Proceedings of the ECAI 2002 Workshop on

Computational Models of Natural Argument

Giuseppe Carenini
Floriana Grasso
Chris Reed

22nd July 2002
Lyon, France

Programme

Computational Models of Natural Argument

Giuseppe Carenini[1] and **Floriana Grasso**[2] and **Chris A. Reed**[3]

The Workshop. The ECAI 2002 workshop on Computational Models of Natural Argument intends to recognise and consolidate the critical mass that research in the field overlapping Argumentation Theory and Artificial Intelligence has developed in recent years.

As representations and processes investigated in philosophical theory of argumentation, in informal logic, and in dialectics can provide the starting point for computational modelling, the opposite is also true. Efforts within AI to build computational models of argument processing can stimulate researchers in argumentation theory to develop more precise and uniform representations.

Already fruits of cross fertilisation between AI and argumentation theory are beginning to ripen and this workshop will surely foster further interaction and collaboration.

The Programme. The workshop's programme includes eight papers, all of which we believe have great potential to stimulate cross-disciplinary discussion.

A common theme that runs through much of the area, and that is duly represented in several of the papers here, is the relationship between arguments, defeasible logics and probabilistic reasoning. Walton & Reed discuss schemes for presumptive (defeasible) arguments, raising interesting issues on their completeness and on how to express them graphically. Gratton explores how the probabilities of possible countexamples for an argument are related to the degree of support to the argument's claim given its premises. Das presents a Toulmin-based formalism to represent argument schemes that integrates logics and probabilities in a computational model of belief networks.

Several other areas of overlap between AI and argumentation are represented in the workshop program.

In AI, the design of systems that can automatically recognise and generate text is ever more frequently based on the the analysis of large text corpora. Concepts from the pragma-dialectical approach discussed in Snoeck-Henkemans could be extremely valuable to focus an empirical analysis of corpora of (symptomatic) natural arguments.

Yuan, Moore & Grierson unify two dialogue typologies developed in Philosophy and Education. From this analysis, promising ideas emerge for the design of a computer debate system that may effectively improve students' critical thinking and debating skills.

The generation of natural language arguments from logical proofs has received considerable attention in AI. Fielder & Horacek outline a computational model that produces a natural proof description from an underlying machine-generated proof. Several techniques are discussed to select the proper degrees of argument granularity and explicitness by taking into account a model of human performance in processing arguments. In the proposed model, once an argument is presented, it can be further explored/expanded interactively by the user.

Argumentation can play a critical role in modelling autonomous agents and their interactions. Kakas & Moraitis propose a framework, grounded in logic programming and non-monotonic reasoning, that can be applied to model how each agent determines its position in an argumentative dialog. The framework focuses on inter-agent deliberation and decision making, and the interaction with agent preferences.

The legal domain is probably the one in which there is more need for systems that can support the management of large sets of arguments. Prakken & Vreeswijk discuss the development of a system for editing, processing and visualizing arguments involved in a legal civil case. The main challenge they face is to devise argumentation schemes that are expressive enough to represent the subtleties of the case and to support useful processing, but are still understandable to the users.

The combination of such a rich and diverse programme on the one hand, and the substantial overlap and common ground in the assumptions and approaches of the contributions, on the other, promises an exciting meeting with the opportunity for all participants to join in lively, stimulating and productive discussion.

Acknowledgements. Finally, we would like to take this opportunity to thank the programme committee members for all their hard work in reviewing the submissions:

- Fiorella de Rosis, Dept. of Informatics, University of Bari, Italy.
- Nancy Green, Dept. of Mathematical Science, University of North Carolina at Greensboro, USA
- Leo Groarke, Dept. of Philosophy, Wilfrid Laurier University, Waterloo, Ontario, Canada.
- Daniel Marcu, Information Sciences Institute, University of Southern California, USA.
- Simon Parsons, Center for Coordination Science, Sloan School of Management, MIT, USA.
- Ehud Reiter, Dept. of Computer Science, University of Aberdeen, UK.
- Richard Rosenberg, Dept. of Computer Science, University of British Columbia, Canada.
- Doug Walton, Dept. of Philosophy, University of Winnipeg, Canada.

and the ECAI organisers for their help, especially the ECAI workshop coordinator: Alun Preece.

Welcome to the ECAI 2002 workshop on Computational Models of Natural Argument!

[1] Department of Computer Science, University of British Columbia, Vancouver, Canada
[2] Department of Computer Science, University of Liverpool, UK
[3] Department of Applied Computing, University of Dundee, UK

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

Session I

Logic of Probabilistic Arguments

Subrata Das[1]

Abstract. We present a logic for reasoning with probabilistic arguments to help decision making under uncertainty. The syntax of the logic is essentially modal propositional, and arguments of decision makers are expressed as sentences of the logic, with associated supports drawn from a probability dictionary. To aggregate a set of arguments for and against some decision options, we construct a Bayesian belief network based on the argument set without requiring any additional information from the decision-maker. Evidence converted from the underlying knowledge of the decision maker is posted at the relevant nodes of the belief network to compute probability distributions, and hence rankings, among the decision options. Decision-making based on such rankings of decision options is therefore guaranteed to be consistent with probability theory. We develop possible world semantics of the logic, and establish soundness and completeness results. We illustrate the proposed decision-making framework in the context of a concrete example.

1 Introduction

Human decision-making can be regarded as a complex information processing activity, which, according to (Rasmussen, 1983), is divided into three broad categories, corresponding to activities at three different levels of complexity. At the lowest level is skill-based sensorimotor behavior, representing the most automated, largely unconscious level of skilled performance such as deciding to brake upon seeing a car ahead. At the next level is rule-based behavior, exemplified by simple procedural skills for well-practiced, simple tasks such as inferring the condition of a game-playing field based on the current weather. Knowledge-based behavior represents the most complex cognitive processing, used to solve difficult and sometimes unfamiliar problems, for making decisions that require dealing with various factors and uncertain data. Examples of this type of processing include determining the status of a game (i.e. a sporting event), given that there is transport disruption. Our focus here is to develop an argumentation framework to support human decision making at the knowledge base level by providing suggestions as to alternative courses of action, and help determine the most suitable. Human decision makers often weigh the available alternatives and select the most promising one based on the associated pros and cons. The proposed argumentation framework, similar to the one developed in (Das et al. 1997; Das and Grecu, 2000; Fox and Das, 2000), therefore naturally supports human decision-makers by augmenting and complementing their own cognitive capabilities.

Two important requirements must be met if we are to develop a practical and useful decision support system: the system must be declarative and robust. The declarative nature of the system ensures a human readable representation of knowledge and human-like reasoning with knowledge. Robustness of the system ensures its ability to cope with uncertain or missing data in situations where the required knowledge is unavailable in the underlying knowledge base. We plan to make our proposed framework declarative via the use of a high-level logical syntax for representing arguments, including probabilities to represent their strengths. The robustness is assured via representations that allow computations over a range of values, and the use of Bayesian belief network technology (Pearl, 1988) to support combining diverse evidence of arguments for and against decision alternatives. The belief network formalism supports probabilistic reasoning over the causal and evidential relations combining knowledge from decision makers and the current set of beliefs, so that the system can derive probability estimates for adopting particular decision options.

To summarize our framework, we use the syntax of modal propositional logic for representing arguments, and include probabilities to represent their strengths. For the purpose of aggregation of arguments, we automatically transform a set of arguments for and against some decision options into a belief network. The generated belief network then forms the basis for computing aggregated evidence for the decision options according to the strengths of the arguments. This hybrid approach has the following advantages:

- Arguments are expressed in a human readable syntax of modal propositional logic, along with a probability dictionary for expressing their strengths.
- The possible world semantics of the logic that we develop is intuitive to decision makers, as decision options simply correspond to various possibilities mapped to possible worlds.
- Aggregation is carried out on a belief network that is automatically constructed out of available arguments, and no additional knowledge needs to be acquired.

The rest of the paper is organized as follows: Section 2 presents an argumentation-based decision-making framework. Section 3 presents the underlying logic of arguments in the proposed framework. Section 4 presents an approach to argument aggregation via Bayesian belief networks. Section 5 presents a concrete example to illustrate the syntax and semantics of the logic and the argumentation and aggregation process. Each of Section 3 and Section 4 can be read independently of the other, but the example in Section 5 requires understanding of both the logic and the aggregation process. Throughout the paper we use the single example of the status of a ball game, which is scheduled to occur sometime today. Proof of theorems and propositions stated in the paper have been omitted due to space limitations. The proofs can be found in (Fox and Das, 2000).

2 Decision Making via Argumentation

This section presents the non-temporal version of the argumentation-based decision-making framework that was developed in (Fox and Das, 2000; Das et al. 1997), but focusing only on probabilistic arguments. We first provide a brief historical background of argumentation. Then we provide a concrete example to illustrate the use of argumentation, followed by the formal 'domino' model of argumentation and a knowledge representation language for expressing decision constructs and beliefs and knowledge in the model.

[1] CHARLES RIVER ANALYTICS, Cambridge, MA 02138, USA

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

2.1 Brief Background in Argumentation

Toulmin in his book (Toulmin, 1956) discussed how difficult it is to cast everyday practical arguments into classical deductive form. He claimed that arguments needed to be analyzed using a richer format than the simple if-then form of classical logic. He characterizes practical argumentation by means of the scheme in Figure 1.

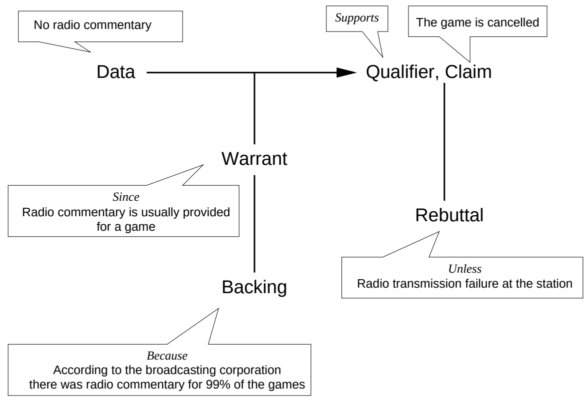


Figure 1. Toulmin's model of argumentation

As shown in Figure 1, Toulmin's model decomposes an argument into a number of constituent elements: 1) Claim: the point a decision maker is trying to make; 2) Data: the facts about a situation provided to support the claim; 3) Warrant: statements indicating general ways of arguing; 4) Backing: generalizations providing explicit support for an argument; 5) Qualifier: phrases showing the confidence an argument confers on a claim; 6) Rebuttal: acknowledges exceptions or limitations to the argument. To illustrate, consider an argument claiming that the game, which was supposed to be held today, has been cancelled. The facts or beliefs (that is, data) on which this claim is made are that there is no radio commentary for the game in question. General principles or rules, such as "radio commentary is usually provided for a game", warrant the argument, based on statistical research published by the broadcasting corporation, which is the backing. Since the argument is not conclusive we insert the qualifier "supports" in front of the claim, and note the possibility that the conclusion may be rebutted on other grounds, such as failure of radio transmission of the commentary. Our approach is to transform Toulmin's work to a more formal setting, much the same way as in (Fox et al, 1992). We too deal with the concepts of warrant and rebuttal, but as very simple prepositional arguments for and against. We do not deal with first-order sentences that are more suitable for representing backings in Toulmin's model. We introduce the use of a single qualifier called 'support'.

2.2 Example Decision Making Process

We explain here the argumentation based decision-making framework in (Fox and Das, 2000), continuing with our ball-game example as shown in Figure 2.

