\{factories, chemicals\}, WMD), (\{spysat\}, WMD)$$

[3] A Prolog implementation of this framework is available at `http://www.csd.abdn.ac.uk/~noren`.

[4] Note that we place no restrictions on the number of arguing agents.

$(\{sanctions, smuggling, factories, chemicals\}, \neg medicine)$

We associate the following costs with literals:

$$\begin{array}{ll}
(spysat, 100) & (chemicals, 30) \\
(news, 0) & (\{medicine, chemicals\}, 50) \\
(smuggling, 30) & (factories, 0)
\end{array}$$

Note that if both medicine and chemicals are present, the agent's utility cost is 50, not 80. Thus, if both $spysat$ and $chemicals$ are admitted to, the agent's utility cost will be 130.

The dialogue might thus proceed as follows:

$$\begin{array}{lll}
(\mathbf{1}) & Agent_0: & (\{\top\}, news), (\{news\}, WMD) \\
(\mathbf{2}) & Agent_1: & (\{\top\}, \neg news) \\
(\mathbf{3}) & Agent_0: & (\{\top\}, factories), (\{\top\}, chemicals), \\
 & & (\{factories, chemicals\}, WMD) \\
(\mathbf{4}) & Agent_1: & (\{\top\}, sanctions), \\
 & & (\{sanctions, factories, chemicals\}, \\
 & & \quad medicine), (\{medicine\}, \neg WMD) \\
(\mathbf{5}) & Agent_0: & (\{\top\}, smuggling), \\
 & & (\{sanctions, smuggling, factories, \\
 & & \quad chemicals\}, \neg medicine) \\
(\mathbf{6}) & Agent_1: & \{\} \\
(\mathbf{7}) & Agent_0: & \{\}
\end{array}$$

Informally, the dialogue proceeds as follows: $Agent_0$ claims that WMDs exist since the news says they do. $Agent_1$ retorts that he has not seen those news reports. $Agent_0$ then points out that factories and chemicals exist, and that these were used to produce WMDs. In response, $Agent_1$ says that due to sanctions, these were actually used to produce medicine. $Agent_0$ attacks this argument by pointing out that smuggling exists, which means that the factories were not used to produce medicines, reinstating the WMD argument. Both agents have nothing more to say, and thus pass. $Agent_0$ thus wins the game.

It should be noted that while $Agent_0$ is aware that spy satellites have photographed the WMDs, it does not want to advance this argument due to the cost of revealing this information. The final utility gained by $Agent_0$ for winning the argument is 20: 100 for winning the argument, less 30 for revealing $smuggling$, and 50 for the presence of the $chemicals$ and $medicine$ literals. Also, note that the fact that $Agent_1$ revealed the existence of medicines cost $Agent_0$ an additional 20 utility. While this makes sense in some scenarios, it can be regarded as counterintuitive in others. Extensions to overcome this behaviour are examined in the next section.

5 Discussion

As mentioned earlier, we created our own underlying framework, and one of our short term research goals involves mapping our heuristic into another, more widely used argumentation framework. Our framework shares much in common with the "sceptical" approach to argumentation; when arguments conflict, we refuse to decide between them, instead ruling them both invalid. The simplicity of our approach means that only specific types of arguments can be represented (namely, those whose premises are a conjunction of literals, and whose conclusion is a single literal). However, as seen in the example, even with this limitation, useful arguments can still emerge.

The way in which we represent the information "leaked" during the dialogue, as well as calculate the agent's net utility, while simple, allows us to start studying dialogues in which agents attempt to hide information. Until now, most work involving utility and argumentation has focused on negotiation dialogues (e.g. [14]). We propose a number of possible extensions to the work presented in this paper.

One simple extension involves the addition of a context to the agent's cost. In other words, given that fact A, B and C are known, we would like to be able to capture the notion that it is cheaper to reveal D and E together than as speech acts at different stages of the dialogue. Without some form of lookahead to allow the agent to plan later moves, this extension is difficult to utilise. Once some form of lookahead exists, the addition of opponent modelling can further enhance the framework. Experimentally, evaluating the effects of various levels of lookahead, as well as different forms of opponent modelling might yield some interesting results.

Currently, we do not differentiate between information which the agent has explicitly committed to, and information that the agent has not yet disagreed with. More concretely, assume that the public knowledge base contains the argument $(\{\top\}, A)$. If an agent makes use of this argument, perhaps by submitting the argument $(\{A\}, B)$, then it is committed to the fact that A is true. If however, it never puts forth arguments making use of the fact, then an opponent cannot know if the agent is actually committed to A or not. We plan to extend our formalism and heuristic to capture this interaction in the near future.

Another extension that emerges from this line of reasoning is the concept of lying. An agent might commit to A to win an argument, even if its knowledge base contains only $\neg A$. How best to deal with this situation is an open question.

6 Conclusions

In this paper, we proposed a heuristic for argumentation based on minimising the cost of information revealed to other dialogue participants. While such an argumentation strategy arises in many real world situations, we are not familiar with any application that explicitly makes use of this technique. To study the heuristic, we proposed an argumentation framework that allowed us to focus on it in detail. Several novel features emerged from the interplay between the heuristic and the framework, including the ability of an agent to win an argument that it should not have been able to win (if all information were available to all dialogue participants). While we have only examined a very abstract model utilising the heuristic, we believe that many interesting extensions are possible.

Acknowledgements

This work is partly funded by the DTI/EPSRC E-Science Core Program and British Telecom, via a grant for the CONOISE-G project (http://www.conoise.org), a multi-university collaboration between Aberdeen, Cardiff and Southampton universities, and BT.

We would also like to thank the anonymous reviewers for their very useful comments.

REFERENCES

[1] Leila Amgoud and Nicolas Maudet, 'Strategical considerations for argumentative agents (preliminary report).', in *NMR*, pp. 399–407, (2002).

[2] Leila Amgoud and Henri Prade, 'Reaching agreement through argumentation: a possiblistic approach', in *Proceedings of KR 2004*, (2004).

[3] Trevor J.M Bench-Capon, 'Specification and implementation of Toulmin dialogue game', in *Proceedings of JURIX 98*, pp. 5–20, (1998).

[4] C. Cayrol, S. Doutre, M.-C. Lagasquie-Schiex, and J. Mengin, '" minimal defence" : a refinement of the preferred semantics for argumentation frameworks', in *Proceedings of NMR-2002*, (2002).

[5] Carlos Iván Chesñevar, Guillermo Ricardo Simari, and Lluis Godo, ‘Computing dialectical trees efficiently in possibilistic defeasible logic programming.’, in *Logic Programming and Nonmonotonic Reasoning, 8th International Conference, LPNMR 2005, Diamante, Italy, September 5-8, 2005, Proceedings (LNCS 3662)*, eds., Chitta Baral, Gianluigi Greco, Nicola Leone, and Giorgio Terracina, Lecture Notes in Computer Science, pp. 158–171. Springer, (2005).

[6] Phan Minh Dung, ‘On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games’, *Artificial Intelligence*, **77**(2), 321–357, (1995).

[7] Antonis C. Kakas, Nicolas Maudet, and Pavlos Moraitis, ‘Layered strategies and protocols for argumentation-based agent interaction.’, in *ArgMAS*, pp. 64–77, (2004).

[8] David Moore, *Dialogue game theory for intelligent tutoring systems*, Ph.D. dissertation, Leeds Metropolitan University, 1993.

[9] Nir Oren, Timothy J. Norman, and Alun Preece, ‘Arguing with confidential information’, in *Proceedings of the 18th European Conference on Artificial Intelligence*, Riva del Garda, Italy, (August 2006). (To appear).

[10] Nir Oren, Timothy J. Norman, and Alun Preece, ‘Loose lips sink ships: a heuristic for argumentation’, in *Proceedings of the Third International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2006)*, pp. 121–134, Hakodate, Japan, (May 2006).

[11] John L. Pollock, ‘Perceiving and reasoning about a changing world.’, *Computational Intelligence*, **14**, 498–562, (1998).

[12] Henry Prakken and Giovanni Sartor, ‘A dialectical model of assessing conflicting arguments in legal reasoning’, *Artificial Intelligence and Law*, **4**, 331–368, (1996).

[13] Chris. A. Reed and Douglas N. Walton, ‘Applications of argumentation schemes’, in *Proceedings of the 4th Conference of the Ontario Society for the Study of Argument (OSSA2001), Windsor, Canada*, eds., H. V. Hansen, C. W. Tindale, J. A. Blair, R. H. Johnson, and Robert C. Pinto, p. CD ROM, (2001).

[14] Katia Sycara, ‘Persuasive argumentation in negotiation’, *Theory and Decision*, **28**(3), 203–242, (May 1990).

[15] Bart Verheij, ‘DefLog: On the logical interpretation of prima facie justified assumptions’, *Journal of Logic and Computation*, **13**(3), 319–416, (2003).

Representing and Querying Arguments in RDF

Iyad Rahwan[1], [2] and **P. V. Sakeer**[3]

Abstract. This paper demonstrates the potential of the Semantic Web as a platform for representing, navigating and processing arguments on a global scale. We use the RDF Schema (RDFS) ontology language to specify the ontology of the recently proposed Argument Interchange Format (AIF) and an extension thereof to Toulmin's argument scheme. We build a prototype Web-based system for demonstrating basic querying for argument structures expressed in the Resource Description Framework (RDF). An RDF repository is created using the Sesame open source RDF server, and can be accessed via a user interface that implements various user-defined queries.

1 Introduction

The theory of argumentation has found a wide range of applications in both theoretical and practical branches of artificial intelligence and computer science [9, 4, 3]. Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge [13, page 5]. In a computational or multi-agent system, the *rational judge* could correspond to a particular choice of rules for computing the acceptable arguments or deciding the agent that wins the argument. Moreover, the *standpoint* may not necessarily be propositional, and should be taken in the broadest sense (e.g. it may refer to a decision or a value judgement). Finally, the term *controversial* should also be taken in the broad sense to mean "subject to potential conflict."

While argumentation mark-up languages, such as AML Araucaria [10], already exist, they are primarily a means to enable users to structure arguments through diagramatic linkage of natural language sentences. Moreover, these mark-up languages do not have clear and rich semantics, and are therefore not designed to process formal logical statements such as those used within multi-agent systems.

In response to the above limitation, an effort towards a standard Argument Interchange Format (AIF) has recently commenced [15]. The aim was to consolidate the work that has already been done in argumentation mark-up languages and multi-agent systems frameworks, and in particular facilitate: (i) argument interchange between agents within a particular multi-agent framework; (ii) argument interchange between agents across separate multi-agent frameworks; (iii) inspection/manipulation of agent arguments through argument visualisation tools; and (iv) interchange between argumentation visualisation tools.

This paper presents preliminary attempts to build a Web-based system for navigating and querying argument structures expressed in the Resource Description Framework (RDF). The RDF representation of arguments conforms to an ontology of arguments, which based on the AIF specification and expressed in the RDF Schema language. By expressing the AIF ontology in a standard format (namely RDF), it becomes possible to use a variety of Semantic Web tools (e.g. RDF query engines) to access and process arguments. This approach opens up many possibilities for automatic argument processing on a global scale.

The rest of the paper is organised as follows. In the next Section, we summarise the current state of the Argument Interchange Format specification. In Section 3, we describe how RDF and RDF Schema can be used to specify argument structures. We conclude the paper with a discussion in Section 4.

2 The Argument Interchange Format Ontology

In this section, we provide a brief overview of the current state of the Argument Interchange Format.[4] The AIF is a core ontology of argument-related concepts. This core ontology is specified in such a way that it can be extended to capture a variety of argumentation formalisms and schemes. To maintain generality, the AIF core ontology assumes that argument entities can be represented as nodes in a directed graph (di-graph). This di-graph is informally called an *argument network* (AN).

2.1 Nodes

There are two kinds of nodes in the AIF, namely, *information nodes* (I-nodes) and scheme application nodes or *scheme nodes* (S-nodes) for short. Roughly speaking, I-Nodes contain content that represent declarative aspects of the the domain of discourse, such as claims, data, evidence, propositions etc. On the other hand, S-nodes are applications of *schemes*. Such schemes may be considered as domain-independent patterns of reasoning, including but not limited to rules of inference in deductive logics. The present ontology deals with two different types of schemes, namely *inference schemes* and *attack schemes*. Potentially scheme types could exist, such as evaluation schemes and scenario schemes, which will not be addressed here.

If a scheme application node is an application of an inference scheme it is called a *rule of inference application node* (RA-node). If a scheme application node is an application of a preference scheme it is called a *preference application node* (PA-node). Informally, RA-nodes can be seen as applications of rules of inference while PA-nodes can be seen as applications of (possibly abstract) criteria of preference among evaluated nodes.

[1] Institute of Informatics, British University in Dubai, PO Box 502216, Dubai, UAE, email: irahwan@acm.org

[2] (Fellow) School of Informatics, University of Edinburgh, Edinburgh, UK

[3] Institute of Informatics, British University in Dubai, PO Box 502216, Dubai, UAE

[4] We will use the AIF specification as of April 2005 [15]).

2.2 Node Attributes

Nodes may possess different attributes that represent things like title, text, creator, type (e.g. decision, action, goal, belief), creation date, evaluation, strength, acceptability, and polarity (e.g. with values of either “pro” or “con”). These attributes may vary and are not part of the core ontology. Attributes may be intrinsic (e.g. “evidence”), or may be derived from other attributes (e.g. “acceptability” of a claim may be based on computing the “strength” of supporting and attacking arguments).

2.3 Edges

According to the AIF core ontology, edges in an argument network can represent all sorts of (directed) relationships between nodes, but do not necessarily have to be labelled with semantic pointers. A node A is said to *support* node B if and only if an edge runs from A to B.[5]

There are two types of edges, namely *scheme edges* and *data edges*. Scheme edges emanate from S-nodes and are meant to support conclusions. These conclusions may either be I-nodes or S-nodes. Data edges emanate from I-nodes, necessarily end in S-nodes, and are meant to supply data, or information, to scheme applications. In this way, one may speak of I-to-S edges (e.g. representing “information,” or “data” supplied to a scheme), S-to-I edges (e.g. representing a “conclusion” supplied by a scheme) and S-to-S edges (e.g. representing one scheme’s attack against another scheme).

	to *I-node*	to *RA-node*	to *PA-node*
from *I-node*		data/information used in applying an inference	data/information used in applying a preference
from *RA-node*	inferring a conclusion in the form of a claim	inferring a conclusion in the form of a scheme application	inferring a conclusion in the form of a preference application
from *PA-node*	applying preferences among information (goals, beliefs, ..)	applying preferences among inference applications	meta-preferences: applying preferences among preference applications

Table 1. Informal semantics of support.

2.4 Extending the Ontology: Toulmin’s Argument Scheme

Philosopher Stephen Toulmin presented a general argument scheme for analysing argumentation. Toulmin’s scheme, which has recently become influential in the computational modelling of argumentation, consists of a number of elements which are often depicted as follows:

$$\begin{tikzcd} D \arrow[r] \arrow[d, dash] & Q, C \arrow[d, dash] \\ \text{since } W \arrow[d, dash] & \text{unless } R \\ B & \end{tikzcd}$$

The various elements are interpreted as follows:

Claim (C): This is the assertion that the argument backs.

Data (D): The evidence (e.g. fact, an example, statistics) that supports the claim.

Warrant (W): This is what holds the argument together, linking the evidence to the claim.

Backing (B): The backing supports the warrant; it acts as an evidence for the warrant.

Rebuttal (R): A rebuttal is an argument that might be made against the claim, and is explicitly acknowledged in the argument.

Qualifier (Q): This elements qualifies the conditions under which the argument holds.

An example of an argument expressed according to Toulmin’s scheme can be as follows. The war in Irat (a fictional country) is justified (C) because there are weapons of mass destruction (WMDs) in Irat (D) and all countries with weapons of mass destructions must be attacked (W). Countries with WMDs must be attacked because they pose danger to others (B). This argument for war on Irat can be rebutted if the public do not believe the CIA reports about Irat possessing WMDs (R). Finally, this argument only holds if attacking Irat is less damaging than the potential damage posed by its WMDs (Q).

Toulmin’s argument scheme may be represented as an extension of the AIF core ontology. In particular, the concepts of *claim*, *data*, *backing*, *qualifier* and *rebuttal* are all expressed as sub-classes of I-Node. The concept of *warrant*, on the other hand, is an extension of RA-Nodes. This is because the former concepts all represent passive declarative knowledge, while the warrant is what holds the scheme together. In addition, since I-Nodes cannot be linked directly to one another, we introduce two new extensions of RA-Nodes. The new *qualifier-application* nodes link qualifier nodes to claim nodes, while *rebuttal-application* nodes link rebuttal nodes to claim nodes.

3 Arguments in RDF/RDFS

In this section, we describe the specification of the AIF ontology, and its extension to Toulmin’s argument scheme, in RDF Schema.

3.1 Background: XML, RDF and RDFS

The Extensible Mark-up Language (XML) is a W3C standard language for describing document structures by *tagging* parts of documents. XML documents provide means for nesting tagged *elements*, resulting in a directed tree-based structure. The XML Document Type Definition (DTD) and XML Schema languages can be used to describe different *types* of XML documents.

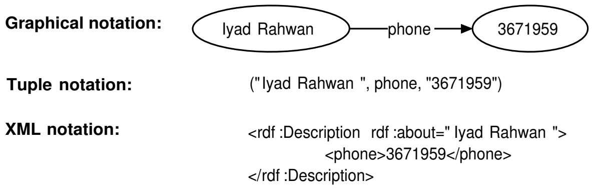
The Resource Description Framework (RDF)[6] is a general framework for describing Internet resources. RDF defines a resource as any object that is uniquely identifiable by an Uniform Resource Identifier (URI). Properties (or attributes) of resources are defined using an object-attribute-value triple, called a *statement*.[7] RDF statements can be represented as 3-tuples, as directed graphs, or using a standard XML-based syntax. The different notations are shown in Figure 1. Attributes are sometimes referred to as *properties* or *predicates*.

Unlike XML, which describes document models in directed-tree-based nesting of elements, RDF’s model is based on arbitrary graphs. This structure is better suited for creating conceptual domain models. RDF provides a more concise way of describing rich semantic information about resources. As a result, more efficient representation, querying and processing of domain models become possible.

[5] Note that this is a rather lose use of the word “support” and is different from the notion of “support between arguments” in which one argument supports the acceptability of another argument.

[6] `http://www.w3.org/RDF/`

[7] Sometimes, an *attribute* is referred to as a *property* or a *slot*.



("Iyad Rahwan ", phone, "3671959")

```
<rdf :Description  rdf :about=" Iyad Rahwan ">
          <phone>3671959</phone>
</rdf :Description>
```

Figure 1. Different notations for RDF statements

RDF Schema (RDFS)[8] is an (ontology) language for describing vocabularies in RDF using terms described in the RDF Schema specification. RDFS provides mechanisms for describing characteristics of resources, such as the domains and ranges of properties, classes of resources, or class taxonomies. RDFS (vocabulary describing) statements are themselves described using RDF triples.

3.2 AIF and Toulmin's Scheme in RDF Schema

We have first specified the AIF core ontology in RDFS using the Protégé ontology development environment.[9] The main class `Node` was specialised to three types of nodes: `I-Node`, `S-Node` and `Conflict-Node`. The `S-Node` class was further specialised to two more classes: `PA-Node` and `RA-Node`. For example, the following RDFS code declares the class `PA-Node` and states that it is a sub-class of the class `S-Node`.

```
<rdfs:Class rdf:about="&kb;PA_Node"
            rdfs:label="PA_Node">
  <rdfs:subClassOf rdf:resource="&kb;S-Node"/>
</rdfs:Class>
```

Next, the following elements from Toulmin's scheme were introduced as I-Nodes: claim, data, backing, rebuttal, and qualifier. All these elements represent passive declarative knowledge. Toulmin's warrant was expressed as an RA-Node, since it holds part of the argument together, namely the data nodes and the claim. Similarly, we introduced two other types of RA-Nodes: `Rebuttal-Application` nodes are used to link rebuttal nodes to claims, while `Qualifier-Application` nodes are used to link qualifier nodes to claims. The resulting ontology is represented in Figure 2.

Note that the concept `ToulminArgument` is a standalone concept. Instances of this concept will express complete arguments expressed in Toulmin's scheme. Such instances must therefore refer to instances of the various elements of the scheme. The ontology imposes a number of restrictions on these elements and their interrelationships. In particular, each Toulmin argument must contain exactly one claim, exactly one warrant, exactly one qualifier, at least one backing, and at least one data. The following RDFS code declares the property `claim` which links instances of `ToulminArgument` to instances of type `Claim`, and states that each `ToulminArgument` must be linked to exactly one `Claim`:

```
<rdf:Property rdf:about="&kb;claim"
    a:maxCardinality="1"
    a:minCardinality="1"
    rdfs:label="claim">
  <rdfs:domain rdf:resource="&kb;ToulminArgument"/>
  <rdfs:range rdf:resource="&kb;Claim"/>
</rdf:Property>
```

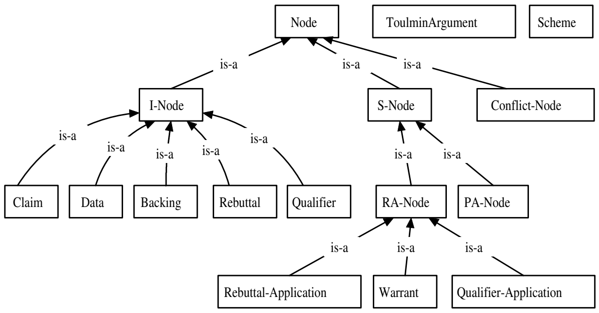


Figure 2. Toulmin argument class hierarchy as an extension of AIF ontology

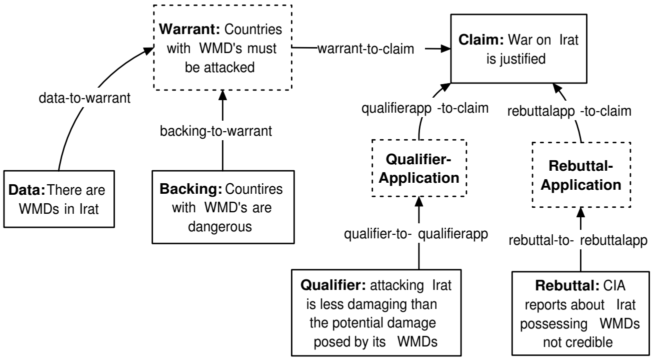

Warrant: Countries with WMD's must be attacked
warrant-to-claim
Claim: War on Irat is justified
data-to-warrant
backing-to-warrant
qualifierapp -to-claim
rebuttalapp -to-claim
Qualifier-Application
Rebuttal-Application
Data: There are WMDs in Irat
Backing: Countires with WMD's are dangerous
qualifier-to- qualifierapp
rebuttal-to- rebuttalapp
Qualifier: attacking Irat is less damaging than the potential damage posed by its WMDs
Rebuttal: CIA reports about Irat possessing WMDs not credible

Figure 3. RDF graph for a Toulmin argument

In our ontology, we defined various types of edges to capture every type of edge, such as those that emanate from backing nodes to warrant nodes, those from warrants to claims, and so on.

Note that according to our ontology, a single claim node can belong to multiple instances of Toulmin arguments. For example, a single claim may be supported by multiple arguments. Moreover, a single data node could contribute to multiple unrelated claims. The RDF graph model enables such flexibility.

With the ontology in place, it is now possible to create instances of the Toulmin argument scheme in RDF. Figure 3 shows an instance representing the argument mentioned above for justifying the war on Irat. In the Figure, we distinguished S-Nodes by dotted boxes although they are treated the same from the point of view of RDF processing.

3.3 Deploying an RDF Repository of Arguments

Our ultimate aim is to provide an infrastructure for publishing semantically annotated arguments on the *Semantic Web* using a language that is semantically rich and amenable to machine processing. The choice of RDF as a representation language was motivated by its expressive power and the availability of tools for navigating and processing RDF statements.

[8] `http://www.w3.org/TR/rdf-schema/`

[9] `http://protege.stanford.edu/`

In order to test our idea, we upladed the argument instances on Sesame:[10] an open source RDF repository with support for RDF Schema inferencing and querying. Sesame can be deployed on top of a variety of storage systems (relational databases, in-memory, filesystems, keyword indexers, etc.), and offers a large set of tools to developers to leverage the power of RDF and RDF Schema, such as a flexible access API, which supports both local and remote access, and several query languages, such as RQL and SeRQL. Sesame itself was deployed on the Apache Tomcat server, which is essentially a Java servlet container.

We have written a number of queries to demonstrate the applicability of our approach. The following query retrieves all warrants, data and backings for the different arguments in favour of the claim that "War in Irat justified."

```
select WARRANT-TEXT, DATA-TEXT, BACKING-TEXT
from {WARRANT} kb:scheme-edge-warrant-to-claim {CLAIM},
     {WARRANT} kb:text {WARRANT-TEXT},
     {DATA} kb:data-edge-data-to-warrant {WARRANT},
     {DATA} kb:text {DATA-TEXT},
     {BACKING} kb:data-edge-backing-to-warrant {WARRANT},
     {BACKING} kb:text {BACKING-TEXT},
     {CLAIM} kb:text {CLAIM-TEXT}
where
     CLAIM-TEXT like "War in Irat justified"
using namespace    kb = http://protege.stanford.edu/kb#
```

The output of the above query returned by Sesame will be the following:

WARRANT-TEXT	DATA-TEXT	BACKING-TEXT
Countries with WMDs must be attacked	There are WMDs in Irat	Countries with WMDs are dangerous

Query results can be retrieved via Sesame in XML for further processing. In this way, we could build a more comprehensive system for navigating argument structures through an interactive user interface that triggers such queries.

4 Discussion and Conclusion

A number of argument mark-up languages have been proposed. For example, the Assurance and Safety Case Environment (ASCE)[11] is a graphical and narrative authoring tool for developing and managing assurance cases, safety cases and other complex project documentation. ASCE relies on an ontology for *arguments about safety* based on *claims*, *arguments* and *evidence* [6].

Another mark-up language was developed for Compendium,[12] a semantic hypertext concept mapping tool. The Compendium argument ontology enables constructing *Issue Based Information System (IBIS)* networks, in which nodes represent *issues*, *positions* and *arguments* [5].

A third mark-up language is the argument-markup language (AML) behind the Araucaria system,[13] an XML-based language [10]. The syntax of AML is specified in a Document Type Definition (DTD) which imposes structural constraints on the form of legal AML documents. AML was primarily produced for use in the Araucaria tool. For example, the DTD could state that the definition of an argument scheme must include a name and any number of critical questions.

ClaiMaker and related technologies [12] provide a set of tools for individuals or distributed communities to publish and contest ideas and arguments, as is required in contested domains such as research literatures, intelligence analysis, or public debate. It provides tools for constructing argument maps, and a server on which they can then be published, navigated, filtered and visualized using the *ClaimFinder* semantic search and navigation tools [2]. However, again, this system is based on a specific ontology called the *ScholOnto* ontology [11].

The above attempts at providing argument mark-up languages share the following limitation. Each of the above mark-up languages is designed for use with a specific tool, usually for the purpose of facilitating argument visualisation. It was not intended for facilitating inter-operability of arguments among a variety of tools. As a consequence, the semantics of arguments specified using these languages is tightly coupled with particular schemes to be interpreted in a specific tool and according to a specific underlying theory. For example, arguments in Compendium are interpreted in relation to a specific theory of *issue-based information systems*. In order to enable true interoperability of arguments and argument structures, we need an argument description language that can be extended in order to accommodate a variety of argumentation theories and schemes. The AIF, as captured in RDF/RDFS, has the potential to form the basis for such a language.

Another limitation of the above argument mark-up languages is that they are primarily aimed at enabling users to structure arguments through diagramatic linkage of natural language sentences [7]. Hence, these mark-up languages are not designed to process formal logical statements such as those used within multi-agent systems. For example, AML imposes structural limitations on legal arguments, but provides no semantic model. Such semantic model is needed in order to enable the automatic processing of argument structures by software agents.

Our future plans include extending the AIF core ontology to other argument schemes, such as Walton's schemes for presumptive reasoning [14]. By doing so, we hope to validate the applicability of our approach and identify the limitations of RDF and RDFS for representing argument structures. It may well be that a more expressive ontology language is needed, such as OWL [8].

Another future direction for our work is to build applications that exploit the rich semantics of arguments provided by Semantic Web ontologies. Such applications could range from sophisticated argument processing and navigation tools to support human interaction with argument content, to purely automated applications involving multiple interacting agents operating on Web-based argument structures.

ACKNOWLEDGEMENTS

The first author is grateful to *AgentLink*[14] and ASPIC [1] for financial support to attend the AgentLink Argumentation Interchange Format Technical Forum in Budapest during September 2005, and to the AIF Technical Forum Group for many stimulating discussions.

REFERENCES

[1] ASPIC. Argumentation service platform with integrated components, a European Commission-funded research project (no. ist-fp6-002307), 2004.

[2] Neil Benn, Simon Buckingham Shum, and John Domingue, 'Integrating scholarly argumentation, texts and community: Towards an ontology and services', in *Proceedings of the Fifth Workshop on Computational Models of Natural Argument (CMNA 2005)*, (2005).

[10] www.openrdf.org/

[11] www.adelard.co.uk/software/asce/

[12] www.compendiuminstitute.org/tools/compendium.htm

[13] http://araucaria.computing.dundee.ac.uk/

[14] www.agentlink.org

[3] D. Carbogim, D. Robertson, and J. Lee, 'Argument-based applications to knowledge engineering', *Knowledge Engineering Review*, **15**(2), 119–149, (2000).

[4] C. I. Chesñevar, A. Maguitman, and R. Loui, 'Logical models of argument', *ACM Computing Surveys*, **32**(4), 337–383, (2000).

[5] Jeffrey Conklin and M. L. Begeman, 'gIBIS: a hypertext tool for exploratory policy discussion', *ACM transactions on office information systems*, **6**(4), 303–331, (1988).

[6] Luke Emmet and George Cleland, 'Graphical notations, narratives and persuasion: a pliant systems approach to hypertext tool design', in *HYPERTEXT 2002, Proceedings of the 13th ACM Conference on Hypertext and Hypermedia, June 11-15, 2002, University of Maryland, College Park, MD, USA*, pp. 55–64, New York, USA, (2002). ACM Press.

[7] *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, eds., Paul A. Kirschner, Simon J. Buckingham Schum, and Chad S. Carr, Springer Verlag, London, 2003.

[8] Deborah L. McGuinness and Frank van Harmelen, 'Web ontology language (owl): Overview', Technical report, W3C Working Draft, (31 March 2003).

[9] Henry Prakken and Gerard Vreeswijk, 'Logics for defeasible argumentation', in *Handbook of Philosophical Logic*, eds., D. Gabbay and F. Guenthner, volume 4, 219–318, Kluwer Academic Publishers, Dordrecht, Netherlands, second edn., (2002).

[10] G. W. A. Rowe, C. A. Reed, and J. Katzav, 'Araucaria: Marking up argument', in *European Conference on Computing and Philosophy*, (2003).

[11] Simon Buckingham Shum, Enrico Motta, and John Domingue, 'ScholOnto: An ontology-based digital library server for research documents and discourse', *International Journal of Digital Libraries*, **3**(3), 237–248, (2000).

[12] Simon Buckingham Shum, Victoria Uren, Gangmin Li, Bartrand Sereno, and Clara Mancini, 'Modelling naturalistic argumentation in research literatures: Representation and interaction design issues', *International Journal of Intelligent Systems, Special Issue on Computational Modelling of Naturalistic Argumentation (to appear)*, (2006).

[13] Frans H van Eemeren, Rob Flanery Grootendorst, and Francisca Snoeck Henkemans, *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Applications*, Lawrence Erlbaum Associates, Hillsdale NJ, USA, 1996.

[14] D. N. Walton, *Argumentation Schemes for Presumptive Reasoning*, Erlbaum, Mahwah NJ, USA, 1996.

[15] S. Willmott, G. Vreeswijk, C. Chesnevar, M. South, J. McGinis, S. Modgil, I. Rahwan, C. Reed, and G. Simari, 'Towards an argument interchange format for multiagent systems', in *Proceedings of the 3rd International Workshop on Argumentation in Multi-Agent Systems (ArgMAS), Hakodate, Japan*, eds., Nicolas Maudet, Simon Parsons, and Iyad Rahwan, (2006).

A critical review of argument visualization tools: do users become better reasoners?

Susan W. van den Braak[1] and **Herre van Oostendorp**[1] and **Henry Prakken**[2] and **Gerard A.W. Vreeswijk**[1]

Abstract. This paper provides an assessment of the most recent empirical research into the effectiveness of argument visualization tools. In particular, the methodological quality of the reported experiments and the conclusions drawn from them are critically examined. Their validity is determined and the methodological differences between them are clarified. The discrepancies in intended effects of use especially are investigated. Subsequently, methodological recommendations for future experiments are given.

1 Introduction

Argument diagramming is often claimed to be a powerful method to analyze and evaluate arguments. Since this task is laborious, researchers have turned to the development of software tools that support the construction and visualization of arguments in various representation formats, for instance, graphs or tables. As a result, several argument visualization tools currently exist [3], such as ArguMed [18], Araucaria [5], ATHENA [6], Convince Me [7], Compendium [8], Belvedere [9], ProSupport [4], and Reason!Able [15]. Typically, these tools produce “boxand arrow”diagrams in which premises and conclusions are formulated as statements. These are represented by nodes that can be joined by lines to display inferences. Arrows are used to indicate their direction.

Although it is often claimed that structuring and visualizing arguments in graphs is beneficial and provides faster learning, experiments that investigate the effects of these tools on the users’ reasoning skills are relatively sparse. Nevertheless, some experiments have been reported and the purpose of this paper is to critically examine their methodological quality and the conclusions drawn from them. Thus we aim to give an assessment of the state-of-the art in empirical research on the use of argument visualization tools, and to make some methodological recommendations for future experiments.

This paper is part of a larger research project on software support for crime investigations. Since reasoning is central to crime investigations and current support tools do not allow their users to make their underlying reasoning explicit, it is important to consider the use of argument visualization during these investigations. In this respect, it is also important to explore the effectiveness of such visualization tools.

The structure of this paper is as follows. Section 2 describes the criteria that will be used to evaluate the methodological quality of the experiments. The methods and results of these experiments are then discussed in Sections 3 and 4. Finally, Section 5 offers methodological recommendations to conduct future research.

2 Investigating the effectiveness of argument visualization tools

Among the tools that were experimentally tested for their effectiveness are Belvedere, Convince Me, Questmap, and Reason!Able. These have in common that they are education-oriented and designed to teach critical thinking or discussion skills, and are tested in an educational setting, for instance, on students during a course. Also, important discrepancies exist, for example, Belvedere and Reason!Able are entirely designed to assist argument construction and analysis, while Convince Me produces causal networks. Questmap has different main purposes, namely collaborative decision making, but it supports the construction of argument structures to a certain degree. Furthermore, Belvedere and Questmap are tested during collaborative reasoning, while Reason!Able is used by a single user. Most importantly, differences exist between the intended effects of use. Obviously, the latter affects the measures of effectiveness used and the tasks to be performed. This paper aims to provide an overview of these discrepancies.