The process starts when the decision maker observes transport disruption on the way to catch a public transport (e.g. a bus) to go to town for the game. The newly discovered transport status then becomes the decision maker's belief. Given that the decision maker "believes" that there is transport disruption, it raises a "goal" of finding the status of the game. It then infers from its common sense

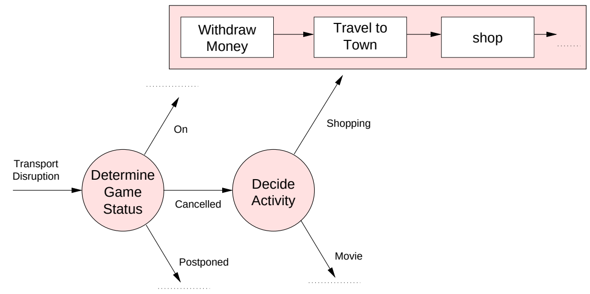


Figure 2. Decision-making flow

knowledge that there are three possible or "candidate" states of the game, On, Cancelled, and Postponed, and so constructs arguments for and against these alternatives. These arguments use other beliefs of his, based on observations such as the weather and radio commentary. In this case the balance of "argument" is in favor of the game being cancelled, and this conclusion is added into the decision maker's database of beliefs. Given this new belief regarding the cancelled status of the game, a new goal is raised, i.e. to plan for alternative activities. As in determining the status of the game, here there are two options for alternative activities, shopping and going to a movie, and the decision maker once again constructs arguments for the alternatives, taking into account transport, cost, etc., and recommends going shopping as the most preferred alternative activity on the basis of the arguments. The adoption of a shopping "plan" leads to an appropriate schedule of "actions" involved in shopping, such as withdrawing money, traveling to town, going to stores, etc. The effects of these actions are recorded in the decision maker's database, which may lead to further goals, and so on.

2.3 The Domino Model

Figure 3, the 'domino' model, captures graphically the decision-making framework, where the outer chain of arrows in the figure represents the above example decision-making process. Within our proposed framework, a decision schema has several component parts: an evoking situation, a goal, one or more candidates, and one or more commitment rules.

A situation describes, as a boolean expression on the database of beliefs, the situation or event which initiates decision making. For example, a belief that an abnormality (e.g. transport disruption) is present may lead to a choice between alternative possible causes/effects of it.

A goal is raised as soon as the evoking situation occurs. In particular, the belief that an abnormality is present may raise the goal of determining its cause or effects. For example, if transport is disrupted then one of its possible effects is the cancellation of the game, so therefore the goal is to determine game status. On the other hand, if there is no radio commentary then a goal is to determine the status of the game, as its cancellation causes no radio commentary. Typically, a goal is represented by a property that the decision maker tries to bring about.

Candidates are a set of alternative decision options, such as on, cancelled, postponed. In principle the set of candidates may be defined extensionally (as a set of propositions) or intentionally (by rules), but we only consider the former case here.

Arguments are modal-propositional rules that define the arguments that are appropriate for choosing between candidates for the

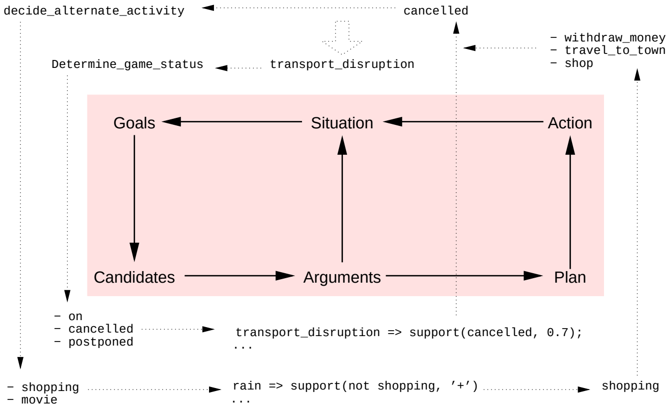


Figure 3. Domino process view of the example

```
decision:: game_status
    situation
        transport_disruption
    goal
        determine_game_status
    candidates
        on;
        cancelled;
        postponed
    arguments
        transport_disruption => support(cancelled, 0.7);
        not radio_commentary => support(not on, 0.9);
        not rain => support(on, 0.95);
        bad_economy => support(not cancelled, 0.6);
        bad_economy & free_slot => support(postponed, 0.7);
    commits
        netsupport(X, U) & netsupport(Y, V) &
        netsupport(Z, W) & U > V & U > W => add(X).
```

Figure 4. Example decision construct

decision. Argument schemas are typically concerned with evidence when the decision involves competing hypotheses (beliefs), and with preferences and values when the decision is concerned with actions or plans.

Commitment rules define the conditions under which the decision may be recommended, or taken autonomously, by the decision maker. It may include logical and/or numerical conditions on the argument and belief databases.

The following section represents a decision schema and its components as described above into a decision construct.

2.4 Decision Constructs

The concept of the domino decision scheme and its components is captured in a high-level declarative syntax. Figure 4 gives the decision construct representing the 'Determine Game Status' decision circle in Figure 2. All decisions have an evoking situation which, if the decision maker believes it to be true, raises the corresponding goal. The three possible paths from the decision circle go to the following three alternative pathways: on, cancelled, and postponed. These candidates are represented explicitly in the decision construct. The arguments and commitments within a decision construct are also represented directly.

The decimal number in an argument represents the probabilistic measure of support given by the argument to the decision candidate. The basic idea is that an argument is a reason to believe something or a reason to act in some way and an argument schema is a rule for generating such reasons during decision making. The more arguments there are for a candidate belief or action, then the more a decision maker is justified in committing to it. The aggregation function can be a simple "weighing up of pros and cons" (netsupport), but it represents a family of more or less sophisticated functions by which we may assess the merit of alternative candidates based on the arguments about them.

In general, an argument schema is like an ordinary inference rule with

```
support(<candidate>, <sign>)
```

as its consequent, where `<sign>` is drawn from a set called a dictionary. The `<sign>` represents, loosely, the confidence that the inference confers on the candidate. The dictionary may be strictly quantitative (e.g. the numbers in the [0,1] interval) or qualitative (e.g. the symbols +, - or pro, con). Here we are dealing with probabilistic arguments and `<sign>` is drawn from the probability dictionary [0,1]. An example argument from the decision construct in Figure 4 is

```
transport_disruption =>
        support(cancelled,0.7)
```

where `<candidate>` is 'cancelled'. Informally, the argument states that if there is transport disruption then there is 70% chance that the game will be cancelled. The rest of the arguments of the decision construct provide support for and against the decision options based on the evidence of radio commentary, weather, and hosting club's economic condition, and availability of free slots for rescheduling the game. A knowledge base for the decision maker consists of a set of definitions of this and other kinds of tasks.

A decision maker considers the decision `game_status` in Figure 4 for activation when the belief `transport_disruption` is added to the database. When the decision maker detects this, it checks whether any of the candidates has already been committed. If not, the decision will be activated and the goal `determine_game_status` is raised; otherwise no action is taken. While the goal is raised, further information about the situation (e.g. the weather) can be examined to determine whether the premises of any argument schemas are instantiated.

A commitment rule is like an ordinary rule with one of

```
add(<property>)
schedule(<plan>)
```

as its consequent. The former adds a new belief to the knowledge base and the latter causes an action to be scheduled as follows (see Figure 5):

See (Fox and Das, 2000) for information on how to deal with a scheduled plan that is committed. When a decision is in progress then, as additional arguments become valid, the decision's commitment rules are evaluated to determine whether it is justified to select a candidate. A commitment rule will often make use of an aggregation function such as 'netsupport' but this is not mandatory. The netsupport function evaluates collections of arguments for and against any candidate to yield an overall assessment of confidence and establish an ordering over the set of candidates; this ordering may be based on qualitative criteria or on quantitative assessment of the strength of the arguments. This function has the form:

```
netsupport(<candidate>, <support>)
```

```
decision:: alternative_activity
    situation
        cancelled
    goal
        decide_alternative_activity
    candidate
        shopping;
        movie
    arguments
        rain => support(no shopping, 0.8);
        ...
    commits
        ... .
```

Figure 5. Example decision construct

In section 4, we implement the 'netsupport' function using an algorithm for evidence propagation in belief networks (Pearl, 1988; Jensen, 1996).

3 Logic of Arguments

The section presents the underlying logic of the argumentation-based decision-making framework, L_{Arg}, as described above, its possible world semantics, and the soundness and completeness results.

3.1 The Syntax

Suppose P is the set of all propositions, representing properties and actions, and includes the special property symbol $\top$ (true). Note that the logic does not distinguish between properties and actions; rather they are treated uniformly as propositions. L_{Arg} is essentially a propositional logic extended with certain modal operators. The modal operators $\langle bel \rangle$ and $\langle goal \rangle$ of L_{Arg} correspond to beliefs (Fagin, 1988; Hintikka, 1962) and goals (Cohen and Levesque, 1990) respectively. Propositions are supported by *collections of arguments*, and the confidence in a proposition or argument is represented by a number between 0 and 1. Suppose D is the dictionary [0, 1] with the top element Δ as 1. In addition, for each dictionary symbol $d \in D$, we have a modal "support" operator $\langle sup_d \rangle$ in L_{Arg}. The *formulae* (or *assertions*) of L_{Arg} extend the domain of propositional formulae to the domain of *formulae* as follows:

- *propositions* are formulae.
- $\langle bel \rangle F$ and $\langle goal \rangle F$ are formulae, where F is a formula.
- $\langle sup_d \rangle F$ is a formula, where F is a formula and d is in the dictionary D.
- $\neg F$ and $F \wedge G$ are formulae, where F and G are formulae.

We take $\perp$ (false) to be an abbreviation of $\neg\top$. Other logical connectives and the existential quantifier are defined using $\neg$ and $\wedge$ in the usual manner.

3.2 Example Sentences and Arguments

We provide here some example sentences of L_{Arg} that are translations of the decision construct shown in Figure 4. The situation and goal portion in the decision `game_status` is translated to the following modal rule:

$$\langle bel \rangle transport\_disruption \rightarrow \langle goal \rangle determine\_game\_status$$

The above L_{Arg} sentence states that if *transport_disruption* is believed, then a goal is *determine_game_status*. A goal is considered to be achieved as soon as it becomes true. In the context of the decision *game_status*, this is reflected in the following formulae:

$$\langle bel \rangle (on \wedge \neg cancelled \wedge \neg postponed) \rightarrow \langle bel \rangle determine\_game\_status$$
$$\langle bel \rangle (cancelled \wedge \neg on \wedge \neg postponed) \rightarrow \langle bel \rangle determine\_game\_status$$
$$\langle bel \rangle (postponed \wedge \neg on \wedge \neg cancelled) \rightarrow \langle bel \rangle determine\_game\_status$$

Figure 6. Translation of the goal in the decision construct shown in Figure 4

The first of the above four sentences (Figure 6) states that if it is believed that the game is on, but neither cancelled nor postponed, then *determine_game_status* is believed. In other words, the earlier goal *determine_game_status* is considered achieved upon believing that the game is on. The L_{Arg} representations for the arguments in the diagnosis decision are (Figure 7):

$$\langle bel \rangle transport\_disruption \rightarrow \langle sup_{0.7} \rangle cancelled$$
$$\langle bel \rangle \neg radio\_commentary \rightarrow \langle sup_{0.9} \rangle \neg on$$
$$\langle bel \rangle \neg rain \rightarrow \langle sup_{0.95} \rangle on$$
$$\langle bel \rangle bad\_economy \rightarrow \langle sup_{0.6} \rangle \neg cancelled$$
$$\langle bel \rangle (bad\_economy \wedge free\_slot) \rightarrow \langle sup_{0.7} \rangle postponed$$

Figure 7. Translation of the arguments in the decision construct shown in Figure 4