In the remainder of this paper, the methods (viz. experimental designs, participants, and procedures) and results of the conducted experiments on argument visualization tools will be described. The aim of this is to find a general pattern or plan that may be followed to conduct research in this area. Moreover, we will determine whether these experiments were able to prove the long existing claim that visualization improves and simplifies reasoning. While describing the experimental methods, two important issues will be addressed, that is, the validity of the experiments and the related problem of finding a measure for the outcome variable, because these may affect their outcomes and the interpretations of their results. For this purpose, a checklist will be presented that allows us to assess their methodological quality. Additionally, this paper provides an overview of the proposed measures and their reliability.

2.1 Validity

If empirical experiments are conducted, it is important to take into account the validity of the experiment. Validity is mainly concerned with the question of whether the experiment really measures what it is supposed to measure. Two important types that we will consider in this paper are internal validity and external validity [2, 19]. Internal validity is the extent to which the differences in values of the dependent variable (the outcome variable) were actually caused by the independent variable (the variable that is manipulated by the experimenter) and not by some other source of variation. The external validity of an experiment is concerned with the following question: how well do the results of the experiment generalize beyond the

[1] Department of Information and Computing Sciences, Utrecht University, the Netherlands

[2] Faculty of Law, University of Groningen, the Netherlands

sample of subjects in the experiment and the particular experimental manipulations to the rest of the possible situations of the real world?

Besides evaluating the validity of an experiment, it is also important to consider the reliability of the measures used and the experiment conducted. If an experiment or measure is reliable, it means that it yields consistent results. In order for a measure to be reliable (or accurate) the results should be reproducible and as little as possible be influencedby chance.

It should be noted that validity implies reliability but not the other way around. Validity refers to obtaining results that accurately reflect the concept being measured, while reliability refers to the accuracy of the scores of a measure.

Generally, internal validity is assured by assigning subjects to treatment groups and control groups randomly. Experiments that use randomization and that are internally valid are sometimes called "true" experiments. Experiments that approximate these internally valid experiments but do not involve randomization are called quasi-experimental. This means that a valid experiment should at the very least have the participants assigned to conditions randomly, so that the external variables are under control and internal validity is maintained.

However, internal validity is not easy to obtain and is dependent on the chosen design. In a *between-subjects* design the participants are used only once and are part of the treatment group or the control group but differences between participants cannot be completely controlled. To cancel out the influence of relevant pre-existing differences between groups on the results, the treatment and control groups have to be matched or homogenized. For this reason, random assignment of subjects to conditions is crucial. Another solution to avoid effects of external variables is the use of a *within-subjects* design. In such a design all participants are used twice, as they receive both treatments. In order to cancel out any carryover effects, such as learning, practice, or fatigue effects, participants have to be assigned in such a way that different subjects receive both treatments in different orders (i.e. counterbalancing). Basically, these methods of randomization, counterbalancing, matching, and homogenization help to ensure internal validity.

External validity is affected by the design and subjects chosen. In order to assure external validity, the experimenter has to make sure that the experiment is conducted with the right participants as subjects, in the right environment, and with the right timing. Therefore, the experimental environment should be as realistic as possible. Additionally, the subjects should be selected from the population randomly. Finally, to check for external validity, the experiment should be replicated in other settings, with other subject populations, and with other, but related variables.

Table 1. Criteria for experimental validity

	Criteria
Reliability	use consistent measures
Internal validity	use at least one control group assign participants to conditions randomly match or homogenize (between-subjects designs) counterbalance (within-subjects designs)
External validity	draw a random sample from a population use real world settings and stimuli replicate the experiment

Obviously, since experimenters try to prove the effectiveness of their tool by justifying causal relations between the use of the tool and the users' reasoning skills, their research should preferably be done through laboratory experiments that are valid; the criteria are summarized in Table 1. Unfortunately, as we will see below, this is not often the case so that valid conclusions cannot be drawn.

2.2 Measures

The goal of the experiments described in this paper is to measure the effectiveness of a tool. The effectiveness describes the effect on the users' ability to reason (e.g. did these tools make their users better reasoners?). However, defining a measure for this is not straightforward. It is even hard to find an objective, reliable measure, that accurately measures the users' progress in reasoning skills. Moreover, to allow for statistical comparison, a quantitative measure has to be used, but such a generally accepted reliable measure is not available yet, as can be concluded from the large amount of different measures used. Generally, scores on critical thinking tests or assignments assessed by experts are used as measures for learning outcomes. These seem to be the only feasible and most reliable ways to measure reasoning skills in a quantitative way. However, as said, not all tools are designed with the same effects of use in mind. In some cases, the effectiveness of a tool is measured by the quality of the constructed argument. In other cases it is measured by the amount of discussion or the coherence of the arguments. It is important to be aware of these differences and their influence on the experimental tasks and the conclusions drawn from them.

3 Methods and results

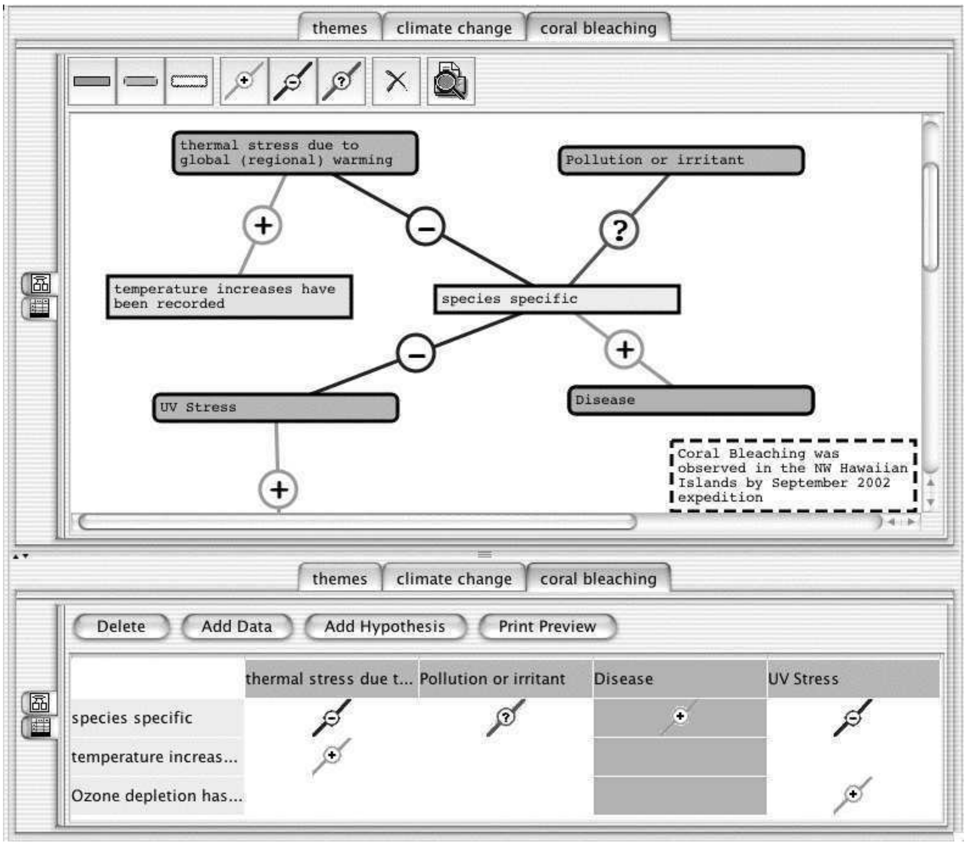
In this section a detailed description of the reported methods and results of the experiments on Belvedere, Convince Me, Questmap, and Reason!Able is given. Their validity will be assessed and their conclusions will be critically examined.

3.1 Belvedere

Belvedere [9] is a tool that is designed to support scientific argumentation skills in students and to stimulate discussions on scientific topics among students. With Belvedere students can build and display "inquiry diagrams" to model argumentation (see Figure 1). These diagrams consist of data nodes, hypothesis nodes, and unspecified nodes. Undirected links can be used to connect these nodes by for, against, and unspecified relations.

3.1.1 Method

Belvedere was tested in laboratory sessions and an in-school study [9] that investigated how well Belvedere facilitated the emergence of critical discussion. In the first set of sessions, the participants worked in pairs, using only one computer. The pairs were asked to resolve a conflict that was presented in textual and graphical form. The participants were also allowed to use a database with a small amount of relevant information. The second set was almost identical to the first set except that the participants worked on individual monitors and a shared drawing space. It should be noted that only two pairs of students participated in these sessions.


themes
climate change
coral bleaching
thermal stress due to global (regional) warming
Pollution or irritant
temperature increases have been recorded
species specific
UV Stress
Disease
Coral Bleaching was observed in the NW Hawaiian Islands by September 2002 expedition
+
−
?
+
−
+
themes
climate change
coral bleaching
Delete
Add Data
Add Hypothesis
Print Preview

	thermal stress due t...	Pollution or irritant	Disease	UV Stress
species specific	−	?	+	−
temperature increas...	+			
Ozone depletion has...				+

Figure 1. Screenshot of the Belvedere programme

The effect of Belvedere on the participants' critical discussion skills was measured by the amount of discussions that arose. This measure was rather a qualitative than a quantitative one, as the researchers mainly described the students' interactions. This experiment was not valid, because the measure was not valid and no control group was used to compare the experimental group to.

Further, to compare the effect of different representations on the learning outcomes, three different representation formats were tested in [10] and also [11] and [12]. This experiment was internally valid as it was based on a between-subjects design with three groups in which the participants were assigned to groups randomly. Moreover, there were no significant pre-existing differences between the groups' gender balance and mean grade point average due to homogenization. External validity was not guaranteed, because of the artificial nature of the task. It was very limited and was completed in a laboratory setting, while the effect was only measured during the initial use and not over a longer period of time.

The groups, consisting of 20 students each, were defined by the software they used, that is, matrix, graph, or text. All groups had to perform the same task of structuring an unsolved science challenge task into data, hypotheses, and evidential relations. Identical background information was presented to all three groups, one page at a time. The students had to work in pairs and were asked to use the given information in their representation of the problem, before continuing to the next page (the showed information would not remain available for later reference). After finishing their representation of the problem, the students had to complete a post-test containing multiple-choice questions and had to write a collaborative essay. These essays were scored according to the following measures:

- Evidential strength: the strength of the evidential relationship, on a scale of 0 to 4, with + indicating a supporting relationship, and − indicating a conflictingrelationship.
- Inferential difficulty: the number of information pages that must be accessed to infer the relationship, with 0 indicating that the relationship is explicitly stated in the material, and > 1 indicating that the relationship has to be inferred.
- Inferential spread: the difference (in pages) between the first and last page needed to infer the relationship. This is a measure of how well participants integrate information given at different pages.

In order to obtain a measure of the quality of the essay that was produced, an expert completed the task himself and his evidential matrix was used to compare the students' essays to. In this way, the students' ability to list the most important data and relations of the problem was measured. It thus measures the students' collaborative scientific discussion skills.

In sum, Belvedere has two aims: to support the amount of critical discussion and to enhance collaborative learning of reasoning skills. The former was tested in an internally invalid study, while the latter was investigated in an internally valid experiment. The tasks involved constructing arguments based on unstructured information in which the students had to identify data for and against their hypothesis.

3.1.2 Results

For the first set of experiments, the researchers only gave qualitative descriptions of the results. In the first set of sessions, the experimenters found an encouraging amount of discussion. In the second set they found that in one pair the students cooperated to a high degree, but that there was no interaction at all in the other pair.

For the in-school study it was found that sensible diagrams were produced, but that the use of shapes and link types was inconsistent. Moreover, it was found that students incorporated several points of the debate into diagrams.

On the basis of these observations, the authors concluded that Belvedere indeed stimulated critical discussions. However, although a tendency was shown, this experiment did not conclusively prove an effect as it was not internally valid. Conclusions drawn based on these studies are therefore premature. In this respect, the second experiment is more promising, because internal validity was achieved. Moreover, the documentation on the second experiment was considerably more detailed.

None of the test in the second experiment yielded a significant difference between the groups. From these results the researchers concluded that there were no significant differences in performance between the users that used matrix or graph representations and the users that used text only. According to the researchers, the lack of significance of the learning outcomes was disappointing, although the researchers noted that this was not surprising given the fact that the total amount of time spent working with Belvedere was too short for learning outcomes to develop.

It must be said that trends were in the predicted direction but not significant. This means that the students who were allowed to use the Belvedere software that contained matrix representations performed better than the students who used graph representations, who in turn performed better than the students who used text only. Therefore, a tendency is shown that visually structured representations can provide guidance for collaborative learning that is not provided by plain text only, while a significant difference could not be proven. This conclusion is legitimate since the experiment was internally valid.

3.2 Convince Me

Convince Me [7] is a tool for generating and analyzing argumentation and is designed to teach scientific reasoning. In addition, Convince Me provides feedback on the plausibility of the inferences drawn by the users as it predicts the user's evaluations of the hypotheses based on the produced arguments. It is based on Thagard's Theory of Explanatory Coherence [13]. Arguments in Convince Me consist of causal networks of nodes and the users' conclusion drawn from them (see Figure 2). Nodes can display either evidence or hypotheses. Explanatory or contradictory relations are represented as the undirected links between these nodes.

3.2.1 Method

The study described in [7] compared the performance of the participants who used Convince Me to the performance of paper and pencil users. In this study, 20 undergraduate students of Berkeley had to complete a pre-test (in which both groups had no access to the software), three curriculum units on scientific reasoning, integrative exercises (one group is allowed to use Convince Me, the other group is not allowed to do so), a post-test (nobody had access to Convince Me), and a questionnaire (to establish relevant differences between groups).

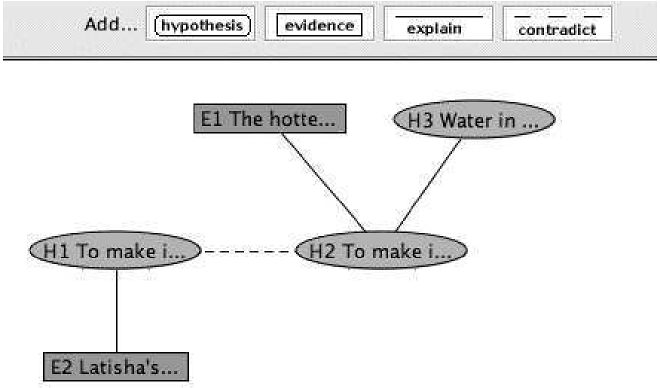


Figure 2. Screenshot of the Convince Me tool

The group that was allowed to use Convince Me consisted of 10 participants, the other 10 participants were part of the group that used paper and pencil only. Both groups received the same instructions and exercises. There were no significant difference between the groups in age, year in school, SAT scores, and total session hours.

This experiment used a between-subjects design. The potential effect of intergroup differences was not an issue here as the experimenters confirmed that the groups were homogeneous with respect to relevant variables. However, they did not mention whether randomization was used while assigning subjects to conditions. Therefore, it will be assumed that this experiment was at least quasi-experimental, but a definitive analysis of the experiments' validity cannot be made.

The following measures were used to measure the utility of the software:

1. How well the participants' beliefs are in accord with their argument structures.
2. The kinds of changes made when arguments are revised.

Only the first measure will be used in the description of the results that will presented below, because this is the most suitable of the two to measure the effectiveness of a tool. The latter only measures the stability of the arguments constructed, not the effect on the users' reasoning skills. The former is a measure of the arguments' coherence, that is, it shows whether people are able to construct arguments that reflecttheir beliefs properly.

So in short, Convince Me attempts to improve the coherence of its users' arguments so that users become more aware of the believability of their arguments. Note that this differs from the learning effect that was claimed by the developers of Belvedere. Required methodological information is missing so that a genuine assessment of the validity of this experiment cannot be made. Moreover, important details about the nature of the task were not reported.

3.2.2 Results

During the exercises, the participants' beliefs were more in accord with the structures of their arguments if they were using Convince Me, than if they were using paper ($p < 0.05$). Also during the post-test, the belief-argument correlations of Convince Me users were significantly higher ($p < 0.05$) and better than during the pre-test (see Figure 3).

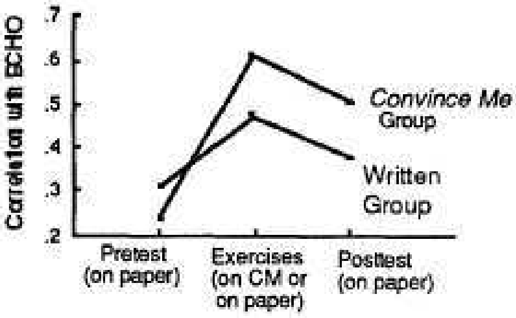


Figure 3. Results of the experimental testing of Convince Me, after [7]

Based on these results the experimenters claimed that the tool improved the users' argumentation skills and made them better reasoners. They also showed that these skills remained when the participants did not have access to the tool and were not supported by it, and that those were still better than the skills of the participants who did not use the tool at all. However, some reservation is appropriate here as the validity of the experiments is unknown.

3.3 Questmap

Questmap is designed to mediate discussions by creating visual information maps (see Figure 4), but is used by [1] to support collaborative argumentation in legal education. It is based on IBIS, an Issue-Based Information System that is designed for collaborative problem identification and solving. IBIS helps multiple users to discuss issues related to a problem and reach a consensus on a solution. Its main procedure involves decomposing the problem into issues. Possible answers to them are recorded as positions. Arguments for and against these positions may be recorded as well. Questmap provides many additional node types, including problems, claims, warrants, backing, and data nodes. By using these nodes, arguments can be constructed.

3.3.1 Method

In [1], the computer-based representational tool Questmap, was tested for its effect on legal argumentation skills.

The most important research question to be answered was: "How does using CSAV, while groups of three or four second-year law students generate arguments throughout the semester, affect the quality and type of arguments generated on a practice final exam (p. 81)". Also, a hypothesis was formulated: "groupsusing CSAV to construct arguments throughout the study will create higher quality arguments on a practice final exam than those who construct written arguments throughout the study. (p. 81)"

The quality of the produced arguments was measured by:

1. the number of arguments, counterarguments, rebuttals, and evidence present in the practice final exam
2. the scores on the final exams as assessed by the professor
3. the richness of arguments saved in Questmap throughout the semester measured by the number of nodes created (to describe the progress in the treatment group only)

The design was a quasi-experimental between-subjects design. The treatment group consisted of 33 law students who completed the assignments using Questmap in groups of three or four. The control group of 40 students completed the exercises individually using conventional methods. Participants were not randomly assigned to groups, because the participants were allowed to choose the group they wanted to participate in. On the other hand, the pre-test revealed that the groups were in fact homogeneous. This means that at least some internal validity was assured.

The students' argumentation skills were tested and trained throughout the semester. They had to complete five assignments that addressed current legal issues in relation to the admissability of the evidence. Both groups of students were allowed access to the same materials, but only the treatment group was allowed to use Questmap. Two of the assignments of the treatment group were analyzed to measure the progress throughout the semester.

At the end of the semester all participants completed a final exam without the use of Questmap. During this exam the students had to construct all relevant arguments to a given problem individually and without the use of legal resources. These exams were graded by the professor.

To sum up, Questmap claims to improve the quality of the users' arguments so that the users become better reasoners. The assignments involved producing answers to the problem that consisted of arguments, counterarguments, and rebuttals. In the experiment, internal validity was only partially assured.

3.3.2 Results

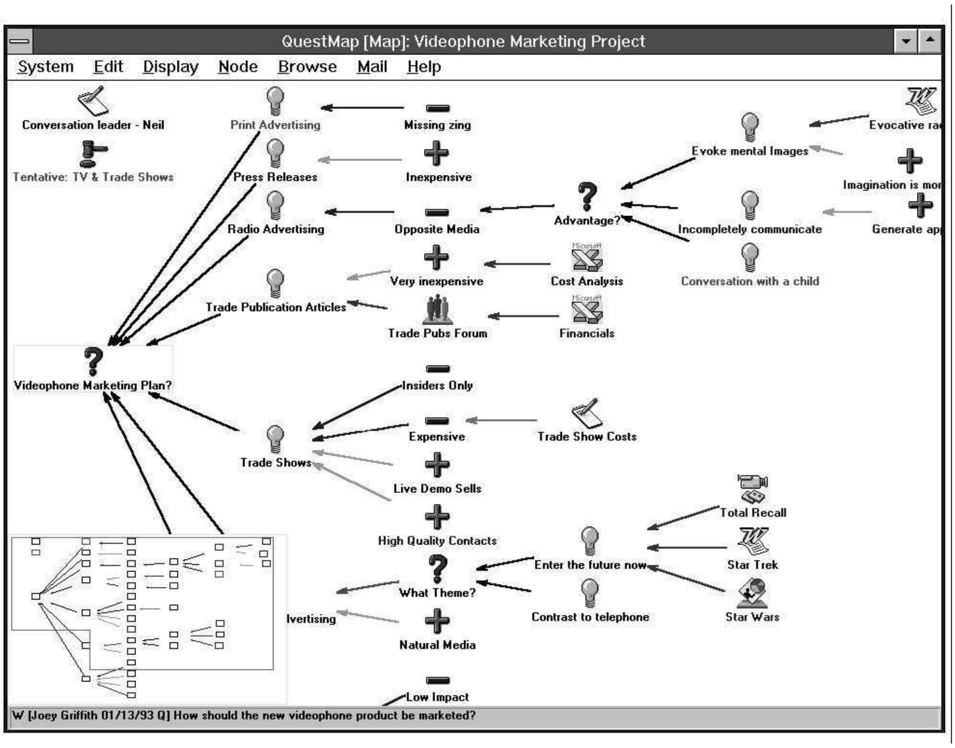
The found results show that there were no pre-existing differences between the groups ($p > 0.05$), that the arguments did not become more elaborate throughout the semester, and that the treatment group did not have a significantly higher score than the control group ($p > 0.05$). Based on these results, the experimenter claimed that the hypothesis did not hold and that law students who were allowed to use a computer supported argumentation tool did not perform better on the exam than students who only used paper and pencil during the course. On the other hand, it must be said that while the differences between the treatment and control group were not significant, a trend was discovered in the predicted direction (cf. mean $= 5.15$ and mean $= 4.50$ respectively, where $0.05 < p < 0.10$). However, the value of these observations is limited, as complete internal validity was not assured.

3.4 Reason!Able

Reason!Able [15] is educational software that supports argument mapping to teach reasoning skills. It provides support to the users by guiding them step-by-step through the construction process. The argument trees constructed by Reason!Able contain claims, reasons, and objections (see Figure 5). Reasons and objections are complex objects that can be unfolded to show the full set of premises and helping premises that are underlying them.

3.4.1 Method

In [15] and [16], the question of "doesit work" was addressed. To answer this question, all students who were part of a one-semester undergraduate Critical Thinking course at the University of Melbourne and used Reason!Able during this project, were asked to complete a pre-test and a post-test that was based on the California Critical



W [Joey Griffith 01/13/93 Q] How should the new videophone product be marketed?

Figure 4. Screenshot of Questmap

Thinking Skills Test. This test consisted of 34 multiple-choice questions. Obviously, this experiment was not internally valid, because no control group was used so that a valid comparison of the results is impossible, although the measure seems to be reliable.

A similar study was reported by [17] in which students were also pre-tested and post-tested using two tests, namely the California Critical Thinking Skills Test and written test in which students had to identify the main conclusions, reformulate the reasoning, and evaluate the reasoning of a short argumentative text. The latter was assessed by two experts. Methodological details were missing so no real assessment of the internal validity can be made. But since no direct control group was available, internal validity will be limited.

Another, more elaborate, study was reported in [14]. Students were learning argumentation skills during a period of 16 weeks; one group of 32 students participated in a traditional course, another group of 53 in a Reason!-based course. The latter was allowed to use Reason! (a predecessor of the Reason!Able programme) to construct argument trees. Both groups were pre-tested and post-tested using the Watson-Glaser Critical Thinking Appraisal; another multiple-choice test. The students in the Reason! group were also asked to complete the written pre-test and post-test. Although two groups were tested, those were not compared to each other. This means that no real between-subjects design was used. Moreover, it was not mentioned whether randomization was used. Therefore, this experiment cannot be considered to be internally valid.

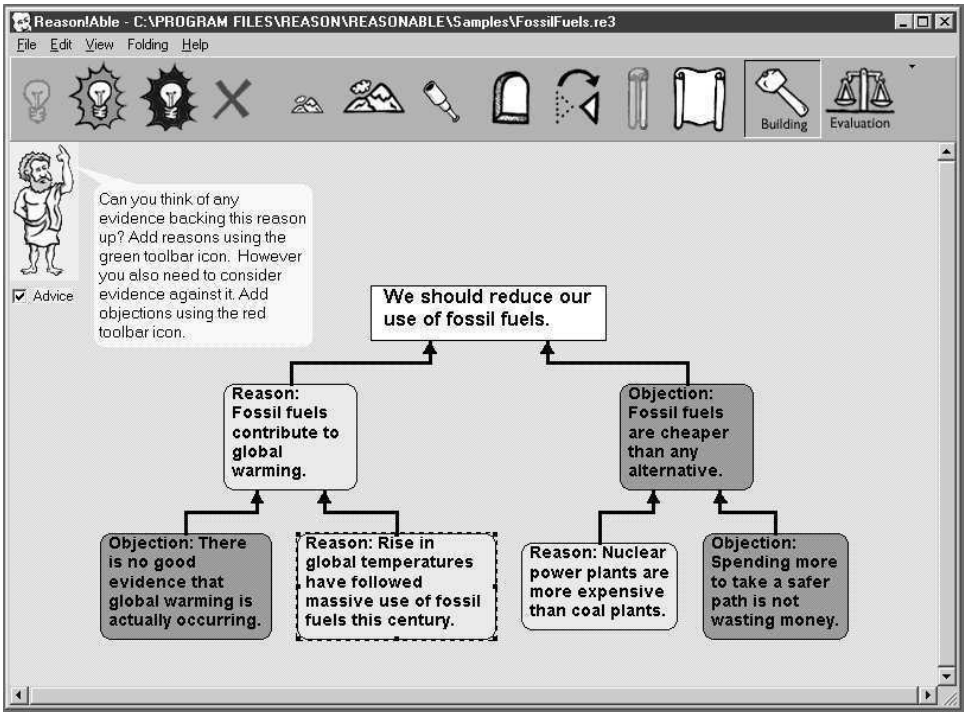
So, similar to Questmap, Reason!Able aims to provide support to make its users better reasoners. Several studies were performed, which were not internally valid. During the course, students had to produce their own arguments but the written pre-test and post-test consisted of the reproduction of an argument from an argumentative text. Similarly, the multiple-choice tests involved identifying proper arguments rather than constructing arguments. This means that the task that measured the students' skills considerably differed from the assignments during the course, although both involved the identification of arguments and counterarguments.

3.4.2 Results

In the first study it was found that the students' scores improved with almost 4 points over the last three years ($SD = 0.8$). Generally, it is assumed that the students' performance in any subject would normally be expected to improve by only 0.5 standard deviation over three years. From this the author concluded that the Reason! approach improved the students' critical thinking skills and was more effective than traditional approaches. Unfortunately, no valid experimental design was used to compare these results statistically.

Similarly in [16] and [17] it was claimed that the approach improved the students' skills more over one semester than traditional approaches that needed the entire undergraduate period to achieve the same result. Reason! was claimed to be three to four times more effective than traditional approaches that do not use the Reason!Able software. However, these claims seem to be premature, as the experiments were not valid.

In the last study, two groups of students were tested but not compared to each other. In [14] significant progress was reported for the Reason! users ($p < 0.05$) on both the multiple-choice and the written test, while the traditional group did not display a significant gain in reasoning skills. But since internal validity was not assured, no safe conclusion can be drawn.


Reason!Able - C:\PROGRAM FILES\REASON\REASONABLE\Samples\FossilFuels.re3
File Edit View Folding Help
Building
Evaluation
☑ Advice

Can you think of any evidence backing this reason up? Add reasons using the green toolbar icon. However you also need to consider evidence against it. Add objections using the red toolbar icon.

We should reduce our use of fossil fuels.

Reason: Fossil fuels contribute to global warming.

Objection: Fossil fuels are cheaper than any alternative.

Objection: There is no good evidence that global warming is actually occurring.

Reason: Rise in global temperatures have followed massive use of fossil fuels this century.

Reason: Nuclear power plants are more expensive than coal plants.

Objection: Spending more to take a safer path is not wasting money.

Figure 5. Screenshot of the Reason!Able software tool

4 Discussion

The experiments described above significantly differ. The most important methodological differences are concerned with the nature of the task that had to be performed, the measures used, and the underlying argumentation theory. These differences are summarized in Table 2.

With respect to the task, the main differences had to do with the intended effect of use. Also the nature of the tasks differed, as in some experiments the participants had to produce the arguments themselves, while in other ones reproduction of arguments based on a argumentative text was asked or multiple-choice test had to be completed. Moreover, sometimes collaboration was mandatory, while in other cases users had to work individually. In most experiments subjects had to establish supporting and attacking (or contradicting) relationships.

The measures that were used also differed. Although most of them involved expert assessment, there was a lack of information about the criteria that were used to assess the quality of the users' reasoning. Similarly, little is known about the contents of the multiple-choice tests. As far as the measures of argument quality are concerned, another important distinction has to be made. Two different aspects are measured, firstly, the quality of the arguments' structure. For example, this is measured by the number of nodes used (is there a sufficient amount of detail) or the validity of the structure. Secondly, the quality of the content of the argument is measured, for instance, by expert assessment.

It was found that most results indicated that the tools have a positive effect on argumentation skills and make the users better reasoners. However, most experiments did not yield significant effects. The observation that different underlying argumentation theories were used is relevant for the conclusions drawn. Results that are not significant may be caused by an underlying theory that is not suitable for the task at hand. For example, an IBIS-based system may not be suitable for the task of constructing legal arguments.

The difference in measured effects means that we have to divide our conclusions into three subconclusions on argument quality, argument coherence, and critical discussion skills. Significant effects were only found for argument coherence. For argument quality the effects were not significant, but trends were shown in the positive direction. These trends both concerned argument structure and content. No quantitative results were reported on discussion skills.

5 Conclusion and future work

This paper has provided a critical review of the most recent research into the effectiveness of argument visualization tools. Although it is promising that some researchers at least subjected their tools to testing, most of the experiments described in this paper were not completely valid. Sometimes it was even impossible to determine the validity of the results at all, as many important details were missing in the description of the experiments; in particular methodological and statistical details were not mentioned. As a consequence, due to a lack of internal validity, the differences found may not be completely caused by the use of the visualization tool but may have additional causes and due to a lack of external validity, the results cannot easily be generalized to other populations. Therefore, it is premature to claim that argument visualization tools cause higher quality arguments, critical discussion, or coherent arguments. But given the fact that most results point in the same direction, we think it is reasonable to assume that these tools have a positive effect on the users' argumentation skills.

However, a lot still remains to be done, because until now experiments have failed to provide significant evidence for the benefits of argument visualization tools. After all, significant differences have been found but only in invalid experiments, while in the internally

Table 2. Overview of methodological differences between experiments

	Experimental tasks				**Experimental measures**	**Argumentation theory**
	Effect of use	**Production**	**Links**	**Collaboration**		
Belvedere	critical discussion skills and quality of argument structure	production	attack and support	yes	amount of discussion, multiple-choice test, and expert assessment of essay by inferential strength, difficulty, and spread	arguments in terms of inference trees
Convince Me	argument coherence (structure)	*unknown*	*unknown*	*unknown*	correlation with ECHO	Thagard's theory of explanatory coherence
Questmap	quality of both argument structure and content	production	attack and support	yes but not mandatory in control group and not during post-test	the number of argument structures, the richness of arguments, and expert assessment of final exam	IBIS
Reason!Able	quality of argument content	reproduction (pre-test and post-test)	attack and support	no	multiple-choice critical thinking skills tests and expert assessment of written test	arguments in terms of inference trees

valid experiment the results have been not significant. More specifically, based on our assessment of the internal validity, we have to further restate our conclusions and say that with respect to the experiments on Belvedere (the first experiment), Questmap, and Reason!Able, no real conclusions can be drawn. Valid conclusions can be drawn from the second experiment on Belvedere that failed to prove a significant effect on argument quality, although a trend was proven in the positive direction.

Nevertheless, the designs of these experiments and their shortcomings are useful to give recommendations for future research on computer-supported argument visualization. First, the experiment has to be valid, so that the results that are found and the conclusions that are drawn are valid and can be generalized to larger populations. More specifically, at least a between-subjects design should be used with one control group. Second, the chosen measure should be reliable. Therefore, a quantitative, objective measure for the effectiveness of a tool should be developed, but it should be noted that this is not straightforward. The most reliable measure found so far seems to be expert assessment, that is, specialists are asked to assess the quality of the argumentation by criteria such as the completeness and validity of the argument constructed.

Now we have come to the point at which an action plan to conduct research into the effectiveness of argument visualization tools can be given:

1. Formulation of hypotheses.
2. Selection of the variables, especially choosing a dependent variable that is based on a valid measurement.
3. Selection of the subjects, especially choosing a representative sample for the population the results have to be generalized to, other important issues include the sample size.
4. Selection of the design, especially choosing between a within-subjects or between-subjects design, other important issues involve randomization, homogenization (between-subjects design), and balancing (within-subjects design).
5. Selection of the appropriate statistical tests in order to draw valid conclusions.

Preferably, the usability and user-friendliness of the visualization tool is tested first, so that it is easy enough for everybody to understand and use, and its complexity does not limit the constructed arguments. Subsequently, other experiments can be conducted that measure its effectiveness.

In short, this paper has made a contribution to the area of empirical research on argument visualization tools, in that it paves the way for a more scientific approach to this research and provides an action plan to conduct experiments. It is also relevant to our research project on crime investigations, since the effectiveness of the tool we plan to develop will be tested. Unfortunately, to our knowledge no experiments focus on the effects of such tools on police investigations. We are cautious to generalize the results described in this paper to the domain of evidential reasoning in police investigations, as external validity was not assured and the domain differs both in the type and setting of the reasoning (cf. teaching versus crime solving). Most of the described experiments did not concentrate on the effects on evidential reasoning but focus on more general reasoning and conflict resolution skills. Critical discussion and collaborative problem solving are other skills that are of use to police investigators. Taking this into consideration the results on Belvedere and Questmap are most relevant here, though no significant effects were demonstrated. This means that a lot remains to be done in this area and that as far as we know the experiment we plan to conduct on police investigators will be the first of its kind.

ACKNOWLEDGEMENTS

This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number 634.000.429. Henry Prakken and Gerald Vreeswijk were also partially supported by the EU under IST-FP6-002307 (ASPIC).

REFERENCES

[1] Chad S. Carr, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, chapter Using computer supported argument visualization to teach legal argumentation, 75–96, Springer-Verlag, London, UK, 2003.

[2] Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton MifflinCompany, 1979.

[3] Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, Springer-Verlag, London, UK, 2003.

[4] Henry Prakken and Gerard A.W. Vreeswijk, 'Encoding schemes for a discourse support system for legal argument', in *Workshop Notes of the ECAI-02 Workshop on Computational Models of Natural Argument*, pp. 31–39, (2002).