3.3 The Axioms

The axioms of L_{Arg} are divided into classical and modal axioms. For classical axioms, we consider every instance of a propositional tautology to be an axiom, and we also have the *modus ponens* inference rule. L_{Arg} adopts a standard set of axioms and inference rules of beliefs and goals in its reasoning and decision making, which can be found in (Cohen and Levesque, 1990; Meyer et al, 1991). A detailed explanation can be found in (Fox and Das, 2000). The L_{Arg} axioms and inference rules are:

$$\neg \langle bel \rangle \perp, \neg \langle goal \rangle \perp$$
$$\langle bel \rangle F \wedge \langle bel \rangle (F \rightarrow G) \rightarrow \langle bel \rangle G$$
$$\langle bel \rangle F \rightarrow \langle bel \rangle \langle bel \rangle F$$
$$\neg \langle bel \rangle F \rightarrow \langle bel \rangle \neg \langle bel \rangle F$$
$$\langle goal \rangle F \wedge \langle goal \rangle (F \rightarrow G) \rightarrow \langle goal \rangle G$$
$$\langle bel \rangle F \rightarrow \langle goal \rangle F$$
$$\text{if} \vdash F \text{ then} \vdash \langle bel \rangle F$$

We now present a set of axioms for the modal operator $\langle sup_d \rangle$. First of all, there can be no support for an inconsistency and this is axiomatized as follows:

$$\neg \langle sup_d \rangle \perp, \text{for every } d \in D$$

The following inference rule states that the support operator is closed under implication. In other words, if F has support d and

$F \rightarrow G$ is valid in then G too has support d.

$$\text{if } \vdash F \rightarrow G \text{ then } \vdash \langle sup_d \rangle F \rightarrow \langle sup_d \rangle G, \text{ for every } d \in D$$

A valid L_{Arg} formula always has the highest support:

$$\text{if} \vdash F \text{then} \vdash \langle sup_{\Delta} \rangle F$$

Support operators can be combined to obtain a single support operator by using the following axiom:

$$\langle sup_{d1} \rangle F \wedge \langle sup_{d2} \rangle G \rightarrow \langle sup_{d1 \oplus d2} \rangle (F \wedge G)$$

where $\oplus : D \times D \rightarrow D$ is the function for computing supports for assertions derived through material implication. The axiom states that if $d1$ and $d2$ are supports for F and G respectively then $\oplus(d1, d2)$ (or $d1 \oplus d2$ in infix notation) is a derived support for $F \wedge G$. Note that $d \oplus \Delta = d$, for every d in D. If $F = G$, then the above axiom basically aggregates two arguments for the decision option F. Such aggregation via belief networks will be presented in the following section. The following axiom says that every level of evidence for an assertion also implies every level of evidence for the assertion lower than the evidence:

$$\langle sup_{d1} \rangle F \rightarrow \langle sup_{d2} \rangle F, \text{where } d2 \leq d1$$

3.4 Possible World Semantics

A *model* of L_{Arg} is a tuple

$$\langle W, V, R_b, R_s, R_g \rangle$$

in which W is a set of possible worlds. A world consists of a set of qualified assertions outlining what is true in the world. V is a valuation that associates each world with a subset of the set of propositions. In other words,

$$V : W \rightarrow \Pi(P)$$

where P is the set of propositions and $\Pi(P)$ is the power set of P. The image of the world w under the mapping V, written as $V(w)$, is the set of all propositions which are true in the world w. This means that p holds in w for each p in $V(w)$.

The relations R_b, R_s and R_g are the accessibility relations for beliefs, supports and goals respectively. For example, the relation R_b relates a world w to a set of worlds considered possible by the decision-maker from w. If there are n candidates for a decision that are active in a world w then there are n possible worlds.

The relation R_s is a *hyperelation* which is a subset of the set

$$W \times D \times \Pi(W)$$

Semantically, if $\langle w, d, W' \rangle \in R_s$ then there is an amount of support d for committing to one of the possible worlds in W' from the world w, where W' is non-empty. In other words, the support d is for the set of assertions uniquely characterized by the set of worlds W'.

An assertion is a *belief* of a decision maker at a world w if and only if it is true in all possible worlds that are accessible from the world w by R_b. Note that the members of R_s have been considered to be of the form $\langle w, d, W' \rangle$ rather than $\langle w, d, w' \rangle$. The main reason is that the derivability of $\langle sup_d \rangle F$ means F is true only in a "subset" of the set of all possible worlds accessible from w. If F is true in all possible worlds accessible from w then we would have had $\langle bel \rangle F$, which implies the highest form of support for F that is greater than or equal to d.

Due to the axioms related to the modal operator $\langle bel \rangle$, the standard set of properties that will be possessed by the accessibility relation R_b is:

Model Property 1: R_b is serial, transitive, and euclidean

The requirement that a decision maker may not believe in something that is inconsistent guarantees the existence of a possible world, which is the seriality property. The explanation for R_b being transitive and euclidean can be found in (Chellas, 1980; Lemmon, 1977).

The hyperelation R_s satisfies the following properties due to the axioms related to the modal operator $\langle sup_d \rangle$:

Model Property 2: For every w, w_1, w_2 in W and d, d' in D, the relation R_s satisfies the following conditions:

- if $\langle w, d, W' \rangle \in R_s$ then $W' \neq \emptyset$.
- if $\langle w, d, W' \rangle \in R_s$ then $\langle w, d', W' \rangle \in R_s$, for every $d' \leq d$.
- $\langle w, \Delta, W \rangle \in R_s$.
- if $\langle w, d1, W_1 \rangle, \langle w, d2, W_2 \rangle \in R_s$ then $\langle w, d_1 \oplus d_2, W_1 \cap W_2 \rangle \in R_s$, provided $W_1 \cap W_2 \neq \emptyset$.

Explanation of each of these restrictions on Rs can be found in (Das and Fox, 2000).

Aggregation of arguments introduces a hierarchy of preferences among the set of all possible worlds accessible from w by the relation R_b. The maximal elements and possibly some elements from the top of the hierarchy of this preference structure will be called goal worlds. The relation R_g, which is a subset of R_b, relates the current world to the set of goal worlds. Only one of the goal worlds is committed for transition from the current world based on the aggregated support. This world will be called the *committed world*.

An assertion is a *goal* in a world w if and only if it is true in every goal world accessible from w by the accessibility relation R_g. Axiom $\neg \langle goal \rangle \perp$ introduces the seriality property on the accessibility relation R_g. Axiom $\langle bel \rangle F \rightarrow \langle goal \rangle F$ restricts R_g to a subset of R_b, that is, the set of goal worlds is a subset of the set of all possible worlds.

Model Property 3 :

- R_g is serial
- $R_g \subseteq R_b$: for every w and w' in W, if wR_gw' then wR_bw'

The semantics of supports, beliefs and goals are as follows. Given a model $M = \langle W, V, R_b, R_s, R_g \rangle$, the truth values of formulae with respect to a world w are determined by the rules given below:

$\models_{M_w} \top$
$\models_{M_w} p$ iff $p \in V(w)$
$\models_{M_w} \langle sup_d \rangle F$ iff there exists $\langle w, d, W' \rangle$ in R_s such that $\models_{M_w} F$, for every $w' \in W'$
$\models_{M_w} \langle bel \rangle F$ iff for every w' in W such that wR_bw', $\models_{M_w} F$
$\models_{M_w} \langle goal \rangle F$ iff for every w' in W such that wR_gw', $\models_{M_w} F$
$\models_{M_w} \neg F$iff $\not\models_{M_w} F$
$\models_{M_w} F \wedge G$ iff $\models_{M_w} F$ and $\models_{M_w} G$

A formula F is said to be true in model M if and only if $\models_{M_w} F$, for every w in W. A formula F is said to be valid if F is true in every model.

Suppose Γ is the class of all models satisfying Model Property 1, Model Property 2, and Model Property 3. Then the soundness and completeness theorem establishes the fact that L_{Arg} is determined by Γ.

4 Aggregation of Probabilistic Arguments via Belief Networks

This section presents our approach to aggregating arguments via Bayesian belief network technology. This aggregation process is a meta-level reasoning that takes the clauses in the underlying knowledge base as input. The reasoning at the object or knowledge base level is carried out using the logic L_{Arg}. We first provide a brief background in the technology and then present the details of the approach.

4.1 Review of Bayesian Belief Networks

A Bayesian belief network (Pearl, 1988; Jensen, 1996) is a graphical, probabilistic knowledge representation of a collection of variables describing some domain. The nodes of the belief network denote the variables and the links denote causal relationships between the variables. The topology encodes the *qualitative* knowledge about the domain. Conditional probability tables (CPTs) encode the *quantitative* details (strengths) of the causal relationships between a node and its parents. In other words, the CPTs are *local* joint probability distributions involving subsets of the whole domain. For example, if a variable, x, is 4-valued and has one parent variable, y, which is 3-valued, then x's CPT can be represented as a 3×4 table where the $(i, j)^{th}$ entry is $p(x_j \mid y_i)$. The belief network of Figure 8 encodes the relationships over a simple domain consisting of the six binary variables, *Injury, Rain, Game, Transport, Electricity*, and *Commentary*.

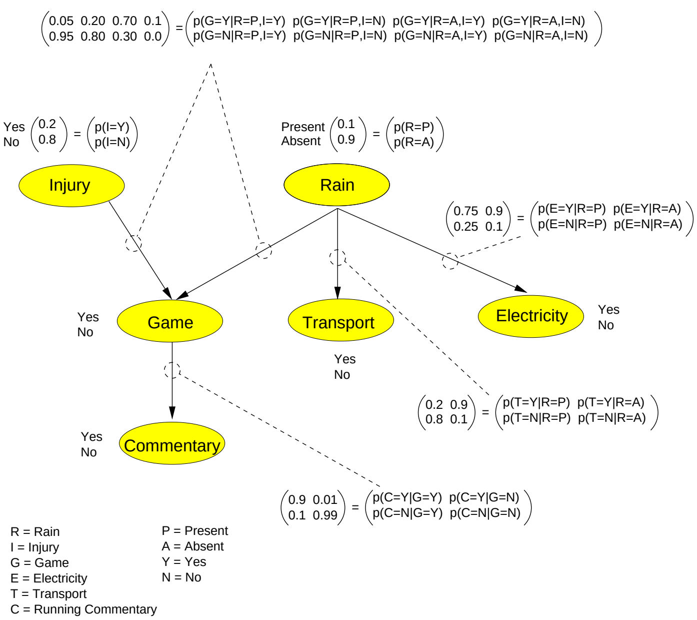


Figure 8. Simple Bayesian belief network

The topology captures the commonsense knowledge that:

1. *Rain* causes *Transport* disruption
2. *Rain* causes *Electricity* failure
3. *Game* causes running *Commentary* on the radio
4. *Injury* and *Rain* prevent *Game* from being played

As shown in Figure 8, the CPT specifies the probability of each possible value of the child variable conditioned on each possible combination of parent variable values. For example, the probability of having electricity given that rain is present is 0.75, whereas the probability of having electricity given clear skies is 0.9.

The structure of a belief network encodes other information as well. Specifically, the lack of links between certain variables represents a lack of direct causal influence, that is, they indicate conditional independence relations. This belief network encodes many independence relations, for example,

1. *Electricity* $\perp$ *Transport* | *Rain*
2. *Commentary* $\perp$ { *Rain, Electricity* } | *Game*

where '$\perp$' is read 'is independent of' and '|' is read 'given.' Once the value of *Rain* is known, the value of *Transport* adds no further information about *Electricity*. Similar conditional independence assertions hold for other variables.

A central feature of the BN formalism is that the belief vector is decomposed as a product of the total *causal* evidence at x, which comes from x's parents, and the total *diagnostic* evidence at x, which comes from x's children. Root nodes are special cases; they require some initial estimate for their causal evidence vectors. Belief vectors generally change as new evidence regarding any of the variables is added to the network. Thus, if we obtain new evidence of electricity being present, our initial belief about rain, i.e. (Present = 0.1, Absent = 0.9), should be revised accordingly, e.g. to (Present = 0.2, Absent = 0. 8). This is an example of *diagnostic* reasoning from effects back to possible causes. This new evidence should also cause us to revise our belief vector for *Game* to reflect a higher probability that the game will be played, e.g. to (Yes = 0.91, No = 0.09). This is an example of *causal* reasoning from causes to effects. Thus, belief nets can support the model-based anomaly diagnosis both by hypothesis generation (diagnostic reasoning) and hypothesis testing (causal reasoning). Additionally, the topologies of the networks themselves can capture the structure and interconnection of the components at hand in an aggregate and easily understood manner.