[5] Chris A. Reed and Glenn W.A. Rowe, 'Araucaria: Software for argument analysis, diagramming and representation', *International Journal on Artificial Intelligence Tools*, **14**(3-4), 961–980, (2004).

[6] Bertil Rolf and Charlotte Magnusson, 'Developing the art of argumentation: A software approach', in *Proceedings of the Fifth Conference of the International Society for the Study of Argumentation*, (2002).

[7] Patricia Schank and Michael Ranney, 'Improved reasoning with Convince Me', in *CHI '95: Conference Companion on Human Factors in computing Systems*, pp. 276–277, New York, NY, (1995). ACM Press.

[8] Albert Selvin, Simon Buckingham Shum, Maarten Sierhuis, Jeff Conklin, Beatrix Zimmermann, Charles Palus, Wilfred Drath, David Horth, John Domingue, Enrico Motta, and Gangmin Li, 'Compendium: Making meetings into knowledge events', in *Proceedings Knowledge Technologies 2001*, (2001).

[9] Daniel Suthers, Arlene Weiner, John Connelly, and Massimo Paolucci, 'Belvedere: Engaging students in critical discussion of science and public policy issues', in *AI-Ed 95, the 7th World Conference on Artificial Intelligence in Education*, pp. 266–273, (1995).

[10] Daniel D. Suthers and Christopher D. Hundhausen, 'Learning by constructing collaborative representations: An empirical comparison of three alternatives', in *European Perspectives on Computer-Supported Collaborative Learning, Proceedings of the First European Conference on Computer-Supported Collaborative Learning*, eds., P. Dillenbourg, A. Eurelings, and K. Hakkarainen, pp. 577–584, Maastricht, the Netherlands, (2001).

[11] Daniel D. Suthers and Christopher D. Hundhausen, 'The effects of representation on students' elaborations in collaborative inquiry', in *Proceedings of Computer Support for Collaborative Learning 2002*, pp. 472–480. Hillsdale: Lawrence Erlbaum Associates, (2002).

[12] Daniel D. Suthers and Christopher D. Hundhausen, 'An empirical study of the effects of representational guidance on collaborative learning', *Journal of the Learning Sciences*, **12**(2), 183–219, (2003).

[13] Paul Thagard, 'Probabilistic networks and explanatory coherence', *Cognitive Science Quarterly*, **1**, 91–114, (2000).

[14] Tim J. van Gelder, 'Learning to reason: A Reason!-Able approach', in *Cognitive Science in Australia, 2000: Proceedings of the Fifth Australasian Cognitive Science Society Conference*, eds., C. Davis, T. J. van Gelder, and R. Wales, Adelaide, Australia, (2000).

[15] Tim J. van Gelder, 'Argument mapping with Reason!Able', *The American Philosophical Association Newsletter on Philosophy and Computers*, 85–90, (2002).

[16] Tim J. van Gelder, 'A Reason!Able approach to critical thinking', *Principal Matters: The Journal for Australasian Secondary School Leaders*, 34–36, (2002).

[17] Tim J. van Gelder and Alberto Rizzo, 'Reason!Able across the curriculum', in *Is IT an Odyssey in Learning? Proceedings of the 2001 Conference of ICT in Education*, Victoria, Australia, (2001).

[18] Bart Verheij, 'Artificial argument assistants for defeasible argumentation', *Artificial Intelligence*, **150**(1-2), 291–324, (2003).

[19] Claes Wohlin, Per Runeson, Martin Host, Magnus C. Ohlsson, Bjorn Regnell, and Anders Wesslen, *Experimentation in Software Engineering: An Introduction*, Kluwer Academic Publishers, Boston, MA, 2000.

A knowledge representation architecture for the construction of stories based on interpretation and evidence

Susan W. van den Braak and **Gerard A.W. Vreeswijk** [1]

Abstract. This paper describes *Stevie*, a knowledge representation architecture for the analysis of complex legal cases. *Stevie* is targeted at legal professionals who may use it to infer stories (plausible and consistent reconstructions of courses of events) from evidence and hypotheses. *Stevie* is based on known argument ontologies and argumentation logics.

1 INTRODUCTION

This paper describes *Stevie*, a knowledge representation architecture for making sense of evidence through stories and their justification. This system is targeted at criminal investigators who may use it to gain a better overview of complex cases. In the process of making sense of large quantities of data, it will enable crime investigators to formulate their hypotheses as stories of what might have happened and to make their underlying reasoning explicit.

In project meetings with crime investigators we learned that in the analysis of crime cases there is a demand for a support tool that offers the ability to search and combine large quantities of data. In fact, crime investigators already use powerful search tools to match possibly relevant data. What they seem to lack is functionality with which search results can be interpreted, explained, and related to each other in a larger context. *Stevie* is a first stab at the realization of such facilities.

With respect to argument visualization, the contribution of *Stevie* is threefold. Firstly, it represents cases (among others) as di-graphs rather than trees. Thus, unnecessary duplication of nodes is avoided. Further, *Stevie* possesses an inferential component to incorporate predefined argumentation schemes. This component also assesses the dialectical status of nodes to suggest plausible stories to analysts. Finally, it represents temporal information and is thus able to rule out stories that are temporally inconsistent.

2 SYSTEM PURPOSE

This section describes the context in which *Stevie* operates. It also describes the functionality that the system provides at its interfaces.

2.1 Context

Stevie provides support during criminal investigations by allowing case analysts to visualize evidence and their interpretation of that evidence in order to construct coherent stories. It allows them to maintain overview over all information collected during an investigation, so that different scenarios can be compared. Moreover, they are able to express the reasons why certain evidence supports the scenarios. In this way it may help them in seeing patterns, discovering inconsistencies and identifying missing evidence.

It must be emphasized that *Stevie* is not meant to be used in the preparation of trials; nor is it intended as a tool for modelling legal cases, since police and prosecution have different responsibilities. Crime analysts are supposed to follow promising leads, without too much concern about proving guilt in court. Once one or more suspects are determined, the prosecution takes over and *Stevie* drops out of the picture.

2.2 System interface

Stevie is presented as a web front-end to an SQL database (Fig. 1). Users log in and create a case record, or select a case which they want to work on. Each case is presented in a split screen where the upper half displays a global overview of the case and the lower half displays the attributes of a node that is selected by the user in the upper half of the screen.

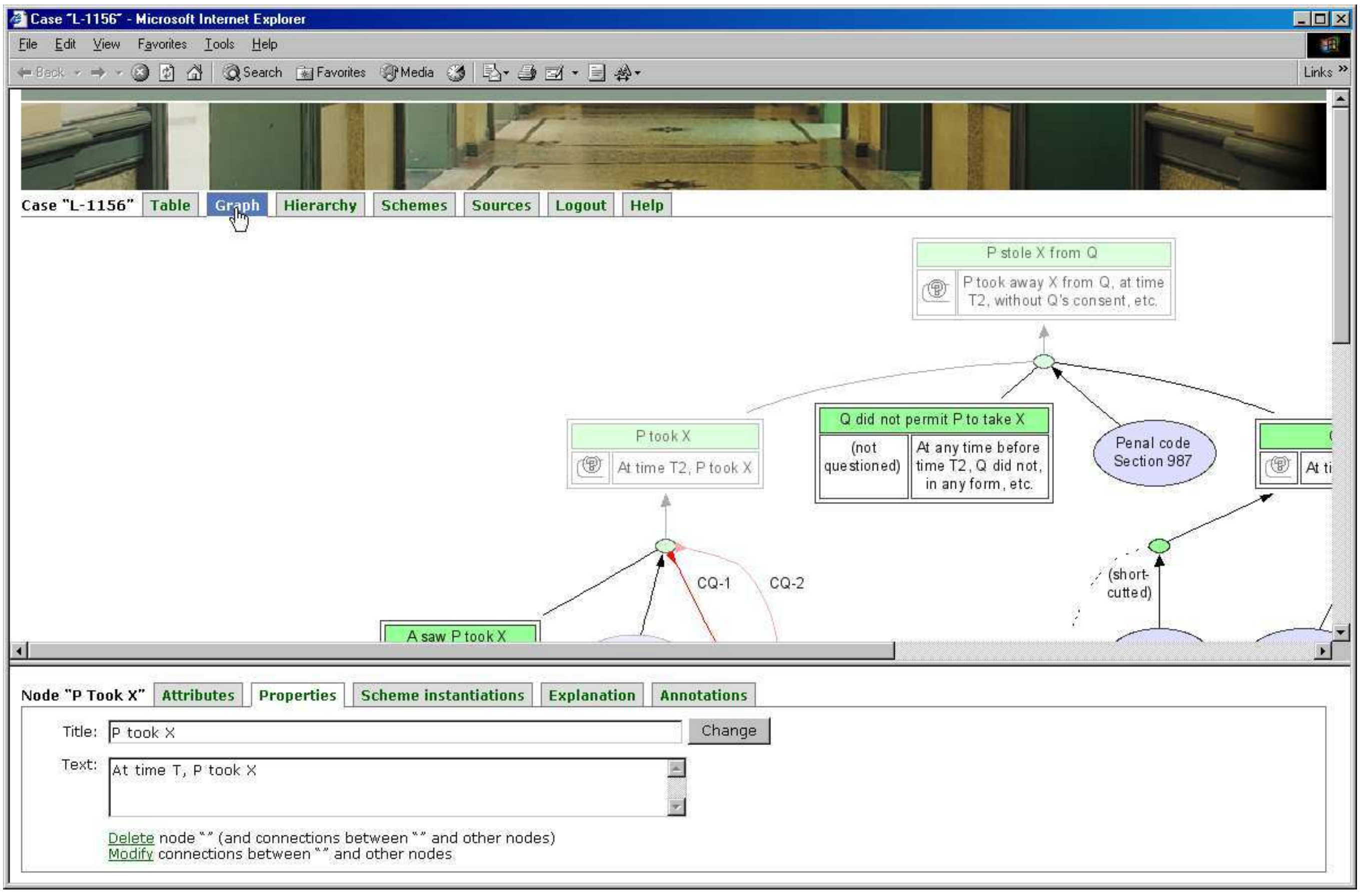
The case can be visually represented through multiple views. These views include a graphical view, a table view, a hierarchical view, a report view, a summary view, and a linear view. The report view is a verbal and linear dump of the case representation and can be used as an official print-out for off-line instantiations (think of the need for paper files and communication by traditional mail).*Stevie* draws heavily on ideas from visualizing argumentation [6, 11]. Therefore, the graphical view is considered to be most representative for on-line uses of *Stevie*.

If a node is clicked in the upper half of the screen, its contents (and some of its other attributes) can be edited in the lower half of the screen. Nodes can be created in isolation (bottom-up) or hierarchically through other nodes (top-down). Thus, a case is built.

2.3 State of implementation

Stevie is prototyped in *Aafje*. *Aafje* is programmed in PHP and stores case data in a PostgreSQL database. *Aafje* has the following functionality: creation of cases, support of multiple users, linkage to quotes in PDF documents, usage of schemes, creation of nodes top-down (from the main claim), bottom-up (from evidence), and by scheme instantiation. Unimplemented features include a properly working labeling algorithm for stories and a faithful incorporation of the AIF ontology (to be explained below).

[1] Department of Information and Computing Sciences, Utrecht University, the Netherlands



Title: P took X

Text: At time T, P took X

Delete node "" (and connections between "" and other nodes)
Modify connections between "" and other nodes

Figure 1. Screenshot of system interface.

3 THEORY

Stevie's conceptual framework is to a large extent based on the core ontology for argument entities and relations between argument entities as described in a recent document on argument interchange formats (AIFs) and ontologies [18]. According to the AIF ontology, knowledge about a (not necessarily legal) case is stored in two kinds of nodes, viz. *information nodes* (I-nodes) and *scheme instantiation nodes* (S-nodes). I-nodes relate to content and represent claims that depend on the domain of discourse. In Fig. 2, I-nodes are rectangles (and conversely).

	Green	*Red*	*Blue*
Rectangle (I-node)	P-node	P-node	Q-node
Ellipse	S-node	S-node	Scheme node

Figure 3. Node visualization.

Schemes

According to the AIF standard, I-nodes may be connected to indicate inferential support, and S-nodes represent justifications for those connections. S-nodes (small red or green ellipses in Fig. 2) are instantiations of general inference schemes (large blue ellipses) and are called scheme instantiation nodes (or instantiation nodes for short). Table 3 summarizes node visualization. Schemes are pre-defined patterns of reasoning, such as rules of inference in deductive logics but then broadened to non-deductive logics or domain dependant patterns such as witness testimony in evidential reasoning [11, 3, 12].

In principle schemes are predefined and may be reused by case analysts. There are many schemes and our system cannot contain them all. Currently, *Stevie* uses the scheme list of Araucaria [10] which to our knowledge is the first system that deals with schemes.

Stories

According to Wagenaar *et al.*'s theory of anchored narratives [16], a story is a credible, coherent, temporally consistent, and defensible set of claims that together describe a possible course of events of a case that is subject to investigation.

Stevie uses defeasible reasoning [4, 9] to distill stories out of large quantities of information. If we use principles of defeasible reasoning to define stories, we may say that stories must be contained in conflict-free and self-defending collection of claims (I-nodes). A set of claims is conflict-free if (and only) if it does not contain a conflicting pair of I-nodes. The meaning of conflict-freeness is further defined in the subsection on stories (p. 4). A set of claims is self-defending if (and only if) every argument (made up of I-nodes and S-nodes) against an element of that story can be countered with an argument made up of I-nodes that belong to that story. In addition to defeasible reasoning principles we add a third constraint on stories, namely that they must be temporally consistent. What this means is defined below. A simple example of a case representation that contains valid stories is shown in Fig. 4.

4 STRUCTURE

The most important elements of *Stevie* are nodes and links between nodes.

Nodes

The basic building block of *Stevie* is a node. A node is an elementary piece of information that is used in modeling cases. Nodes can be facts in a case or claims about a case and are typically displayed in a GUI. Every node possesses two mandatory attributes, viz. a title field and a text field. Additionally nodes possess optional (scalar) attributes such as slots indicating time and location, the name of the analyst who created the node, and a list of records of all edits. Finally, a node can refer to zero or more real-world objects, such as persons, institutions, locations and cars.

I-nodes fall apart into two categories, namely, quotation nodes (Q-nodes, colored blue) and interpretation nodes (P-nodes, colored green and red, depending on the party of interest, cf. Fig. 2).

Quotation nodes

A quotation node represents information from outside the system, such as quotes from testimonies, reports, minutes and other original source documents, but also plain data such as car registration details, addresses, and telephone numbers. The text field of a quotation node is a literal transcription of the selected fragment and cannot be further edited. Once imported, the content of a quotation node is fix ed, and its status is incontestable within the system.

There are two types of quotation nodes: information nodes and scheme quotation nodes (scheme nodes, for short). Information quotation nodes (blue rectangles) are ordinary quotations from external source documents. Scheme nodes (blue ellipses) represent a special type of external information, namely, (quoted) argumentation schemes.

Interpretation nodes

A P-node represents an observation or claim made by a user for the purpose of making sense out of quoted data. Nodes that (indirectly) support the main thesis are colored green; nodes that (indirectly) contest the main thesis are colored red, and nodes that may serve both interests are colored yellow. In the present example, yellow nodes do not occur but they may occur in more complicated cases.

Interpretation nodes can be questioned by users and can be supported by other nodes. Unquestioned interpretation nodes provide support of themselves. Questioned interpretation nodes (indicated by the blue question mark on the left) need further support from other nodes in order to be “believed” or “IN”(the evaluation of nodes is described below). Whether this support indeed exists depends on further input of case analysts.

Thus, an I-node may contain a quote from a source document (Q-node), or it may contain an explanation or interpretation of such a quote (P-node).

Schemes

Schemes belong to a special group of nodes that represent predefined patterns of reasoning. A single scheme describes an inference, the necessary prerequisites for that inference, and possible critical questions that might undercut the inference. A scheme may be instantiated to one or more scheme instances (S-nodes). Graphically, an S-node is depicted as a small ellipse that is red or green depending on the side of interest. Every S-node springs from a scheme node (blue ellipse) and uses zero or more antecedent nodes to justify a consequent node (cf. Fig. 2).

As an example of how schemes may be applied, consider Fig. 2. If a case analyst wishes to support the claim that “P stole X from Q”, *Stevie* will present one or more inference schemes from which this conclusion follows. In this case, the analyst chose the scheme entitled “Penalcode Section 987”. According to this scheme, in order to prove “P stole X from Q”, it is necessary to prove three subclaims, viz. “Q owns X”, “Q did not permit P to take X”, and “P took X”. In this case, these three claims suffice to conclude that “P stole X from Q”.

Schemes can also be instantiated the other way around, from quotation (or interpretation) nodes to conclusion nodes. Consider again Fig. 2. If an analyst wants to find out which conclusion follows from the testimonial evidence “A: “I saw P took X””, he may chose the “Quote instantiation” scheme and will be automatically presented with the conclusion that follows being “A said: “I saw P took X””.

Most schemes incorporate a pre-defined list of so-called *critical questions*. A critical question is a possible circumstance that may invalidate a particular scheme instantiation [11, 12]. Thus, critical questions are latent rebutters of S-nodes or, put differently, latent undercutters. Fig. 2 shows examples of critical questions for some schemes. For instance, the inference from “A saw P took X”” to “P took X” through “Perception” may be rebutted by the knowledge that A is short-sighted and did not wear glasses.

Links

To create a network of inferential and temporal interdependencies, nodes can be linked through two types of connections, that is, inferential connections (arrows and arrows with reversed arrowheads in Fig. 4) and temporal connections (arrows with solid dots as arrowheads).