When new evidence is posted to a variable in a BN, that variable updates its own belief vector, then sends out messages indicating updated predictive and diagnostic support vectors to its children and parent nodes respectively. These messages are then used by the other nodes to update their belief vectors and propagate their own updated support vectors. The separation of evidence yields a propagation algorithm (Pearl, 1988) in which update messages need only be passed in one direction between any two nodes following posting of evidence. Thus, the algorithm's complexity in a polytree type of network is proportional to the number of links in the network. This separation also automatically prevents the possibility of double-counting evidence.

In summary, a Bayesian Belief Network (Pearl, 1988; Lauritzen and Spiegelhalter, 1988) offers these principal advantages compared to other probabilistic reasoning methods:

1. Its use of cause/effect relationships is intuitive.
2. Its probability estimates are guaranteed to be consistent with probability theory.

The following section details our use of belief network technology for aggregating arguments for and against decision options.

4.2 Aggregation of Arguments

An argumentation based decision-making framework like the one described here is functionally similar to classical rule-based experts systems, with the following exceptions:

- It deals with more expressive knowledge in the form of arguments, than simply rules and a variety of dictionaries.
- It incorporates an inference mechanism which is capable of aggregating arguments for and against decision options and therefore more general than simple forward chaining.

While various types of classical, modal, and temporal logics can be used to represent and reason deductively with arguments, inferencing schemes within logics are insufficient for aggregating arguments, as the typical aggregation process is a meta-level reasoning involving sets of arguments. We propose here a scheme for aggregating arguments via Bayesian belief networks. The evidence propagation mechanism in belief networks implements both abductive and deductive inference schemes. While it is easier to elicit a set of arguments, constructing a belief network involves a more methodical approach to knowledge elicitation, and is usually much more time consuming. But a major advantage of an argumentation based framework is that support can be provided for making decisions even with a very few arguments, making the framework highly robust. But the propagation algorithm in a belief network fails to work even if a single entry within a CPT of the network is missing.

As pointed out in (Korver and Lucas, 1993), due to differences in the type of knowledge represented and in the formalism used to represent uncertainty, much of the knowledge to building an equivalent belief network could not be extracted from a rule-based expert system. In our approach, we will be able to extract the network structure fully, but cannot extract every entry in the conditional probability tables. The missing probabilities for variable states are assumed by default to be equally distributed. There are various approaches (Krause, 2000) to learning belief networks from sample data sets. For example, the approach taken in (Heckerman, 1996; Ramoni and Sebastiani, 1997) considers cases where both network structures and probabilities can be learned. The major assumption for learning probabilities from a complete data set is that the distribution for the variable representing probability vectors is considered to be *Dirichlet*. On the other hand, the *Gibbs sampling* technique is often employed to deal with incomplete data sets. Such techniques can be easily incorporated within our approach to estimate the probabilities that were assumed by default, provided relevant sample data sets are available.

Jitnah et. al. (2000) generates rebuttals in a Bayesian argumentation system based on normative and user models, represented in belief networks, that are manually constructed beforehand. The tutoring system proposed in (Conati et. al., 1997) automatically generates and updates belief networks during its interaction with the student for solving a problem. However, these approaches are only vaguely related to our approach to building a belief network, which is to be used for aggregating arguments, and does not seek for additional knowledge from the decision maker. We first construct fragments of networks using the arguments relevant to the decision-making task at hand. Note that, given a network fragment with a variable, and its parents and CPT, the fragment can be equivalently viewed as a set of arguments. For example, consider the network fragment in Figure 9, which states that player injury and rain together can determine the status of the game.

Each column of the CPT yields an argument for and an argument against a state of the variable *Game*. For example, if there is player injury and it rains then there is an argument for a game with support 0.05.

```
injury & rain =  support(game, 0.05)
```

Since the arguments are probabilistic, corresponding to the above argument there will be another argument which states that if there is

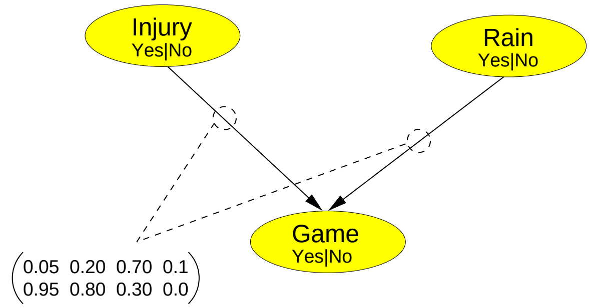


Figure 9. Example belief network fragment

player injury and it rains then there is an argument against the game with support 1 (0.05, that is, 0.95, yielding the following:

```
injury & rain =>  support(not game, 0.95)
```

The rest of the entries of the CPT can be translated to arguments in a similar manner.

Continuing with our illustration of the network construction process from a set of arguments, consider the decision construct shown in Figure 4. Each argument with a single antecedent is translated to a network fragment containing two random variables corresponding to the antecedent and the consequent of the argument. For example, the argument

```
transport_disruption =>
             support(cancelled, 0.7)
```

is translated to the network fragment on the left of Figure 10, which has two nodes or random variables: one for the antecedent `transport_disruption` and the other one for the decision option in the consequent. Since a particular decision option may occur in consequents of many arguments, their corresponding nodes in the network fragments are numbered to avoid ambiguity. Thus, the consequent of the above argument is translated to a node labeled Cancelled-1.

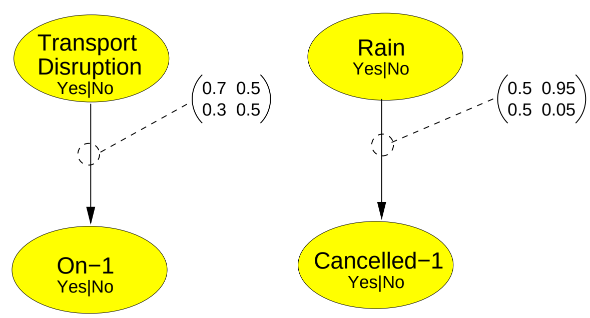


Figure 10. Belief network fragments by converting arguments

The following entry in the CPT comes directly from the argument:

P(Cancelled-1 = Yes | Transport Disruption = Yes) = 0.7
P(Cancelled-1 = No | Transport Disruption = Yes) = 0.3

The above type of probabilities will be equivalently written as the following:

P(Cancelled-1 | Transport Disruption) = 0.7
P(not Cancelled-1 | Transport Disruption) = 0.3

In case of no transport disruption, we have no information relating it to the cancellation of the game. Therefore, the probability distribution among the cancellation and non-cancellation states is even (uniform) given there is no transport disruption:

P(Cancelled-1 | not Transport Disruption) = 0.5
P(not Cancelled-1 | not Transport Disruption) = 0.5

Similarly, the network fragment on the right of Figure 10 is obtained by translating the argument

```
not rain =>  support(on, 0.95)
```

In this case, the above argument generates the following entries of the CPT:

P(On-1 | not Rain) = 0.95
P(not On-1 | not Rain) = 0.05

Since we cannot say anything about the state of the game given rain, the other two entries of the CPTs are as follows:

P(On-1 | Rain) = 0.5
P(not On-1 | Rain) = 0.5

An argument with multiple conditions is translated into a network fragment in a similar manner. Consider the following argument for postponing the game that has two conditions:

```
bad_economy & free_slot =>
        support(postponed, 0.7)
```

The translated network is shown in Figure 11. Observe that we are only able to fill in only one column of the CPT and each of the rest of the columns is uniformly distributed.

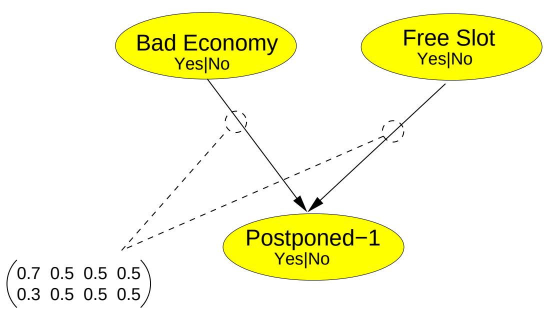


Figure 11. Belief network fragment by converting arguments with multiple conditions

After translating each individual argument to a belief network fragment, the next task is to aggregate arguments for and against each decision option. The heuristic used here is that the probability distribution of the two states of the variable corresponding to a decision option after the aggregation is proportional to the number of arguments for and against the decision option. For example, if we have three arguments for the decision option On via the three nodes On-1, On-2, and On-3, and no arguments against then we have the following probabilities for and against On:

P(On | On-1, On-2, On-3) = 1.0
P(not On | On-1, On-2, On-3) = 0.0

On the other hand, for example, if we have two arguments for the decision option On via the two nodes On-1 and On-2 and one argument against via the node On-3 then we have the following:

P(On | On-1, On-2, not On-3) = 2/3
P(not On | On-1, On-2, not On-3) = 1/3

This is illustrated in Figure 12.

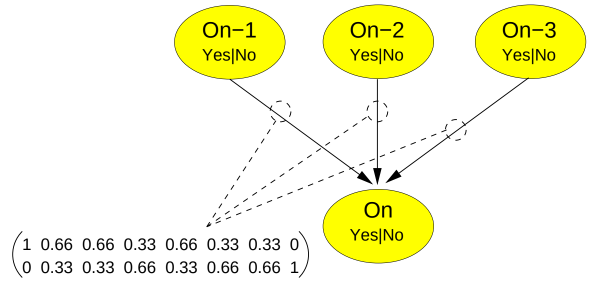


Figure 12. Belief network fragments by converting arguments for/against a decision option

Now that we have network fragments for arguments for and against individual decision options, we need to combine these arguments to rank the decision options. For this, we create a random variable with the states corresponding to the decision options for the task at hand. In the context of our example, we create a random variable called Game with three states On, Cancelled, and Postponed. The variable has three parents corresponding to the three decision options. The decision options are ranked based on the aggregation of arguments for and against the decision options; the values of the CPT are determined accordingly. For example, if we have aggregated evidence for each of the three decision options On, Cancelled, and Postponed, then the probability distribution of the variable Game is evenly distributed as follows:

P(Game = On | On, Cancelled, Postponed) = 0.33
P(Game = Cancelled | On, Cancelled, Postponed) = 0.33
P(Game = Postponed | On, Cancelled, Postponed) = 0.33

Note that we have the same probability distribution when we have aggregated evidence against each of the three decision options. On the other hand, for example, if we have aggregated evidence for each of the two decision options On and Cancelled, and aggregated evidence against the decision option Postponed, then the probability distribution on the states of the variable Game is as follows:

P(Game = On | On, Cancelled, not Postponed) = 0.5
P(Game = Cancelled | On, Cancelled, not Postponed) = 0.5
P(Game = Postponed | On, Cancelled, not Postponed) = 0.0

This is illustrated in Figure 13.

Figure 14 shows the combined network for aggregating the arguments of the decision construct in Figure 4. Such a network has three blocks: the Argument Block, the Axiom Block, and the Aggregation Block. The Argument Block is constructed out of the network fragments obtained by translating the arguments in the decision construct. The Axiom Block, to some extent, implements a specific case of axiom $\langle sup_{d1}\rangle F \langle sup_{d2}\rangle G \rightarrow \langle sup_{d1 \oplus d2}\rangle (F \wedge G)$ (when $F = G$). The Aggregation Block implements the commitment rule in the decision construct. Mismatch is expected between the network in Figure 8 and that of in Figure 14 as any complete network of the

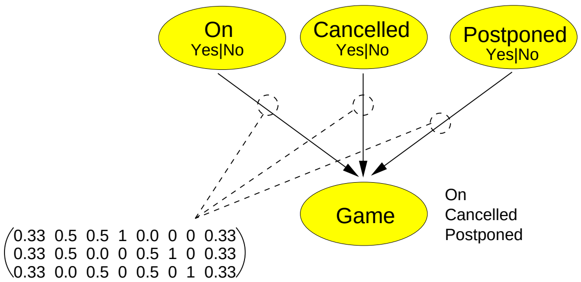


Figure 13. Belief network fragment for aggregating arguments for/against decision options

former type is carefully constructed via a knowledge elicitation effort. (One can always incorporate additional knowledge from experts into the constructed network for improved prediction.)