Inferential links

Inferential connections can be created by instantiating schemes. Thus, although inference links and S-nodes look different, they are actually the same. Supporting connections are displayed by arrows, attacking connections by reversed arrowheads.

Temporal links

Temporal connections are made when two nodes possess sufficient information to relate them temporally, or else when a case analyst decides that two nodes must be connected temporally. Once temporal connections exist it is possible to represent stories of what might have happened as a sequence of temporally structured nodes.

Two nodes receive a temporal connection automatically if they both possess an explicit time stamp. Nodes can be connected manually as well. If a case analyst decides that node A precedes node B in time, he creates a temporal link between A and B. In doing so, the case analyst must qualify that link by indicating his own confidence in that link. This qualification can be selected from a predefined set of modalities (for example: "certainly," "beyond a reasonable doubt," and "likely").

Stories

The objective of *Stevie* is to create, on the basis of quotes and interpretations, possible stories that indicate what might have happened. In *Stevie*, a story is a set S of nodes that satisfies the following two postulates:

1. S is conflict-free and self-defending.
2. The underlying temporal digraph T of S is internally consistent (i.e., acyclic) and consistent with temporal and causal orderings implied by scheme instantiation nodes.

Thus, S must be conflict-free, self-defending, and temporally consistent. Since all information available in a case together is almost always inconsistent, it is usually the case that a single case yields room for multiple stories. Based on inferential connections, nodes can be evaluated as being "IN" or "OUT". Quotation nodes and unquestioned interpretation nodes are "IN".

There exist several semantics for node evaluation. *Stevie* uses the grounded and the admissibility semantics, respectively [5, 9]. For the sake of simplicity, only the admissibility semantics is briefly quoted here [5]. This semantics enforces the two properties that are mentioned under (1) above.

Nodes can be either "IN," "OUT," or "UNDEC" (undecided).

1. A questioned interpretation node N is "IN", if it satisfies the following two conditions.
   (a) N is supported by an S-node that is "IN"
   (b) All S-nodes that attack N are "OUT"
2. A questioned interpretation node N is "OUT", if it satisfies one of the two following conditions.
   (a) All S-nodes that support N are "OUT"
   (b) N is attacked by an S-node that is "IN"
3. A questioned interpretation node N is "UNDEC", otherwise.

More complex configurations possess more than one valid labeling, and in some configurations the empty story (all nodes "UNDEC") is also a valid labeling. When instantiating a scheme, newly created antecedent elements cannot have been questioned yet so that they are "IN", until either the corresponding S-node or else one of its antecedent nodes is either questioned or attacked. In Fig. 2 the node "Q sold X to P" is out since it is undercutted by "P is a party concerned". As a result, the node "Q owns X" is "IN", because its rebutter is "OUT".

A detailed description of the algorithms used for graph "consistency checking" (as it is called by one of the reviewers) is beyond the scope of this paper. More detailed descriptions a various such algorithms can be found in the formal argumentation literature [4, 9].

5 RELATED WORK

As remarked in Sec. 2.2, *Stevie* draws heavily on ideas from visualizing argumentation. Compared to traditional issue-based information systems (IBISs) and argument visualization tools, however, *Stevie* is more directed towards the construction of stories than to visualization as a goal in itself. Further, *Stevie* uses a node ontology that is in line with the current standards on representation formats for argument interchange (AIF).

Because of its graphic interface, *Stevie* is strongly connected to FLINTS [7, 8, 19]. FLINTS (Forensic Led Intelligence System) is a methodology and software system that helps analysts to identify relevant information in large amounts of data. The difference between FLINTS and *Stevie* other than that FLINTS is a much more matured system, is that FLINTS is not centered around the construction of stories as *Stevie* is.

With respect to the data model, *Stevie* follows the same approach as case analysis tools such as Araucaria [11] and Legal Apprentice [17]. Araucaria is a software tool for the analysis and visualization of arguments. It supports argumentation schemes, and depicts arguments as trees of nodes, where nodes consists of quotes from a fixed text that is displayed in the left margin. Legal Apprentice (LA) is a case analysis system that visualizes evidence in so-called legal implication trees. Those are AND/OR tree-structures where nodes can receive a true, false or undefined status from case analysts. The main conceptual differences between *Stevie* and these systems is that *Stevie* uses a logic and ontology of which basic principles such as scheme instantiation [11, 3, 12] and admissibility [5] have a solid theoretical underpinning in the theory of formal argumentation [4, 9, 18].

With respect to argumentation and legal narratives, *Stevie* is also strongly connected to MarshalPlan [13], a formal tool to prepare legal cases for trial. The main difference between *Stevie* and MarshalPlan is that *Stevie* is more directed towards investigation than towards the preparation of legal trials.

Particularly relevant to mention is DAEDALUS [2], a tool that may help Italian magistrates and prosecutors in their work; it is not, like *Stevie* graphically oriented but its usefulness resides in the facility that it may be requested to validate and document steps made by the magistrate and the police.

A last approach that is interesting to mention is the coherentist approach as advocated by Thagard *et al.* such as ECHO [14, 15] and especially ConvinceMe [1]. The latter is an artificial pedagogical assistant to help students structure, restructure, and assess their knowledge about often controversial situations. Like *Stevie* it is a sense-making tool to formulate hypotheses based on evidence, but then based on principles of coherence rather than being based on principles of argument.

Acknowledgement. The authors like to thank Floris Bex, Henry Prakken and the anonymous reviewers for their constructive help. They also gratefully acknowledge the support from the Netherlands Organisation for Scientific Research (NWO) through project 634.000.429 and from the EU through project IST-FP6-002307 (ASPIC).

REFERENCES

[1] Stephen T. Adams, ‘Investigation of the ”Convince Me”computer environment as a tool for critical argumentation about public policy issues’, *Journal of Interactive Learning Reseach*, **14**(3), 263–283, (2003).

[2] Carmelo Àsaro, Ephraim Nissan, and Antonio A. Martino, ‘Daedalus: An integrated tool for the italian investigating magistrate and the prosecutor. a sample session: Investigating an extortion case’, *Computing and Informatics*, **20**(6), 515–554, (2001).

[3] Floris J. Bex, Henry Prakken, Chris A. Reed, and Douglas N. Walton, ‘Towards a formal account of reasoning about evidence: argumentation schemes and generalisations’, *Artificial Intelligence and Law*, **11**, 125–165, (2003).

[4] Carlos I. Chesñevar, Ana G. Maguitman, and Ronald P. Loui, ‘Logical models of argument’, *ACM Comput. Surv.*, **32**(4), 337–383, (2000).

[5] Phan Minh Dung, ‘On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games’, *Artificial Intelligence*, **77**(2), 321–357, (1995).

[6] *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, eds., Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr, Springer-Verlag, London, 2002.

[7] Richard M. Leary, *Evidential Reasoning and Analytical techniques in Criminal Pre-Trial Fact Investigation*, Ph.D. dissertation, University College, London, 2004.

[8] Richard M. Leary, John Zeleznikow, and Wim VanDenBerghe, ‘User requirements for financial fraud modeling’, in *Proceedings of BILETA 2003: British and Irish Law, Education and Technology Association 18th Annual Conference*, (2003).

[9] Henry Prakken and Gerard A.W. Vreeswijk, ‘Logical systems for defeasible argumentation’, in *Handbook of Philosophical Logic*, volume 4, 218–319, Kluwer, 2nd edn., (2002).

[10] Chris A. Reed and Glenn W.A. Rowe, ‘Araucaria: Software for argument analysis, diagramming and representation’, *Int. Journal of AI Tools*, **14**(3-4), 961–980, (2004).

[11] Chris A. Reed and Douglas N. Walton, ‘Applications of argumentation schemes’, in *Proceedings of the 4th OSSA Conference*, Ontario, Canada, (2001).

[12] Chris A. Reed and Douglas N. Walton, ‘Towards a formal and implemented model of argumentation schemes in agent communication’, *Autonomous Agents and Multi-Agent Systems*, **11**(2), 173–188, (2005).

[13] David Schum, ‘Evidence marshaling for imaginative fact investigation’, *Artificial Intelligence and Law*, **9**(2/3), 165–188, (2001).

[14] Paul Thagard, *Coherence in Thought and Action*, MIT Press, Cambridge, MA, 2000.

[15] Paul Thagard, ‘Causal inference in legal decision making: Explanatory coherence vs. bayesian networks’, *Applied Artificial Intelligence*, **18**(3/4), 231–249, (2004).

[16] Willem A. Wagenaar, Hans F.M. Crombag, and Peter J. van Koppen, *Anchored Narratives: Psychology of Proof in Criminal Law*, St Martin’s Press / Prentice-Hall, 1993.

[17] Vern R. Walker. Early legal apprentice software (published as screencast). `http://people.hofstra.edu/faculty/vern_r_walker/LegalReasoning.html`, 2006.

[18] Steven Willmott, Gerard Vreeswijk, Matthew South, Carlos Chesñevar, Guillermo Simari, Jarred McGinis, and Iyad Rahwan, ‘Towards an argument interchange format for multiagent systems’, in *Proc. of the 3rd Int. Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2006)*, (2006).

[19] John Zeleznikow, Giles Oatley, and Richard M. Leary, ‘A methodology for constructing decision support systems for crime detection’, in *Proc. of the Ninth Int. Conf. on Knowledge-Based and Intelligent Information and Engineering Systems (KES)*, pp. 823–829, (2005).

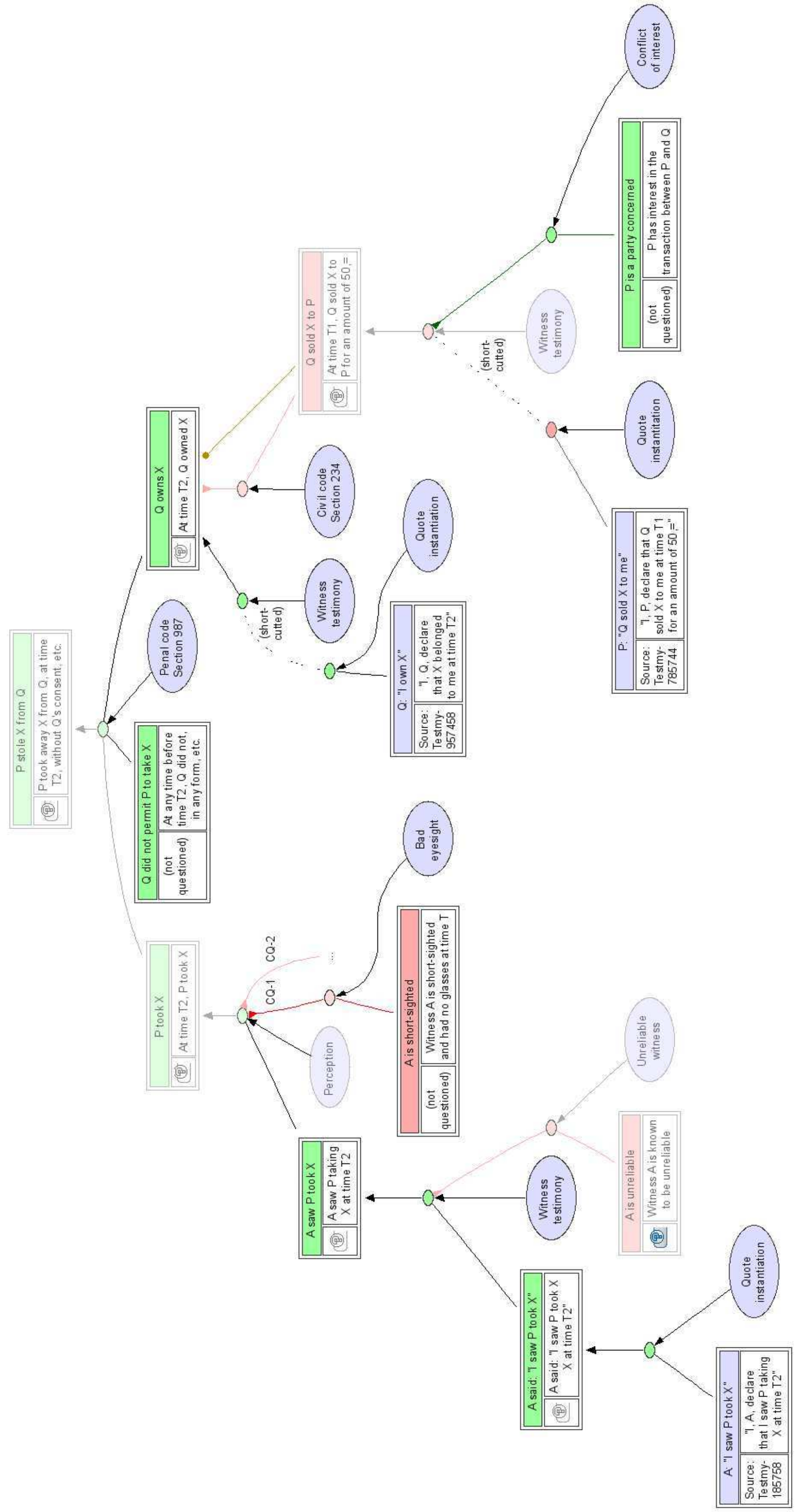


Figure 2. Graph view of theft case.

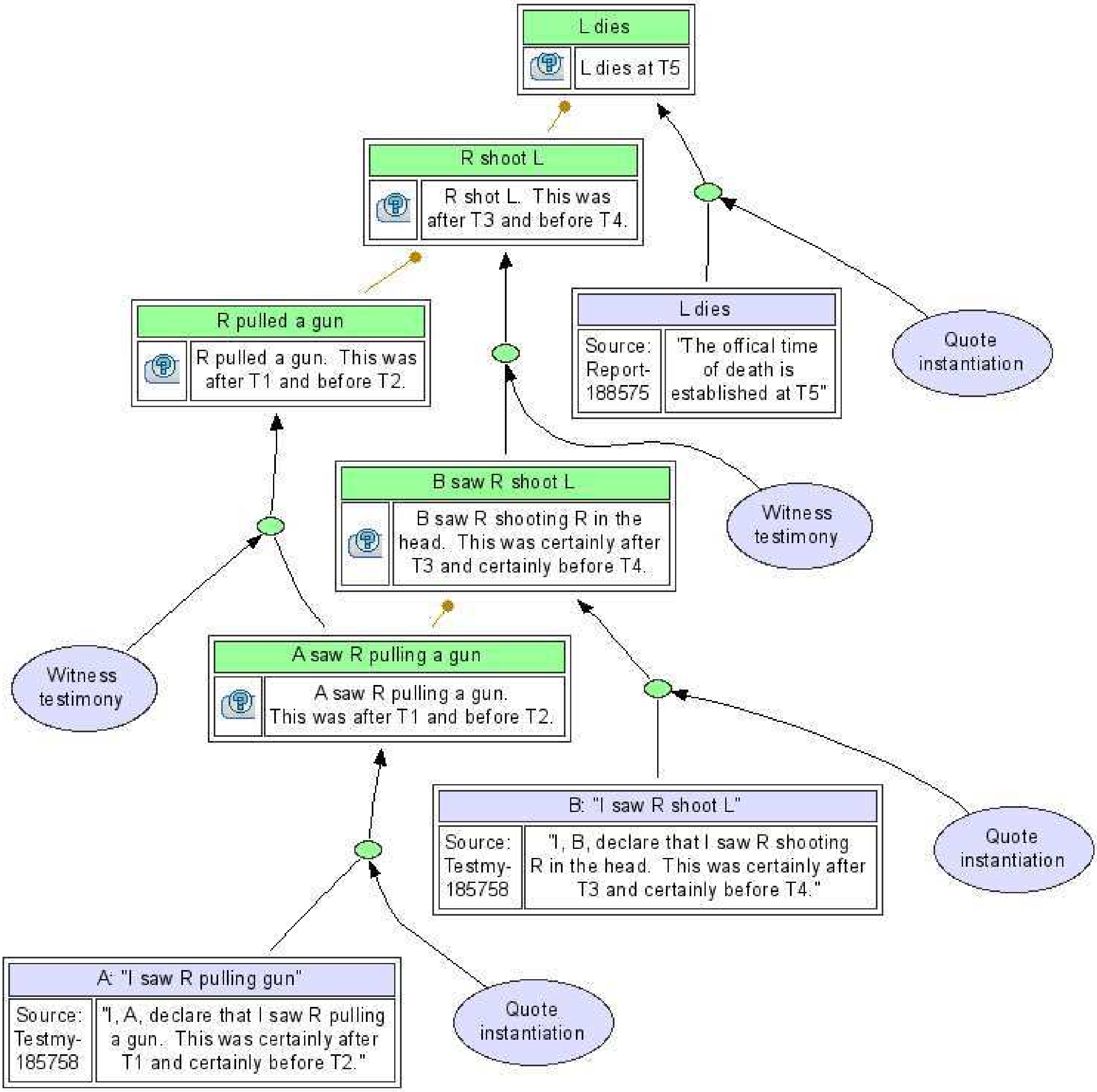

L dies
L dies at T5
R shoot L
R shot L. This was after T3 and before T4.
R pulled a gun
R pulled a gun. This was after T1 and before T2.
L dies
Source: Report-188575
"The offical time of death is established at T5"
Quote instantiation
B saw R shoot L
B saw R shooting R in the head. This was certainly after T3 and certainly before T4.
Witness testimony
Witness testimony
A saw R pulling a gun
A saw R pulling a gun. This was after T1 and before T2.
B: "I saw R shoot L"
Source: Testmy-185758
"I, B, declare that I saw R shooting R in the head. This was certainly after T3 and certainly before T4."
Quote instantiation
A: "I saw R pulling gun"
Source: Testmy-185758
"I, A, declare that I saw R pulling a gun. This was certainly after T1 and certainly before T2."
Quote instantiation

Figure 4. Graph view of shooting case.

Knowing when to bargain: The roles of negotiation and persuasion in dialogue

Simon Wells and Chris Reed

Abstract. In this paper two formal dialectic systems are described, a persuasion protocol (PP_0) and a negotiation protocol (NP_0), together with a method for shifting from an instance of a persuasion dialogue to an instance of a negotiation dialogue. The rationale for this kind of shift is explored in the context of the fallacy of bargaining. Such a dialectical shift is proposed as a valuable way to enable the participants in an argumentative dialogue to proceed towards a practical settlement when they are otherwise unable to persuade each other and thereby bring about a resolution of their conflicts.

1 Introduction

A typical situation in argumentative dialogue occurs when one party attempts to persuade another party to accept some standpoint. This involves notions of attack and defence as the parties attempt to justify their own position whilst refuting that of their opponent. However, because the participants are autonomous entities they will each evaluate the proffered arguments on their own terms. An argument that party A believes is sufficient to persuade party B isn't necessarily the same argument that B would accept and would thus be persuaded. What should occur when A cannot persuade B? If getting B to accept the standpoint is important to A, then A should have available an alternative tactic for reaching agreement in those situations where a sufficiently persuasive argument cannot be brought to bear.

In real-world argument many people resort to bargaining when they are unable to persuade their opponent. For example, Harry and Sally are arguing about who should do the washing up. Both have stated that they will not do the washing up and that the other should do it. Sally tries to persuade Harry to do the washing up and defends her position, when it is inevitably attacked, by stating that she always does the washing up and asks why Harry can't do it for a change. Harry justifies his refusal to do the washing up with the defence that he has just hoovered the living room and so he shouldn't have to do both jobs. Domestic conflictssuch as this are a common occurrence that are often resolved when an offer is made, for example, Harry concedes he will do the washing up if Sally will take the rubbish out. This is not a concession based upon Sally's superior persuasive argument but based upon a wider view of the situation and the need to reach a practical settlement. The fact that the rubbish needed to be taken out was not an issue that was raised in the preceding persuasion dialogue but was an issue that could be raised during a negotiation dialogue.

As demonstrated in the domestic strife example, when a party cannot get their standpoint accepted through justification of that standpoint an alternative tactic is to enter into some sort of negotiation over the issue to determine;

1. what it would take to get the standpoint accepted by the other party, and, failing that,
2. to determine what alternative (possibly reduced) standpoint B might accept if it turns out that the original standpoint is unlikely ever to be acceptable.

This kind of situation can be characterised as the movement within a dialogue from a persuasion-type sub-dialogue to a negotiation-type sub-dialogue. This paper introduces two formal dialectic systems named Persuasion Protocol $_0$ (PP_0) and Negotiation Protocol $_0$ (NP_0), together with a method for moving from a persuasion sub-dialogue carried out in accordance with PP_0 to a negotiation sub-dialogue carried out in accordance with NP_0. The aim is to demonstrate that this particular shift, from persuasion to negotiation, can be a useful way to proceed when a persuasion dialogue is unlikely to reach a stable agreement. These results can then be applied to computational models of argument such as those for use in multiagent systems. Agents may have many more capabilities than those that are relevant to the current persuasion dialogue. If $agent_1$ cannot persuade $agent_2$ then $agent_1$ may use the opportunity to shift to a negotiation dialogue in which a concession might be won.

2 Background

This paper deals with a number of topics in argumentation including the use of formal dialectic systems to model the interactions between participants in an argumentative dialogue, the recognition that dialogues conform to a number of distinct types, and that given a formal dialectic system which models the interactions in a particular type of dialogue, there will arise the need to shift from a dialogue of one type to a dialogue of another type, and hence transition from one dialectic system to another.

Formal Dialectic Systems Dialogue games have been proposed as a means to model the interactions between participants during argumentative dialogues. One branch of dialogue game research is into the formal dialectic system [5]. Formal dialectic systems are two-player, turn-taking games in which the moves available to the players represent the locutional acts or utterances made by the participants of a dialogue. Many dialectic systems have been proposed based on the characterisations of a range of dialogical situations, for example, Hamblin's system [5] and Mackenzie's DC [6] are targeted towards fallacy research whilst Walton and Krabbe's system PPD_0 [15] models the interactions between parties in a permissive persuasion dialogue. Girle introduces a number of systems which are aimed at modelling belief revision in A.I. systems [2, 3, 4]. McBurney and Parsons specify some games for use in communication between agents in multiagent systems [8]. Bench-Capon *et al.* introduce a system for modelling dialectical argument called the Toulmin Dialogue Game [1] that is based upon the argument schema of Toulmin [12].

Dialogue Typologies Dialogue can categorised into types and are distinguished based upon a range of characteristics such as initial situation, the overall goal and the participant's individual aims. An influential but partial typology of such dialogue types which includes information-seeking, persuasion, negotiation, deliberation, and inquiry can be found in [15]. This paper is concerned primarily with the negotiation and persuasion types of dialogue although the findings can be extended to incorporate the other dialogue types identified by Walton and Krabbe.

Negotiation Dialogues In multiagent systems research, negotiation is often characterised as a means to distribute limited resources between competing agents. Negotiation dialogues can be used to determine the distribution of those resources between the conflicting parties. In the Walton and Krabbe typology negotiation dialogues are characterised by a conflict of interests and a need for cooperation leading to a practical settlement.

Persuasion Dialogue Persuasion dialogues occur when there is a conflict and the participants attempt to reach a stable agreement or resolution of the issue that gave rise to the conflict. Walton and Krabbe specify a formal dialectic system to model the interactions during persuasion dialogues name PPD_0.

Progression Between Dialogue-types The notion of embedding an instance of one type of dialogue within an instance of another type of dialogue was proposed in [15] and various other approaches have been proposed including Reed's Dialogue Frames [10], and the layer model of McBurney and Parsons [7]. The core idea is to enable the participants in a dialogue to move from a sub-dialogue of one type to a sub-dialogue of another type where each sub-dialogue has its own specification of rules governing how a dialogue of that type should progress. The notion of embedding persuasion sub-dialogues within an ongoing negotiation dialogue has been explored quite extensively by Sycara in relation to the PERSUADER system [11], and by Rahwan [9] in relation to argument-bsaed negotiation in multiagent systems. However the converse situation of embedding negotiation sub-dialogues within a persuasion dialogue has not been explored specifically except as a by-product of enabling embeddings and shifts in general.

3 The fallacy of Bargaining

Walton and Krabbe identify in [15] that shifts from one type of dialogue to another may be either licit or illicit. A licit shift occurs when the shift is constructive and agreed to by all parties. When a shift is concealed or otherwise inappropriate then it is illicit. Walton argues that a characteristic of many fallacies is that they occur where shifts in the dialogue are illicit [14]. In [15] the fallacy of bargaining is identified as occuring when participants are engaged in a dialogue which starts out as a persuasion but that at some point during the course of the dialogue an illicit shift occurs from persuasion to negotiation.

The example of the fallacy of bargaining used by Walton and Krabbe involves a government minister of finance who has been caught profiting from certain tax exemptions. The minister argues that those tax exemptions should be allowed temporarily and not be penalized. The minister then goes on to propose to his critics that if they abstain from moving for penalties for the exemptions, then he will not oppose a bill that the critics will benefit from. In this case, instead of satisfying his burden of proof with respect to his position on the tax exemptions, the minister substitutes an offer for an argument, a move which is not permissible in persuasion dialogues. By making an offer during the persuasion dialogue the minister has reneged on his commitment to defend his position, *vis a vis* the tax exemptions, and caused an illicit shift to a negotiation dialogue.

However, the shift from persuasion to negotiation need not always be an instance of the fallacy of bargaining. As Walton and Krabbe recognise, illicit shifts occur when the shift is concealed or inappropriate and a fallacy can occur as a result, If the shift occurs in an open way, and is demonstrated to be appropriate then there is no need to characterise it as fallacious. Where conflicting participants in a dialogue have exhausted their persuasive arguments and are in a position that is unlikely to be resolved through continuation of the persuasion dialogue then it is acceptable for the participants to try some other way to break the deadlock. In this case, the persuasion dialogue has failed because a stable agreement has not been reached. Given that both participants actually wish to resolve the conflict, which is the reason why they are still engaged in the dialogue at this point, a shift to another type of dialogue enables the participants to continue. If the shift is from a persuasion dialogue to a negotiation dialogue then the participants may be able to reach a practical settlement and so be able to move forward.

The dialogue protocols presented in this paper together with the associated machinery to effect dialogue shifts are aimed at demonstrating two points. Firstly that not all shifts from persuasion to negotiation dialogues need be instances of the fallacy of bargaining, and secondly that these kinds of shifts can be utilised to enable participants who would otherwise have reached an impasse to continue.

4 The systems: PP_0 and NP_0

The two formal dialectic systems, PP_0 and NP_0 are represented using the unified specification format introduced in [16]. This representation is part of a unified framework for representing, rapidly implementing and deploying formal dialectic systems called the Architecture for Argumentation (A4A). To facilitate this, the framework incorporates a range of general machinery for representing dialectic systems. This machinery is then tailored to the needs of a specific dialectic system. The dialectic system itself is designed to model the interactions between participants during a particular dialogical situation. In this case PP_0 is formulated to model persuasion dialogues and NP_0 is formulated to model negotiation dialogues.

The reason for the A4A representation is twofold; to simplify and unify the representation of formal dialectic systems and to enable the construction of a common engine for running those systems so represented. The traditional layout of formal dialectic involves specifying a number of groups of rules that govern a range of capabilities of the system such as commitment store updates and legal sequences of moves. This approach is adequate but can obscure comprehension of which moves are legal at any given point in a dialogue and the exact effect of playing any of those moves. The A4A approach specifies the range of rules which can be used to layout a dialectic system. These rules are grouped together to facilitate understanding and transparency of the overall system. The gross structure of an A4A layout involves specification of the type and capabilities of a number of basic components, followed by a prescription of global rules. Finally a collection of moves is laid out. Basic components include a unique identifier for the system, a turn-structure, identifiers for the participants and the setting up of stores for any artifacts created during the dialogue. Global rules are used to identify a range of conditions that can arise during a dialogue and specify what should be done when those conditions arise. In the case of PP_0 and NP_0 these include rules that hold when a new dialogue is entered, rules that govern transitions between sub-dialogues, e.g. from a PP_0 sub-dialogue to an NP_0 sub-dialogue, and rules that specify when a dialogue should terminate. The rules that concern individual moves are grouped together so that it is immediately apparent when the move can legally be played and what the effect of playing that move is.

PP_0 is a protocol tailored towards persuasion-type dialogues.

System Name PP_0

Turn Structure = ⟨Determinative, Single-Move⟩

Participants = {init, resp}

Artifact Stores :
⟨CStore, init, Mixed, Set, Light, Global⟩
⟨CStore, resp, Mixed, Set, Light, Global⟩

Global Rules :

Initiation
Requirements:
$\mathrm{T}_{current} = 0$
Effects:
$\mathrm{T}^{init}_{next\_move} = \langle\mathrm{Request, (goal)}\rangle$

Progression
Requirements:
$\mathrm{S} \in \mathrm{CStore}^{init}_{1} \wedge \mathrm{S} \in \mathrm{CStore}^{init}_{current} \wedge$
$\mathrm{T}^{resp}_{last} = \langle\ \mathrm{Reject, (S)}\ \rangle$
Effects:
$(\mathrm{System}=NP_0) \vee (\mathrm{System}=PP_0)$

Termination
Requirements:
$\mathrm{S} \in \mathrm{CStore}^{init}_{1} \wedge (\mathrm{S} \notin \mathrm{CStore}^{init}_{current} \vee$
$\mathrm{S} \in \mathrm{CStore}^{resp}_{current}) \vee$
$\mathrm{T}_{last\_move} = \langle\mathrm{Withdraw(–)}\rangle$
Effects:
$\mathrm{Dialogue}_{status}$ = complete

Moves :

⟨**Request, (S)**⟩
Requirements:
Ø
Effects:
$\mathrm{T}^{listener}_{next\_move} = \langle\ \mathrm{Accept, (S)}\ \rangle \vee \langle\ \mathrm{Reject, (S)}\ \rangle \vee$
$\langle\ \mathrm{Challenge, (S)}\ \rangle \wedge$
$\mathrm{CStore}^{speaker}_{current} + \mathrm{S}$

⟨**Accept, (S)**⟩
Requirements:
$\mathrm{T}^{listener}_{last\_move} = \langle\ \mathrm{Request, (S)}\ \rangle$
Effects:
$\mathrm{CStore}^{speaker}_{current} + \mathrm{S} \wedge \mathrm{CStore}^{speaker}_{current} - \neg\mathrm{S}$

⟨**Reject, (S)**⟩
Requirements:
$\mathrm{T}^{listener}_{last\_move} = \langle\ \mathrm{Request, (S)}\ \rangle$
Effects:
$\mathrm{T}^{listener}_{next\_move} = \langle\ \mathrm{Challenge, (S)}\ \rangle \vee \langle\ \mathrm{Withdraw, (–)}\ \rangle \wedge$
$\mathrm{CStore}^{speaker}_{current} + \neg\mathrm{S} \wedge \mathrm{CStore}^{speaker}_{current} - \mathrm{S}$

⟨**Challenge, (S)**⟩
Requirements:
$\mathrm{T}^{listener}_{last\_move} = \langle\ \mathrm{Request, (S)}\ \rangle \vee \langle\ \mathrm{Reject, (S)}\ \rangle \vee$
$\langle\ \mathrm{defence, (S'{\rightarrow}S)}\ \rangle$
Effects:
$\mathrm{T}^{listener}_{next\_move} = \langle\ \mathrm{defence, (S'{\rightarrow}S)}\ \rangle \vee \langle\ \mathrm{Reject, (S)}\ \rangle \vee$
$\langle\ \mathrm{Withdraw, (–)}\ \rangle$

⟨**defence, (S′→S)**⟩
Requirements:
Ø
Effects:
$\mathrm{T}^{listener}_{next\_move} = \langle\mathrm{Challenge, (S)}\rangle \vee \langle\mathrm{Challenge, (S')}\rangle \vee$
$\langle\mathrm{Challenge, (S'{\rightarrow}S)}\rangle \vee \langle\mathrm{reject, (S'{\rightarrow}S)}\rangle \vee \langle\mathrm{reject, (S)}\rangle \vee$
$\langle\mathrm{reject, (S')}\rangle \vee \langle\mathrm{accept, (S'{\rightarrow}S)}\rangle \vee \langle\mathrm{accept, (S)}\rangle \vee$
$\langle\mathrm{accept, (S')}\rangle$
$\mathrm{CStore}^{speaker}_{current} + \mathrm{S} \wedge \mathrm{CStore}^{speaker}_{current} + \mathrm{S'} \wedge$
$\mathrm{CStore}^{speaker}_{current} + \mathrm{S'{\rightarrow}S}$

⟨**Withdraw, (–)**⟩
Requirements:
$\mathrm{T}_{last\_move} = \langle\mathrm{Challenge(S)}\rangle \vee \langle\mathrm{Reject(S)}\rangle$
Effects:
Ø

PP_0 enables two players named *init* and *resp* to engage in a persuasion dialogue. Players can make one move per turn, starting with init. The turn structure means that turns procede automatically, after one player makes their move, the next player has their turn and so on, such that it can be seen from examination of the current turn index which players move it is. The actual moves that are played cannot influence which player is assigned the speaker role in the next turn and thus cannot influence whose turn it is. Each player is assigned an artifact store named CStore. The remaining parameters specify that the store can contain a mixture of commitment types, for example a player can incur commitment to just the content of a move or to the entire move, that the store is a light side store [13] which stores a set of commitments and that the stores are to be shared between sub-dialogues of differing types. PP_0 incorporates three types of global rule. These rules specify the requirements for starting a new instance of a PP_0 sub-dialogue, the requirements for initiating a progression from an instance of a PP_0 sub-dialogue to a new instance of another sub-dialogue type, and the conditions for terminating a PP_0 dialogue.

When a new sub-dialogue of type PP_0 is begun the initiation rules require only that the very next move, in this case the first move of the new sub-dialogue, must be a request. For a progression to to be legal it is required that the player who initiated the PP_0 instance still be committed to their initial thesis and that the last move played in the immediate previous turn was a rejection of that initial thesis by the respondent. These conditions establish that a progression is legal at this point in the dialogue, and that the next move may be from the set of moves allocated to the NP_0 system. The current player may elect to continue in the current dialogue without progressing to another dialectic system. For example, the progression rules of PP_0 only establish that a transition is legal, not that it must occur. To actually initiate a progression at this point requires the player to make a legal move from the NP_0 move set according to the initiation rules for NP_0.

It should be noted that the particular formulation of progression rules in PP_0 could be folded into the effects of the reject move but that in the wider context of the A4A this approach increases the flexibility of the overall system. This flexibility allows systems to be created in which the conditions for a legal progression between sub-dialogues can occur based on the state of the system's components regardless of the actual move which has just been played.

It is important that a computational model of argument include a clear formulation for when the system should terminate. This helps avoid the implementational problems that can occur when adopting a dialectic system which has no formulation for termination rules.In these case the implementors must add rules to the core system to determine when a dialogue should terminate. This can lead to many variations on the core system. The termination rules of PP_0 require that either the withdraw move has been played, or that the initial thesis of the initiator has either been withdrawn by the initiator or accepted by the respondent.