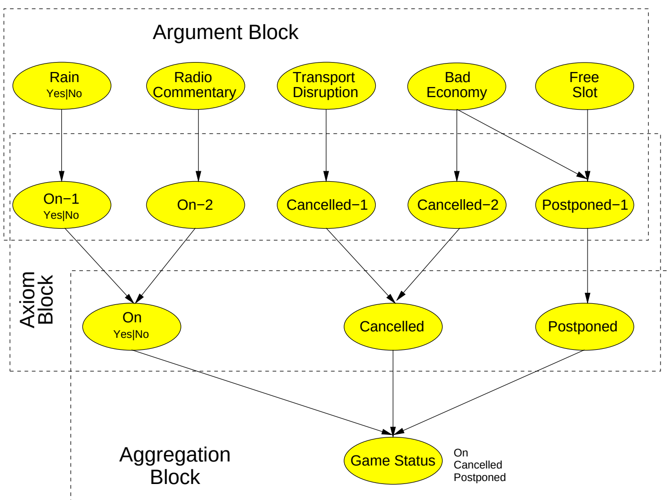


Figure 14. Combined belief network for argument aggregation

In the absence of any evidence, no arguments are generated and the *a priori* probabilities of the decision options are as follows:

P(Game = On) = 0.33
P(Game = Cancelled) = 0.32
P(Game = Postponed) = 0.35

No evidence in the network has been posted at this stage, not even for any prior beliefs on the variables. Now, given that there is transport disruption and rain, the network ranks the decision options based on the following posterior probabilities (as shown in the figure):

P(Game = Postponed | Transport Disruption, Rain) = 0.37
P(Game = Cancelled | Transport Disruption, Rain) = 0.37
P(Game = On | Transport Disruption, Rain) = 0.26

The dilemma occurs between the two decision options Cancelled and Postponed. If we now receive information about the unavailability of free slots then the network ranks the decision options as follows:

P(Game = Cancelled | Disruption, Rain, not Free Slot) = 0.38
P(Game = Postponed | Disruption, Rain, not Free Slot) = 0.34
P(Game = On | Disruption, Rain, not Free Slot) = 0.28

Based on the above probability distribution, the decision maker may decide to commit to the decision option Cancelled.

5 An Example

We present here a concrete example illustrating the proposed argumentation based decision-making process and belief network based aggregation.

Suppose the current world consists of the sentences in the syntax of L_{Arg}, shown in Figure 6 and Figure 7, obtained by translating the specification of the `game_state` decision, shown in Figure 4. In addition, we consider the following set of beliefs and knowledge (knowledge is defined as $F \wedge \langle bel \rangle F$) as part of the decision maker's knowledge base at w0:

$$\{rain, \langle bel \rangle transport\_disruption\}$$

We cannot uniquely define the valuation on as the set of formulae that characterize if it contains assertions that are only believed, such as $\langle bel \rangle transport\_disruption$. An example valuation S on w_0 is the following:

$$S = \{rain, transport\_disruption, cancelled\}$$

Since there are 3 candidates in the `game_state` decision (*on*, *cancelled*, and *postponed*) and we are dealing with probabilistic arguments, these three options will be considered mutually exclusive and exhaustive (which is not the case in general) for the purpose of aggregation:

$$C1 = on, C2 = cancelled, C3 = postponed$$

Consequently, there will be 3 possible worlds w_1, w_2, and w_3, whose valuations are as follows (see figure):

$$V(w_1) = S \cup \{on, determine\_game\_status\}$$
$$V(w_2) = S \cup \{cancelled, determine\_game\_status\}$$
$$V(w_3) = S \cup \{postponed, determine\_game\_status\}$$

Note that the presence of in the knowledge base along with the argument

$$\langle bel \rangle transport\_disruption \rightarrow \langle sup_{0.7} \rangle cancelled$$

derives $\langle sup_{0.7} \rangle cancelled$ from the knowledge base. Now the argument $\langle bel \rangle \neg rain \rightarrow \langle sup_{0.95} \rangle on$ states that P(On | not Rain) = 0.95. But we have $rain$ in the knowledge base and our implicit assumption is P(On | Rain) = 0.5. Therefore, $\langle sup_{0.5} \rangle on$ can be derived from the knowledge base.

The relations R_b and R_s in the model definition are defined as follows:

$$R_b = \{\langle w_0, w_1 \rangle, \langle w_0, w_2 \rangle, \langle w_0, w_3 \rangle\}$$
$$R_s = \{\langle w_0, 0.95, \{w_2\} \rangle, \langle w_0, 0.5, \{w_1\} \rangle\}$$

Note that *determine_game_status* is true in each of the possible worlds and therefore this is a goal - since the set of goal worlds is a subset of the set of possible worlds. This corresponds to the provability of $\langle goal \rangle determine\_game\_status$ in the current world using $\langle bel \rangle transport\_disruption$ in conjunction with the formula

$$\langle bel \rangle transport\_disruption \rightarrow \langle goal \rangle determine\_game\_status$$

The goal is active in w_0 since game status is not yet determined or *determine_game_status* is not yet believed. We are assuming here that the L_{Arg} theorem prover is able to derive the negation of $\langle bel\rangle determine\_game\_status$ from the current world by a mechanism similar to negation by failure. Belief network based aggregation process (as described in the last section) computes the supports for the candidates C1, C2, and C3 as follows:

Total support for: C1 = 0.26, C2 = 0.37, C3 = 0.37

The preference relation $\ll$ among the set of possible worlds is derived as $w_1 \ll w_2$ and $w_1 \ll w_3$. The maximally preferred possible worlds are w_2 and w_3. The relation R_g in the model definition is now defined as follows (Figure 15):

$$R_g = \{\langle w_0, w_1\rangle, \langle w_0, w_3\rangle\}$$

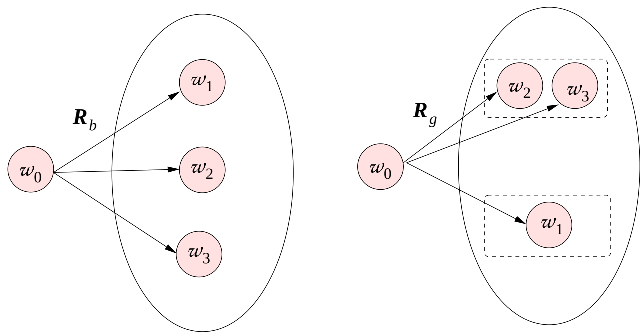


Figure 15. Relations between the current and possible worlds

This produces a dilemma. If the decision maker cannot gather any more evidence it may commit to w_2 by preferring w_2 to w_3. This involves adding the beliefs cancelled, not on, and not postponed to the current state of the database depending on the strength of support for them. In the new situation the goal to determine the status of the game will no longer be active, as *determine_game_status* it will be believed due to the presence of

$$\langle bel\rangle(cancelled \wedge \neg on \wedge \neg postponed) \rightarrow \langle bel\rangle determine\_game\_status$$

and the beliefs in *cancelled*, $\neg on$, and $\neg$*postponed*. Alternatively, if additional evidence is available to the decision-maker about the hosting club's financial situation, say $\langle bel\rangle\neg bad\_economy$, that will increase the total support for C1 as follows:

Total support for: C1 = 0.26, C2 = 0.41, C3 = 0.33

The revised valuation on each w_i will be as before except the additional evidence $\neg bad\_economy$ changes its truth value. The relations R_s and R_g may be redefined as follows:

$$R_s = \{\langle w_0, 0.95, \{w_2\}\rangle, \langle w_0, 0.5, \{w_1\}\rangle, \langle w_0, 0.5, \{w_2\}\rangle\}$$
$$R_g = \{\langle w_0, w_2\rangle\}$$

Since w_2 is the only goal world, the decision-maker considers w_2 as the committed world. Changing to the committed world from the current world involves adding $\langle bel\rangle cancelled$ and $\langle bel\rangle\neg on$, $\langle bel\rangle\neg postponed$ to the database as the decision-maker's beliefs. Adding $\langle bel\rangle cancelled$ to the database will trigger the decision for alternative activity (shown in Figure 5) and the decision making process continues as before.

6 Conclusion

In this paper, we have presented L_{Arg}, a logic for reasoning with probabilistic arguments, along with an approach for aggregating arguments via Bayesian belief networks. The semantics of L_{Arg} is given by enhancing the traditional possible world semantics with a new accessibility relation for support, and the soundness and completeness result is established. In the future, we plan to deal with more general forms of arguments than just propositional sentences, and enhance our proposed aggregation algorithm to aggregate temporal arguments via dynamic belief networks.

ACKNOWLEDGEMENTS

The author would like to thank Dave Lawless and John Fox for many helpful discussions on the topic, and the two anonymous reviewers for their constructive criticisms, that have helped to improve the paper enormously.

REFERENCES

[1] Chellas, B. 1980. "Modal Logic." Cambridge, U.K.: Cambridge University Press.

[2] Cohen, P. R. and Levesque, H. J. (1990). "Intention is choice with commitment." *Artificial Intelligence*, Vol. 42, 13-361.

[3] Conati, C., Gertner, A., van Lehn, K., and Druzdzel, M. (1997). "On-Line Student Modeling for Coached Problem Solving Using Bayesian Networks." *Proceedings of the Sixth International Conference on User Modeling*.

[4] Das, S., Fox, J., Elsdon, D., and Hammond, P. (1997). "A flexible architecture for autonomous agents", *Journal of Experimental and Theoretical AI*, 9(4): 407-440.

[5] Das, S. and Grecu, D. (2000). "COGENT: Cognitive agent to amplify human perception and cognition." *Proceedings of the 4th Int. Conf. On Autonomous Agents*, Barcelona, June.

[6] Fagin, R. 1988. "Belief, Awareness, and Limited Reasoning." *Artificial Intelligence*, 34(1):39-76.

[7] Fox, J. and Das, S. K. "Safe and Sound: Artificial Intelligence in Hazardous Applications," AAAI-MIT Press, June 2000.

[8] Fox, J., Krause, P, and Ambler, S. (1992). "Arguments, contradictions, and practical reasoning." *Proceedings of the Tenth European Conference on Artificial Intelligence*, Vienna, August, pp. 623-626.

[9] Heckerman, D. (1996). "A tutorial on learning with Bayesian networks." Microsoft Technical Report MSR-TR-95-06.

[10] Hintikka, J. 1962. "Knowledge and Belief." Ithaca, N.Y.: Cornell University Press.

[11] Jensen, F.V. (1996). "An Introduction to Bayesian Networks." Springer-Verlag.

[12] Jitnah, N., Zukerman, I., McConachy, R., and George, S. (2000). "Towards the generation of rebuttals in a Bayesian argumentation system." *Proceedings of the 1st Int. Natural Language Generation Conf.*, pp. 39-46.

[13] Korver, M. and Lucas, P. (1993). "Converting a rule-based expert system into belief network." *Medical Informatica*, Vol. 18(3), pp. 219-241.

[14] Krause, P. J. (1998). "Learning probabilistic networks." *The Knowledge Engineering Review*, Vol. 13:4, pp. 321-325

[15] Lauritzen, S. L. and D. J. Spiegelhalter (1988). "Local computations with probabilities on graphical structures and their applications to expert systems." *Journal of the Royal Statistical Society*, B 50 (2), pp.154-227.

[16] Lemmon, E. J. 1977. "An Introduction to Modal Logic." Basil, U.K.: Blackwell.

[17] Meyer, J.-J., and Vreeswijk, G. A. (1991). "Epistemic Logic for Computer Science: A Tutorial. Part 1." *Bulletin of European Association for Theoretical Computer Science* (EATCS) 44(4): 242-270.

[18] Pearl, J. (1988). "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference." San Mateo, CA, Morgan Kaufmann.

[19] Ramoni, M. and Sebastiani, P. (1997). "Learning Bayesian networks from incomplete databases." Technical Report KMI-TR-43, The Open University, UK.

[20] Rasmussen, J. (1983). "Skills, Rules and Knowledge: Signals, Signs and Symbolism, and Other Distinctions in Human Performance Models." *IEEE Transactions on Systems, Man, and Cybernetics*, 12: 257-266.

[21] Toulmin, S. 1956. The Uses of Argument. Cambridge, U.K.: Cambridge University Press.

Educational Human-computer Debate: a Computational Dialectics Approach

Tangming Yuan and **David Moore** and **Alec Grierson** [1]

Abstract. Theories of learning suggest that dialogue is important in shaping conceptual development. However, there is widespread debate as to the forms of dialogue and which are effective in an educational context. In addressing these issues, we have analysed current knowledge concerning dialectics in philosophy and education. We propose to adopt a computational dialectical approach to study the issues related to the development of an intelligent debating system, which is argued to have potential educational benefit. This approach focuses on using models of dialogue developed in the area of informal logic, which prescribe rules to regulate the evolving dialogue. Our proposed research concerns three main issues in the area of computational dialectics: dialogue model, debating heuristic theory and dialectical relevance.

1 Introduction

The recent development of Computer Based Learning Systems and the emergence of the World Wide Web and the Internet have changed the study life of many people. However, the usual assumption underlying these computer based educational systems is that the computer does all the informing, the student being merely a passive receiver of the information. The type of teaching interaction, that is, may become unduly didactic [13]. There is therefore a need for dialogue within interactive computer systems. Further, theories of learning have long suggested that dialogue has an important role to play in shaping conceptual change and developing reasoning skills [18]. There are many different uses of dialogue in an educational context. For example, Grasso et al.'s [5] "Daphne", a computational agent conducts an advice giving dialogue with the user to provide healthy nutrition education. Maudet and Moore's [10] human computer debate prototype will enable a student and computer to conduct a fair and reasonable debate on a controversial issue. Ravenscroft and Matheson [17] introduce two kinds of asymmetric dialogues to support learning. One is the computer being a "facilitating tutor" and the student the "explainer": the tutor raises some questions, students answer the questions, and the tutor solves the contradictions of the student's commitments and helps the students to reach the correct answer rather than directly tell them. Ravenscroft and Matheson's second dialogue type is similar to the first, but includes further didactic features. Bench-Capon et al. [3] investigated the computer mediated dialogue in legal educational context, which is explanation based, both participants adopting symmetric roles [2]. Pilkington's study of simulation-based learning identified two types of dialogue, an inquiry dialogue with asymmetric roles and a more collaborative game generating cognitive conflict and reflection ([15], [16]). However, there is widespread debate as to the forms of dialogue in general and which are effective in educational contexts in particular. We therefore review two approaches to characterising dialogue types, that of Walton and Krabbe [21] and Baker [1], and then, we make a proposal for human computer debate using a dialectical approach.

[1] School of computing, Leeds Metropolitan University, Leeds, LS6 3QS, United Kingdom

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

2 Dialogue Typology

2.1 Walton and Krabbe's typology

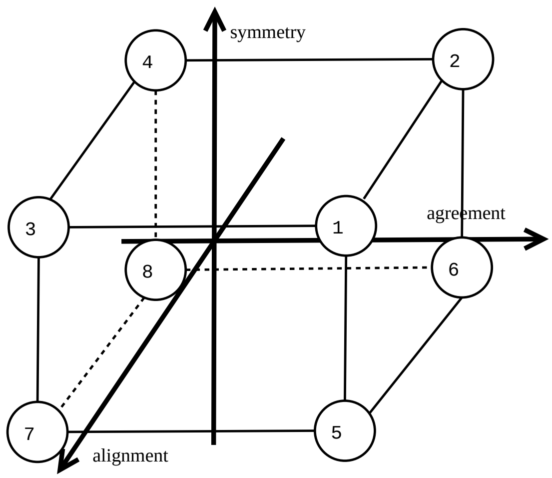
Type of dialogue	Initial situation	Participant's goal	Goal of dialogue
Persuasion	Conflict of opinion	Persuade other party	Resolve or clarify issue
Inquiry	Needs to have proof	Find and verify evidence	Prove (disprove)
Negotiation	Conflict of interest	Get what you most want	Reasonable settlement that both can live with
Information-seeking	Need information	Acquire or give information	Exchange information
Deliberation	Dilemma or practical choice	Co-ordinate goals and actions	Decide best available course of action
Eristic	Personnel conflict	Verbally hit out at opponent	Reveal deeper basis of conflict

Figure 1. Walton and Krabbe's dialogue typology

The most influential dialogue typology is probably Walton and Krabbe's [21] dialogue model developed in the area of argumentation theory. This model provides a broad typology of dialogue types and their rationale. It is based on three factors: "(i) the initial situation, (ii) the private aims of the participating agent, (iii) the joint aims to which all participants implicitly subscribe". Six dialogue types are included in this model: persuasion, negotiation, inquiry, deliberation, information seeking and eristic. See figure 1 (citing from [21]). Reed examined the above dialogue model in some depth in agent communication research [19]. He suggests that 'eristic' dialogue is unlikely to play a significant role in current computer science research. He also suggests that persuasion, inquiry and information-seeking dialogues handle belief, while negotiation dialogue raises a contract and deliberation dialogue forms a plan. He further notes that information-seeking dialogue is asymmetric. According to [8], only persuasive, negotiation and eristic dialogue are argumentative, but deliberation, inquiry and information seeking are seen as non-argumentative, although reasoning is believed to occur in all of them.

2.2 Baker's typology

Baker's problem solving model claims that there are eight basic forms of interactions in co-operative problem solving activity in learning situations, see figure 2 (citing from [1] p131).



1. co-construction
2. apparent co-construction
3. co-argumentation
4. apparent co-argumentation
5. acquiescent co-elaboration
6. apparent acquiescent co-elaboration
7. one-side argumentation
8. apparent one-side argumentation

Figure 2. Baker's dialogue model

Dialogue type	Initial situation	(non)- collaborative	(a) symmetry	Examples
Co-argumentation	conflict	collaborative	symmetric	Negotiation [14], [19]
one-side co-argumentation	conflict	collaborative	asymmetric	Auction or bid
argumentation	conflict	non-collaborative	symmetric	Debate [10] Complex critical discussion [20] Symmetric persuasion [21]
one-side argumentation	conflict	non- collaborative	asymmetric	Asymmetric persuasion [21] Simple critical discussion [20]
co-construction	ignorance	collaborative	symmetric	Deliberation, inquiry [21] Discovery [11]
one-side co-construction	ignorance	collaborative	asymmetric	Facilitating dialogue[18]
information-exchange	ignorance	non- collaborative	symmetric	Information-exchange [6]
Information seeking	ignorance	non- collaborative	asymmetric	Information-seeking [21]

Figure 3. Integrated dialogue typology

Baker's model is based on three dimensions: [1] degree of (dis) agreement, [2] degree of (a) symmetry, [3] degree of alignment. Baker's explanation of the degree of (a) symmetry is "either each participant has an alternative proposal, or else one participant simply contests another's proposal" [**?**]. In a computational context, "symmetry" is often taken to suggest that each participant makes more or less equal contributions to the dialogue and follows the same dialogue rules, while "asymmetric" suggests that participants play different roles in dialogue and follow different dialogue rules [10]. For example one participant simply contests or acquiesces to another's proposal [1]. Baker's notion of "alignment" is the same as 'collaborative', which means the desired end goals are the same for both players, while non-collaborative means they do not have identical end goals [10].

2.3 Integration of the two dialogue typologies

Walton and Krabbe's identification focuses on the philosophical study of dialogue, whereas Baker's model is based on co-operative problem solving activity in learning situations. Walton and Krabbe admit the incompleteness of their identification. Actually, some existing educational dialogues are outside Walton and Krabbe's dialogue typology. For example Ravenscroft and Matheson's two kinds of asymmetric dialogues [17], and Pilkington and Mallen's inquiry dialogue with asymmetric roles[16]. Further, in agent communication research, McBurney and Parsons identify two kinds of dialogue: discovery and command dialogue [11], which are outside Walton and Krabbe's dialogue typology [21]. It might be thought that Baker's model is more general and can subsume Walton and Krabbe's. However, some dialogue types can not be distinguished by Baker's model, for example, Walton and Krabbe's deliberation and inquiry dialogues both fall into one category (co-construction dialogue) of Baker's. Therefore, we integrate Walton and Krabbe's and Baker's dialogue typology, form a broad dialogue typology based on initial situation, collaboration and symmetry (i.e., three dimensions).

2.3.1 Co-argumentation dialogues and one-side co-argumentation dialogue

Co-argumentation dialogues start from conflict, but both participants' aims are identical, with symmetric roles. Examples such as negotiation can be seen in [19] and [14]. The difference between one-side co-argumentative dialogue and co-argumentative dialogue is that the participants of one-side co-argumentative dialogue adopt asymmetric roles, for example auction or bid. The following dialogue shows an example of a one-side co-argumentative dialogue interaction (B: buyer, S: seller).

B: how much is the Chinese leaf? (information seeking)
S: two pounds.
B: it is too expensive, how about one pound? (negotiation)
S: no, it is not expensive. (unsatisfied with the price)
B: it is raining, if you do not sell, it may go bad, how about 1.2 pound? (active negotiation)
S: no (still unsatisfied).
B: 1.5 pounds? (active toward the deal)
S: ok (deal).

It is worth noting that the buyer and seller adopt different roles in negotiation dialogue, the buyer actively negotiates, while the seller just contests rather than actively negotiates, until the end of the dialogue.

2.3.2 Argumentation and one-side argumentation dialogue

Argumentation dialogue starts from conflicts, but both sides attempt to persuade the other to accept their thesis, e.g Maudet and Moore's [10] debating dialogue, Van Eemeren et al.'s [20] complex critical discussion, and Walton and Krabbe's [21] permissive persuasion dialogue (PPD). One-side argumentative dialogue has different roles for both participants, one side builds its position, the other side attacks or contests, e.g Walton and Krabbe's rigorous persuasion dialogue (RPD) [21].

2.3.3 Co-construction dialogue and one-side co-construction dialogue

Co-construction dialogue starts from an open problem or question, two participants contribute more or less equally to solve the problem e.g. McBurney and Parsons's [11] discovery dialogue. It is interesting that Walton and Krabbe's [21] deliberation and inquiry dialogue all fall into this category. The participants of one-side co-construction dialogue have different roles, one side provides the solution, the other side may criticise or point out mistakes, but both parties have identical goals to solve the problem e.g. Ravenscroft and Pilkington's [18] facilitating dialogue.

2.3.4 Information exchange and information seeking dialogue

Such dialogue does not start from conflict. The participants have different dialogue roles and obligations, one side lacks information, the other side provides information, hence the dialogue is asymmetric in nature (cf. Hamblin's information-oriented dialogue [6]). Given this dialogue typology, our question becomes which of the diverse dialogue types are effective in educational contexts. Answers to this question gained from empirical research have yet been only partial [18]. However, the debating style of dialogue interaction is argued by Maudet and Moore [10] to be important in critical thinking and developing debating and reasoning skills, and also suggested by Pilkington and Mallen's [16] educational discourse analysis to be effective and to have rich educational benefit. A particular concern with our research therefore is to investigate issues surrounding a computer based system for educational debate.