PP_0 allows six distinct moves. Each move specification incorporates a formulation of requirements for when the move is legal, and a formulation of effects that must be applied when the move is played. The request move is an utterance of the form "Will you S?", and has no requirements. The effects of playing the request move are that the content of the move is added to the speaker's commitment store and that the legal responses are the accept, reject and challenge moves. The accept move enables a player to agree to a request and is of the form "OK S". Conversely the reject move enables a player to disagree with a request and is of the form "Not S". The challenge move is formulated to enable a player to get justification for a previous request, reject or defence move and is of the form "why S?". The defence move enables a player to defend their challenged position by providing a supporting statement of grounds and by stating an inferential link between the challenged position and the justifying statement. The withdraw move is essentially an utterance of the form "I withdraw from this dialogue", and the rationale is to allow either player the opportunity to withdraw from the dialogue. If either player determines that the dialogue is unlikely to end successfully then it is more computationally efficient to leave the dialogue cleanly at the first subsequent opportunity rather than continue.

PP_0 only allows a player to incur commitment on their own behalf. This is achieved through the formulation of effects for each move which only update the commitment store of the speaker. The only moves which incorporate a commitment effect are the request, accept, reject and defence moves. The challenge move does not incorporate a commitment effect, like the commitment to challenges of DC [6], but rather allows the receiver of the challenge to immediately withdraw from the dialogue without penalty. This enables the participants to produce a number of different justifications in response to a challenge by engaging in several iterations of the challenge-defence sequence. This enables some tactical play to emerge in PP_0 persuasion dialogue whereby a player can repeatedly challenge a statement to uncover the underlying justificatio ns for that statement, but if the player is too persistent then their opponent may choose to withdraw from the dialogue entirely. To avoid withdrawal, it is incumbent upon the challenging player to determine when they are unlikely to be able to persuade their opponent and may have more success engaging in a negotiation dialogue instead. As established earlier, the progression rules set out only when it is legal to transition to a new sub-dialogue, not that that transition must occur.

This particular formulation of progression rules does not wholly alleviate the possible charge of a fallacy of bargaining being committed. However some effort is made to avoid that situation. A progression is only legal, at the very earliest, after a request has been made and that request has been rejected outright by the respondent. The respondent could have challenged the request and the initiator would have been obliged to provide a defence to justify their initial request. It may actually be in the interests of the initiator for the persusasion dialogue to continue because, so long as they have some argument to support their position they may be able to persuade the respondent whereas conversely it can be in the interests of the respondent to enter into negotiation to get some concessions from the initiator. It is only in the event that the initiator has no argument to justify their position and must make an offer in lieu of a defence or withdraw from the dialogue, that it is in the initiators interests to move straight to a negotiation dialogue. A stronger formulation of progression rules would require that the initiator had previously provided at least one defence of their initial thesis before a progression could become legal. This would require the progression rules to check that $CStore^{init}$ contains at least one defence of the initial thesis. This would avoid the kind of fallacy of bargaining attributed to the minister of finance in the Walton and Krabbe example discussed earlier because the initiator would have actually provided a defence in support of their request so the initiator is fulfilling the commitment to defend their position rather than resorting immediately to bargaining.

NP_0 is a protocol tailored towards negotiation-type dialogues. PP_0 is aimed at persuading a player to accept a request through successive rounds of challenge and justification. This type of dialogue requires that arguments be brought to bear which hold direct relations to the issue in question. For example, it is assumed that the defence of a challenged request lends at least some support to the request which was challenged in the first place. Likewise, an argument that is extended in defence of a request should provide relevant support for why that request should be accepted. In a negotiation the players may make offers in support of their goal. The offers however need not pertain directly to the goal. Walton and Krabbe recognise in [15] that the swapping of one concession for another is a characteristic of negotiation. In the context of a multiagent system implementation, the agents may have many different capabilities, many of which are not pertinent to the issue at hand but which may be offered as part of a deal in order to get the goal accepted. This kind of dialogue is characterised by offer-counter offer sequences. The rules of NP_0 are as follows;

System Name NP_0

Turn Structure = ⟨Determinative, Single-Move⟩

Participants = {init, resp}

Artifact Stores :
⟨CStore, init, Mixed, Set, Light, Global⟩
⟨CStore, resp, Mixed, Set, Light, Global⟩

Global Rules

Initiation

Requirements:
$S \in CStore_1^{init} \wedge S \in CStore_{current}^{init} \wedge S \notin CStore_{current}^{resp}$

Effects:
$T_{next\_move}^{speaker} = \langle Offer, (S, proposal)\rangle$

Termination

Requirements:
$S \in CStore_1^{init} \wedge (S \notin CStore_{current}^{init} \vee S \in CStore_{current}^{resp}) \vee$
$T_{last\_move} = \langle Withdraw(–)\rangle$

Effects:
$Dialogue_{status} = complete$

Moves

⟨Offer, (goal, proposal)⟩

Requirements:
$\langle Offer, (goal, proposal)\rangle \notin CStore_{current}^{speaker}$

Effects:
$(T_{next\_move}^{listener} = \langle Accept, (proposal)\rangle \vee \langle Reject, (proposal)\rangle \vee$

$\langle$Offer, (goal, proposal$'$)$\rangle \vee \langle$Offer, (goal$'$, proposal)$\rangle \vee$
$\langle$Offer, (goal$'$, proposal$'$)$\rangle \vee \langle$Withdraw, (–)$\rangle$) $\wedge$
CStore^{speaker} + goal $\wedge$
CStore^{speaker} + proposal $\wedge$
CStore^{speaker} + offer(goal, proposal)

$\langle$Accept, (goal, proposal)$\rangle$
Requirements:
$T^{listener}_{last\_move}$ = $\langle$Offer, (goal, proposal)$\rangle$
Effects:
CStore^{speaker} + goal $\wedge$
CStore^{speaker} + proposal $\wedge$
CStore^{speaker} + offer(goal, proposal)

$\langle$Reject, (goal, proposal)$\rangle$
Requirements:
$T^{Hearer}_{last\_move}$ = $\langle$Offer, (goal, proposal)$\rangle$
Effects:
($T^{listener}_{next\_move}$ = $\langle$Offer, (goal, proposal$'$)$\rangle \vee$
$\langle$Offer, (goal$'$, proposal)$\rangle \vee \langle$Offer, (goal$'$, proposal$'$)$\rangle$)

$\langle$Withdraw, (–)$\rangle$
Requirements:
T_{last_move} = $\langle$Offer(goal, proposal)$\rangle \vee$
$\langle$Reject(goal, proposal)$\rangle$
Effects:
Ø

The initial setup for an NP$_0$ dialogue is similar to that for a PPC$_0$ dialogue. Both systems utilise the same number of and types of commitment store, the contents of which are preserved between progressions from one sub-dialogue to another. Both players retain their participant identifiers in an NP$_0$ sub-dialogue that were established in the preceding PP$_0$ sub-dialogue. The similar setups are necessary to enable a clean progression from one sub-dialogue to the next, and a possible subsequent return to the original dialogue type. This approach also enables a consistent representation of supporting machinery between the two systems as required by the A4A.

The global rules for NP$_0$ specify initiation and termination rules. The initiation rules establish that the initiator has some initial thesis in their commitment store and that that same initial thesis is not present in the respondent's commitment store. The initiation rules also establish that an NP$_0$ dialogue must begin with an offer move in which the initiator states the goal that they are trying to achieve, in this case the goal is actually the initial thesis which was established at the very beginning of the encompassing persuasion dialogue, along with a proposal that they are willing to concede to get the goal accepted. An NP$_0$ dialogue can terminate when either the initiator has withdrawn their initial thesis, or the respondent has accepted the initial thesis, or the withdraw move is uttered.

Because of the formulation of the initiation rules, the profiles of dialogues carried out according to NP$_0$ are slightly assymetrical. Although all the moves are conceivably available to all participants, i.e. there are no moves that can only be played by either the initiator or the respondent, an NP$_0$ dialogue will always start with the initiator making an offer that is based upon the initial thesis instantiated at the beginning of the prior PP$_0$ dialogue.

NP$_0$ incorporates four moves which enable basic bargaining behaviour. The offer move, in the context of a negotiation over action, can be assumed to have the following form, "If you accept X, I will concede Y", where X is some goal that the offerer wants the offeree to achieve and Y is the concession that the offerer is willing to make to achieve X. The offer move requires that the speaker has not previously made the same bid. In the case above, all of X, Y, and the utterance *offer(X, Y)* will be added to the speakers commitment store, so NP$_0$ allows commitment to offers as well as commitment with respect to the individual statements that comprise the offers. The requirements for this move stop the speaker from repeating a bid that they have already offered.

The offer move is designed to be recursive and can be followed in a subsequent turn by a counter offer. NP$_0$ recognises four varieties of offer. The first is the initial offer in a negotiation. The remainder are various types of counteroffer in which either, the goal remains the same and the proposal is altered, the goal is altered and the proposal remains the same, or the goal and the proposal are both altered. In the two instances of counteroffers where the goal is altered, it is assumed that the goal is a reduced or related version of the initial goal but the rules do not enforce this. Given the initial offer, "If you accept X, I will concede Y", it should be noted that in the counter-offers the participants are inverted so that the offer should be read as the inversion of the previous offer; for example the first variety of counteroffer is of the form, "I will accept X, If you concede Y$'$", the second variety is of the form, "I will accept X$'$, if you concede Y", and lastly the final type of counteroffer is of the form, "I will accept X$'$, If you concede Y$'$". Notice that because NP$_0$ dialogues are not entirely symmetrical it is always the case that the goal refers to something that the respondent should accept and that the proposal refers to something that the initiator is conceding. After an initial offer is made the next move can be either outright acceptance or rejection of the offer, or one of the varieties of counteroffer. The accept move enables a player to agree to a given offer and adds the components of the offer and the offer itself to the speakers commitment store so that a player actively commits themself to accept an offer. The reject move enables a player to not accept a proposed offer. Finally the withdraw move is similar to that for withdraw in PP$_0$.

It should be noted that NP$_0$ includes no progression rules to govern either return to the parent persuasion dialogue or to enter a new instance of persuasion or negotiation dialogue as a child of the current NP$_0$ dialogue. This was a purposeful omission partly to aid clarity and partly because although a nice capability it is not required to demonstrate either the use or the utility of the progression from persuasion to negotiation during a dialogue. The machinery of the A4A architecture is sufficiently flexible to enables such transitions to be specified as required either in a manner similar to that used for PP$_0$ or by specific ation of a particular move which leads to a progression as part of the effects of playing that move.

5 Example Dialogue

The following dialogue fragment illustrates the canonical embedding of an NP$_0$ sub-dialogue within a PP$_0$ dialogue. The dialogue is situated within a multiagent distributed computation scenario. Each agent has various capabilities, tasks that it can perform. A key aspect is that no single agent knows all other agents within the system or has complete knowledge of the system. The dialogue fragment is as follows:

The fragment involves two agents, agent$_1$ and agent$_2$. The dialogue is initiated by agent$_1$ who becomes the initiator and requests of agent$_2$ who becomes the respondent to perform task S$_1$. S$_1$ is added to the initiator's commitment store. In turn 2 the respondent challenges the request which, because of the burden of proof required by a persuasion dialogue, means that the initiator must defend the standpoint established in turn T$_1$. At T$_3$ the initiator defends their stand-

[hbtp]

Turn	Player	Move	CStore^{init}	CStore^{resp}
1	init	Request(S$_1$)	S$_1$	–
2	resp	Challenge(S$_1$)	–	–
3	init	Defence(S$_2$$\rightarrow$S$_1$)	S$_2$, S$_2$$\rightarrow$S$_1$	–
4	resp	Challenge(S$_1$)	–	–
5	init	defence(S$_3$$\rightarrow$S$_1$)	S$_3$, S$_3$$\rightarrow$S$_1$	–
6	resp	Reject(S$_1$)	–	–
7	init	Offer(S$_1$, S$_4$)	S$_4$, Offer(S$_1$, S$_4$)	–
8	resp	Offer(S$_5$, S$_6$)	–	S$_5$, S$_6$, Offer(S$_5$, S$_6$)
9	init	Offer(S$_1$, S$_7$)	S$_7$, Offer(S$_1$, S$_7$)	–
10	resp	Accept(S$_1$, S$_7$)	–	S$_1$, S$_7$, Offer(S$_1$, S$_7$)

point and the defence is added to the initiator's commitment store. At T$_4$ the respondent is not pursuaded by the initiator's defence and again challenges S$_1$. The initiator responds at T$_5$ with another defence of S$_1$ and the initiator's commitment store is again updated. In T$_6$ the respondent rejects the initiator's standpoint S$_1$. At this point the requirements of the progression rules of PP$_0$ are met and a shift can legally occur from the PP$_0$ dialogue to an NP$_0$ dialogue. The initiator need not utilise this progression however. If the initiator, for some reason, still has an argument that it can use to support S$_1$ then the PP$_0$ dialogue can continue. In this case though the initiator does not have a further argument to support S$_1$ so takes the opportunity to shift to an NP$_0$ dialogue. The initiator achieves this by playing the offer move at T$_7$ in accordance with the initiation rules of NP$_0$. Following the shift to the NP$_0$ dialogue and the initiatory offer move, the respondent responds in T$_8$ with a counter-offer which includes both a different goal and a different proposal to that offered in T$_7$. At T$_9$ the initiator makes another counter-offer again involving the initiators original standpoint, but this time including a new concession S$_7$. the concessions extended in the offer moves may, in the context of the multiagent system scenario, correspond to particular capabilities of the participating agents who offer to perform certain actions in exchange for acceptance of the initial standpoint. At T$_{10}$ the respondent accepts the offer extended in T$_9$ which incorporates the standpoint originally established in T$_1$. At this point the termination rules of NP$_0$ are met and the status of the dialogue is complete.

This fragment illustrates the use of PP$_0$ to engage in a persuasion dialogue followed by a shift to a negotiation dialogue when the arguments of the initiating player are rejected. This is a very useful capability because it means that once the participant's persuasive arguments are exhausted they still have techniques which can allow them to reach an agreement. Without the negotiation protocol and the mechanism for shifting from a persuasion dialogue to a negotiation dialogue the dialogue would have ended much sooner without an acceptable outcome.

6 Conclusions

In this paper a situation was characterised in which the participants in an argumentative dialogue are unable to resolve their conflict through persuasive arguments. The notion of the fallacy of bargaining was introduced as a real-world tactic that is used to get agreement whereby instead of defending their standpoint from attack, the defendent makes an offer to their challenger which involves some unrelated concession. Such a fallacy involves an illicit shift from a persuasion dialogue to a negotiation dialogue. The proposal was made that so long as the shift is licit, i.e. that the shift is clearly and transparently made, and that the shift is not made in order to escape the burden of proof of defending a standpoint, then such a shift does not lead necessarily to a fallacy of bargaining ocurring.

Given this, then in the failed persuasion scenario the participants could shift from a persuasion dialogue to a negotiation dialogue once they ran out of arguments, either to persuade their opponent or to justify their own position. Once in the negotiation dialogue the participants could make offers to each other in relation to the original issue. Such offers, instead of involving persuasive justifications of their standpoints, involve proposing concessions that could be made which aren't necessarily related to the issue at hand. To illustrate the situation, a pair of formal dialectic systems named PP$_0$ and NP$_0$ were introduced along with a mechanism for facillitating the required dialogue shift.

The next step is to refine the formulations of PP$_0$ and NP$_0$ into PP$_1$ and NP$_1$ to enable bi-directional shifts between PP and NP dialogues as well as shifts to sub-dialogues of other types.

REFERENCES

[1] T. J. M. Bench-Capon, T. Geldard, and P. H Leng, 'A method for the computational modelling of dialectical argument with dialogue games', *Artificial Intelligence and Law*, **8**, 223–254, (2000).

[2] R. A. Girle, 'Dialogue and entrenchment', in *Proceedings Of The 6th Florida Artificial Intelligence Research Symposium*, pp. 185–189, (1993).

[3] R. A. Girle, 'Knowledge organized and disorganized', *Proceedings of the 7th Florida Artificial Intelligence Research Symposium*, 198–203, (1994).

[4] R. A. Girle, 'Commands in dialogue logic', *Practical Reasoning: International Conference on Formal and Applied Practical Reasoning, Springer Lecture Notes in AI*, **1085**, 246–260, (1996).

[5] C. L. Hamblin, *Fallacies*, Methuen and Co. Ltd, 1970.

[6] J. D. Mackenzie, 'Question begging in non-cumulative systems', *Journal Of Philosophical Logic*, **8**, 117–133, (1979).

[7] P. McBurney and S. Parsons, 'Agent ludens: Games for agent dialogues', in *Game-Theoretic and Decision-Theoretic Agents (GTDT 2001): Proceedings of the 2001 AAAI Spring Symposium*, (2001).

[8] P. McBurney and S. Parsons, 'Dialogue games in multi-agent systems', *Informal Logic*, **22**(3), 257–274, (2002).

[9] I. Rahwan, *Interest-based Negotiation in Multi-Agent Systems*, Ph.D. dissertation, University of Melbourne, 2004.

[10] C. Reed, 'Dialogue frames in agent communication', in *Proceedings of the 3rd International Conference on Multi Agent Systems*, pp. 246–253. IEEE Press, (1998).

[11] K. Sycara, 'Persuasive argumentation in negotiation', *Theory And Decision*, **28**, 203–242, (1990).

[12] S. Toulmin, *The Uses Of Argument*, Cambridge University Press, 1958.

[13] D. N. Walton, *Logical Dialogue-Games And Fallacies*, University Press Of America, 1984.

[14] D. N. Walton, 'Types of dialogue, dialectical shifts and fallacies', in *Argumentation Illuminated*, pp. 133–147, (1992).

[15] D. N. Walton and E. C. W. Krabbe, *Commitment in Dialogue*, SUNY series in Logic and Language, State University of New York Press, 1995.

[16] S. Wells and C. Reed, 'Formal dialectic specification', in *First International Workshop on Argumentation in Multi-Agent Systems*, eds., I. Rahwan, P. Moraitis, and C. Reed, (2004).