3 A Proposal for Human-Computer Debate

There are at least two main areas of research dealing with dialogue: linguistic discourse analysis and dialectics. The former approach emphasises empirical research into natural language, its structure and processing and concerns actual conversational exchange, but there are well known difficulties in the application of such an intentional account to make dialogue computationally tractable. The latter approach - dialectics - involves a logical account of interaction in terms of rules for particular kinds of responses and interaction, and utilises ”Dialogue Game Theory” models developed within the field of Informal Logic to prescribe how dialogue should be regulated. There is an increasing use of a computational dialectics approach in the area of human computer interaction (e.g. [5]), agent communication (e. g. [7]), mediation of legal reasoning (e. g. [2]) and Artificial Intelligence in general [22]. In some literature, computational dialectics is seen as a new sub-field of Artificial Intelligence [4]. There are, however, many open research issues within computational dialectics, and an investigation of what are believed to be the most important in adopting the computational dialectical approach to develop a human computer debating system will form the basis of this research. Previous research in this application area ([12], [10]) has revealed several important issues that need further investigation.

3.1 Dialogue model

The most important issue concerns the choice or development of a suitable dialectical model. This is fundamental, because it forms the dialogue model that the computer system will use to rule as to the acceptability of user input and to delineate possible dialogue contributions it can make. The dialogue model is therefore the fundamental element underlying the proposed computer debate system. There are however many normative dialogue game systems that have been proposed in the area of informal logic and dialectics [10]. It is necessary therefore to select or develop a suitable dialectical model given the pre-requisites for a competitive human-computer debate on controversial issues such as capital punishment. Next, the appropriateness of the dialectical model needs to be established. The proposed experimental work required for this, aimed at iteratively building a computational realisation of the model and establishing whether the model can be readily assimilated and used to generate good discourse, will form part of the unique contribution of this research. It is anticipated that this part of the work will contribute towards developments in human computer dialogue and also help to illuminate research issues in the field of dialectic itself.

3.2 Debating strategic heuristics

In dialectical systems, the dialogue regulations usually leave some room for choices as to permissible move type and substantive content [12]. It is crucial therefore that the computer has some means of selecting between the available possibilities. This choice must be based on some suitable strategy, and the research will therefore seek to develop a theory of debating heuristics usable by the debating system. A dialogue strategy is a set of moves designed to cumulate in the achievement of one's objective in the dialogue game. A strategic heuristic in a dialogue game can be seen as a decision about what to do next and may involve forms of argument such as argument from analogy, argument from popularity and argument from consequence. Suitable computational strategies are currently not known, but are essential if the computer is to produce high quality dialogue contributions. To determine the appropriateness of strategies generated by the theory, further technical and user studies will be required, aimed at testing whether the strategy is effective. Analysis of results will illuminate the theory of debating heuristics and hence make a major contribution to the field of computational dialectics.

3.3 Dialectical relevance

A related problem for dialectical systems is that no rule actually controls the relevance of the dialogue moves [9]. Without relevance rules to govern the dialogue, however, it may lose focus, e.g. if the student inputs an irrelevant move, then a computer system without a relevance ruling will follow the student into an irrelevant dialogue. Given the importance of relevance in dialectical system, existing literature concerning the notion of relevance (e. g. [23]) will be investigated and used to derive relevance measures for use within the computer debating system. Further experimental work will then be conducted, aimed at testing the effectiveness of the proposed measures. The research will therefore contribute to our knowledge of how to create more useful dialectical models.

4 Conclusion

We have reviewed two key philosophical and educational dialogue typologies, proposed a broad dialogue typology and argued that debating style dialogue is potentially effective in critical thinking and development of student's debating skills (cf. [12], [16]). A proposal is made to research issues in building an intelligent debating system using a computational dialectical approach. Three important issues are discussed and proposed for further research.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] M. Baker, 'The role and models in Artificial intelligence and Education research: a prospective view', *International Journal of Artificial Intelligence in Education*, **11**, 122–143, (2000).

[2] T.J.M. Bench-Capon, 'Specification and implementation of Toulmin Dialogue Game', in *Proceedings of JURIX 98*, GNI, Nijmegen, pp 5-20, (1998).

[3] T.J.M. Bench-Capon, Leng P. H. and Standford, G., . 'A computer supported environment for the teaching of Legal argument', *Journal of Information, Law, and technology (JILT)*, **3**, `http://www.law.warwick.ac.uk/jilt/98-3/capon.htm`, (1998).

[4] T. Gordon, 'Computational Dialectics'. In P Hoschka (ed.). *Computers as Assistants, A new Generation of Support System*, New Jersey: Lawrence Erlbaum Associates, (1996).

[5] F. Grasso, Cawsey, A. and Jones, R. 'Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition', *International Journal of human computer studies*, **53**, 1077–1115, (2000).

[6] C.L. Hamblin, 'Mathematical models of dialogue', *Theoria*, 130–155, (1971).

[7] N.R. Jennings, 'Commitment and Conventions: The foundation of coordination in Multi-Agent Systems'. *The Knowledge Engineering Review*, **8**, 223–250 (1993).

[8] E. Krabbe, E. 'Symposium on Argument and Computation. Argument and Computation: Position Paper' (2000).

[9] N. Maudet, 'Notes on relevance in dialectical systems'. in *Proceeding of the ECSQARU'2001 Workshop-"Adventures in Argumentation"*, Toulouse, September 2001.

[10] N. Maudet, and Moore, D. (2002). 'Dialogue games as Dialogue Models for Interacting with, and via, Computers'. In Press for *Informal Logic*.

[11] P. McBurney, and Parsons, S. Agent Ludens: Games for Agent Dialogues, in *AAAI Spring Symposium on: Game Theoretic and Decision Theoretic Agents (GTDT2001)*, 2001.

[12] D.J. Moore, *Dialogue game theory for intelligent tutoring systems*, Unpublished doctoral dissertation, Leeds Metropolitan University. (1993)

[13] D.J. Moore, 'A framework for using multimedia within argumentation systems', *Journal of Educational Multimedia and Hypermedia*, **9(2)**, 83–98, (2000).

[14] S. Parsons, and Jennings, N. R. 'Negotiation Through Argumentation-A Preliminary Report', in *Proceedings of the International Conference on Multi-agent Systems*, Kyoto, Japan, 267–274 (1996).

[15] R.M. Pilkington, 'Dialogue games in support of qualitative reasoning', *Journal of Computer Assisted Learning*, **14**, 308–320 (1998).

[16] R.M. Pilkington, and Mallen, C. 'Dialogue Games to Support Reasoning and Reflection in Diagnostic Tasks', In P. Brna, A. Paiva and J. Self (Eds.) *Proceedings of the European Conference on Artificial Intelligence in Education*, September 1996. Lisbon, Portugal: Fundacao Calouste Gulbenkian (1996).

[17] A. Ravenscroft, and Matheson 'Developing and evaluating dialogue games for collaborative e-learning', *Journal of Computer Assisted Learning*, **18** 93–101 (2002).

[18] Ravenscroft, A. and Pilkington, R. M. 'Investigate by Design: Dialogue Models to Support Reasoning and Conceptual Change', *International Journal of Artificial Intelligence in Education*, **11**, 237–298, (2000).

[19] C. Reed, (1998). 'Dialogue Frames in Agent Communication', *Proceeding of the Third International Conference on Multi-Agent Systems*, IEEE Press, 246–253 (1998).

[20] F.H. van Eemeren, Grootendorst, R and Henkemans, F. S. *Fundamentals of Argumentation Theory, A Handbook of Historical Background and Contemporary Development*. Lawrence Erlbaum Associates Publishers (1996).

[21] D. Walton, and Krabbe, E. *Commitment in Dialogue: Basic concept of interpersonal reasoning*. Albany NY: State University of New York Press (1995).

[22] D. Walton, *The New Dialectics: Conversational Context of Argument*. University of Toronto Press. (1998).

[23] D. Walton, D. 'Dialectical Relevance in Persuasion Dialogue', *Informal Logic*, **19(2&3)**, 119–143 (1999).

Session II

Argumentation Schemes and Defeasible Inferences

Doug N. Walton[1] and **Chris A. Reed**[2]

1 Introduction

Argumentation schemes are argument forms that represent inferential structures of arguments used in everyday discourse, and in special contexts like legal argumentation, scientific argumentation, and especially in AI. Deductive forms of inference like *modus ponens* and disjunctive syllogism are very familiar. But some of the most common and interesting argumentation schemes are neither deductive nor inductive, but defeasible and presumptive. You may not be familiar with these. To introduce them, some background may be useful.

Perelman and Olbrechts-Tyteca, in *The New Rhetoric* (1969) identified and defined many distinctive kinds of arguments used to convince a respondent on a provisional basis. Arthur Hastings' Ph.D. thesis (1963) made an even more systematic taxonomy by listing many of these schemes, along with useful examples of them. Hastings presented a form for each scheme, and a set of critical questions matching the form of argument. In each instance, Hastings presented one premise of the form (scheme) as a conditional or generalization expressed as a Toulmin warrant. These features turned out to be very significant in the subsequent development of argumentation schemes. Many argumentation schemes are mentioned or described in the work of van Eemeren and Grootendorst (1984; 1992). Kienpointner (1992) has developed a comprehensive account of argumentation schemes that includes deductive and inductive ones as well as presumptive ones. A list of presumptive argumentation schemes given in (Walton, 1996) is not complete, and the analysis of each scheme is still in rough form. But this list identifies many most common forms of defeasible argumentation. In some important respects, the treatment of schemes follows Hastings' style, especially in having with a set of critical questions matching each form. The latest development is that argumentation schemes are being handled and represented in Araucaria to help with argument diagramming.

But the history of the study of these presumptive argumentation schemes is ancient. Many of these forms of argument were identified and discussed by Aristotle in three of his books especially, *Topics, On Sophistical Refutations and Rhetoric*. Aristotle called these forms of argument "topics" (topoi) or places. Warnick (2000, pp. 120-128) drew up a detailed table comparing twenty-eight topics identified in Aristotle's *Rhetoric* to thirteen of the argumentation schemes in Perelman and Olbrechts-Tyteca. The traditional problem with topics is that it seemed hard for commentators to appreciate what role the topics were supposed to have. Perhaps because of the dominance of deductive logic, the role of the topics seemed obscure. What has been taken to be their most useful purpose is to help a speaker think up new arguments to support rhetorical presentation in a speech. In medieval logic, topics were also sometimes taken to be useful for the purpose of testing the inferential link between a set of premises and a conclusion. But this use never really caught on. The topics had some appeal in rhetoric from time to time, but were never much of a useful tool there. In logic, topics remained marginal.

[1] Department of Philosophy, University of Winnipeg, Canada
[2] Department of Applied Computing, University of Dundee, UK

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

2 Examples of Schemes

For those who are not familiar with argumentation schemes it is good to examine a few examples. Argument from position to know is based on the assumption by one party that another party has information that the first party needs. For example someone lost in a foreign city asks a stranger where the Central Station is. The questioner needs this information, and does not have it. If the respondent gives and answer by citing a location, what reason does the questioner have to think that she can act on this information, or take it as true? The rationale is given by argument from position to know. The version of the argumentation scheme in (Walton, 1996, pp. 61-63) is given below.

Argument from Position to Know (Version I)

Major Premise: Source a is in a position to know about things in a certain subject domain S containing proposition A.
Minor Premise: a asserts that A (in Domain S) is true (false).
Conclusion: A is true (false).

When a proponent puts forward an argument in a dialogue and it meets the requirements indicated above, then it carries some weight as a presumption. But it is defeasible by questioning. Matching the argument from position to know are three critical questions (Walton, 1996, p. 62).

CQ1: Is a in a position to know whether A is true (false)?
CQ2: Is a an honest (trustworthy, reliable) source?
CQ3: Did a assert that A is true (false)?

When the proponent in a dialogue has put forward an argument from position to know, the respondent can ask any one of these three critical questions. Once the question has been asked the presumptive weight the argument had before is withdrawn. But if the proponent gives an acceptable answer to the question, the weight is restored.

Appeal expert opinion is a subtype of argument from position to know where one party has expert knowledge that the other wants to use. This scheme is represented in (Walton, 1997, p. 210) as follows.

Appeal to Expert Opinion (Version I)

Major Premise: Source E is an expert in subject domain S containing proposition A.
Minor Premise: E asserts that proposition A (in domain S) is true (false).
Conclusion: A may plausibly be taken to be true (false).

Appeal to expert opinion is a defeasible form of argument that should not be taken as beyond challenge. There is a natural tendency to respect an expert, and thus we find it hard to question the word of an expert. Still, appeal to expert opinion is best seen as subject to critical questioning. Six basic critical questions are proposed in (Walton, 1997, p. 223).

1. *Expertise Question:* How credible is E as an expert source?
2. *Field Question:* Is E an expert in the field that A is in?
3. *Opinion Question:* What did E assert that implies A?
4. *Trustworthiness Question:* Is E personally reliable as a source?
5. *Consistency Question:* Is A consistent with what other experts assert?
6. *Backup Evidence Question:* Is A's assertion based on evidence?

The two devices of the scheme and the critical questions work together. The scheme is used to identify the premises and conclusion. The critical questions are used to evaluate the argument by probing into its potentially weak points.

Many argumentation schemes are associated with traditional informal fallacies. Appeal to popular opinion is a separate scheme from argument from argument from position to know, but is often connected with it. But in many cases the two are connected. An example would be, "Everybody in Lyon says that the Metro is a good way to get around." This argument is an appeal to popular opinion but its worth is bolstered by the intertwined argument that people who live in Lyon are (presumably) in a position to know about such things.

Argumentum ad hominem, or use of personal attack to criticize somebody's argument, has several interconnected argumentation schemes associated with it. The circumstantial *ad hominem* is a subtype of argument from commitment. In law, circumstantial *ad hominem* arguments are used to raise doubt about the credibility of the witness by attacking his testimony as inconsistent. Several argumentation schemes have to do with meanings of words and phrases. One is argument from classification. Legal arguments are often about how something like a contract can be classified. Other schemes are based on definitions. One is to attack an argument from definition claiming that the definition is too vague.

The sunk costs argument, or argument from waste, as Perelman and Olbrechts-Tyteca called it, runs as follows. I have already sunk such an effort into trying to attain this goal, it would be wasteful for me to stop now. The sunk costs argument also seems to be a species of argument from commitment, as recognized by the growing literature on the notion of precommitment in the literature on decision making in economics and banking. Generally, the presumptive schemes represent types of argument that would be widely seen in AI as abductive. The scheme most closely related to abduction, however, is argument from sign.

As noted above, the schemes as formulated in (Walton, 1996) are in a rough form designed to be useful. They need more work to adopt some standard notation to put them in a consistent structure that could be useful for formalization and computing. For example, consider the two schemes above. They can be reformulated in a way that makes the structure of the inference in them more explicit. Consider argument from position to know first.

Argument from Position to Know(Version II)

Major Premise: Source a is in a position to know about things in a certain subject domain S containing proposition A.
Minor Premise: a asserts that A (in Domain S) is true (false).
Conditional Premise: If source a is in a position to know about things in a certain subject domain S containing proposition A, and a asserts that A is true (false), then A is true (false).
Conclusion: A is true (false).

In version II, the conditional premise plays a role comparable to the general premise in Hastings' formulation of schemes. In this formulation, as noted above, the premise was expressed as a Toulmin warrant. It is a defeasible rule that can default in the face of exceptions to the rule in a given case.

A reformulation of the appeal to expert opinion along the same Hastings-style lines is set out below.

Appeal to Expert Opinion (Version II)

Major Premise: Source E is an expert in subject domain S containing proposition A.
Minor Premise: E asserts that proposition A (in domain S) is true (false).
Conditional Premise: If source E is an expert in a subject domain S containing proposition A, and E asserts that proposition A is true (false), then A may plausibly be taken to be true (false).
Conclusion: A may plausibly be taken to be true (false).

Versions I and II of these schemes are not that different. Version II is a more explicit account of the structure of the inference that makes the warrant that the argument is based on more visible. But version II leads to a certain controversy that now needs to be discussed.

3 *Modus Ponens* and Schemes

The more explicit presentation of the presumptive argumentation schemes, revealing the warrant, often seems to come very close to assuming that inferences have the *modus ponens* form. But this seems inconsistent, because we all know that MP is deductively valid, and yet these presumptive schemes are not supposed to represent deductively valid forms of argument. Blair (1999, p. 341), as quoted in the sentence below, detected an inconsistency in the treatment of schemes in (Walton, 1996).

"(S)everal of the formulations of argumentation schemes (in Walton, 1996) represent valid argument forms, whereas Walton is quite explicit throughout the book that presumptive arguments are not deductive entailments."

As an example, Blair (p. 341) cited the argumentation scheme for appeal to popular opinion as formulated by Walton.

Appeal to Popular Opinion If a large majority (everyone, nearly everyone, etc.) accept A as true, then there exists a (defeasible) presumption in favor of A.

A large majority accept A as true.
Therefore, there exists a presumption in favor of A.

Blair found a contradiction here. He wrote (p. 341), "this scheme has the form of *modus ponens*." And then he wrote, "yet Walton says that this kind of argumentation is deductively invalid!" These comments suggest that there is much to be puzzled about with the account of argumentation schemes ventured in (Walton, 1996). We all know that *modus ponens* is a deductively valid form of argument, and thus that all arguments having the *modus ponens* form are deductively valid. So if presumptive argumentation schemes can be cast in the *modus ponens* form, the outcome seems to be a bad sort of contradiction that needs to be resolved. How can this problem be dealt with?

The problem can be addressed by drawing a distinction between two types of inference after a fashion proposed by Verheij (2000, p. 5).

Modus Ponens

Premises:

As a rule, if P then Q
P

Conclusion:

Q

Modus Non Excipiens

Premises:

As a rule, if P then Q
P
It is not the case that there is an exception to the rule that if P then Q

Conclusion:

Q

As far as terminology is concerned, we would like to call *modus non excipiens* defeasible *modus ponens*. The strict form can then just be called *modus ponens*. Or if the contrast needs to be emphasized, it could be called deductive *modus ponens* or strict *modus ponens*. This distinction, whatever terms you use to draw it, seems to address Blair's problem. But it poses another one. How can one tell in a given case whether a *modus ponens* argument is better formalized using the one form or the other? Verheij (2000, p. 5) proposed policies to enable us to distinguish between cases. But we won't pause on this more practical aspect of the problem. Each case needs to be dealt with individually to examine the claim presumably made by an arguer. Even if this practical problem can be solved, Blair's problem resurfaces in another guise by raising a general theoretical problem. It is a controversial issue that goes to the heart of applied logic.

The reason this issue is so controversial is that logic textbooks have become accustomed to telling students that all arguments having the *modus form* are deductively valid. This statement can be misleading however. It seems to suggest that even arguments of defeasible *modus ponens* form have to be deductively valid. It seems to make deductive logic all-encompassing. It the supposed applicability of deductive logic to arguments that, many of us would say, it doesn't properly apply to. This expansionist approach is evident in many of the standard logic textbooks. For example, in the very widely used textbook Introduction to Logic (Copi and Cohen, 1998, p. 363) the reader is told that the following argument has the *modus ponens* form, and is therefore deductively valid.

If he has a good lawyer then he will be acquitted.
He has a good lawyer.
Therefore he will be acquitted.

Copi and Cohen (p. 363) tell their readers that the first premise should be translated into symbolic form using the material conditional, and that the argument can then be proved to be valid using propositional logic. But is it deductively valid? The problem is that it could be true that you could have a good lawyer, but it could also be true that the other side has a better one. At this point Blair's problem resurfaces as the firestorm of controversy begins (to mix two metaphors). The deductivist camp will maintain that if you mean the first premise to be really true, then the argument can be seen as deductively valid. The problem with this approach is that deductive logic has been expanded so widely that seeing the above argument as having any inferential link or warrant is excluded. In particular this expansionist approach excludes the possibility of seeing the argument as having the defeasible *modus ponens* form. And so it excludes the possibility of using defeasible *modus ponens* as a resource for the study of argumentation schemes.

For those in the computing field, who are used to dealing with defeasible inferences, Blair's problem is easily circumvented. All we need to do is to recognize the distinction between strict and defeasible *modus ponens* and then classify the lawyers argument from Copi and Cohen as having the defeasible form. But those used to deductive logic as presented in the standard textbooks may not give up so easily. One of the issues which brings the two camps closer together is the need to diagram such arguments. Diagramming is of interest both to those in argumentation as a tool in the analytical toolbox, and to computer scientists as a precursor to implementable formalisation.

As explicit *modus ponens* arguments are so rare in everyday conversation (we return to this below), it is not often that one encounters diagrams of such arguments. Given that the conventional, deductive form of *modus ponens* relies on both its two premises, one appropriate diagram would be a linked structure as follows:

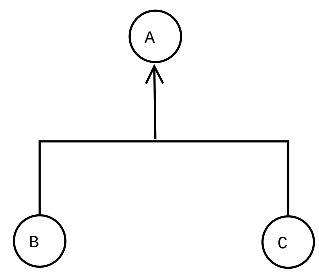


Figure 1. Linked structure diagram

Which maps on to the deductive *modus ponens* with A representing the conclusion Q, B representing the major premise *If P then Q*, and C the minor premise P. Of course, the diagram works equally well as an analysis of the Copi and Cohen argument:

A. He will be acquitted

B. He has a good lawyer

C. If he has a good lawyer then he will be acquitted

So, the apparent similarity in form is mirrored by similarity in diagramming. Yet, if the forms of *modus ponens* and *modus non excipiens* are to be distinguished, then the diagrammatic analysis too should be able to handle the difference.

The approach proposed and implemented in the Araucaria system (Reed and Rowe, 2001) is to mark instantiations of schemes explicitly. If we want to distinguish *modus ponens* and *modus non excipiens* by seeing the latter as a scheme, or if we want to indicate that the Copi and Cohen argument is an instantiation of a particular scheme, the diagram in Figure 2 would be appropriate.

Thus, the part of an argument covered by, or encapsulated in, an argumentation scheme is demarcated by a coloured area - which may then be labelled.

This approach has the benefit of providing a common diagramming technique for both deductivists and those advocating a somewhat smaller remit for deductive logic. In this approach to diagramming, the rich variety of real arguments can be catered for without needing a resolution to that discussion, and, further, it provides a starting point for formlisation of argument structure within computer science. At the moment, the structures in Figures 1 and 2 are constructed within the Araucaria software, and saved using an Argument

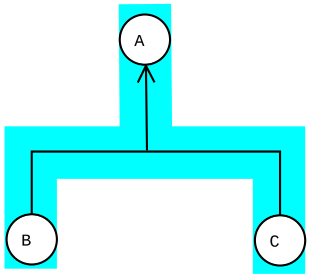


Figure 2. Argument scheme diagram

Markup Language (AML), based upon the industry standard XML approach. There are a range of benefits associated with using XML, but perhaps the most important here is that as an open standard, it supports a wide variety of different techniques for accessing and manipulating the data. Some of these techniques have applications, such as computer supported collaborative work and multi-agent systems communication, which lie squarely within computer science and for which closely defined, formal descriptions of argument are crucial.

4 The Completeness Problem for Argumentation Schemes

What could be called the completeness problem for argumentation schemes is expressed in the following question. When all the appropriate critical questions matching a scheme been answered satisfactorily, must the respondent then accept the argument? Or can he continue to ask critical questions? Or the question can put another way. When is a presumptive argument complete, meaning that if the respondent commits to the premises he must also commit to the conclusion? These questions ask how argumentation schemes are binding so to speak. Arguments based on presumptive schemes are not binding in the same way that a deductively valid is, or even in the same way that an inductively strong argument is. The respondent is only bound to tentatively accept the conclusion of a argument fitting a presumptive scheme, given that he accepts the premises of such an argument. Such arguments are plausible but inherently weak. Only when taken along with other arguments in a mass of evidence do they shift a balance of considerations.

It would be tempting to jump to the following hypothesis. Once all the critical questions matching a scheme have been satisfactorily answered, the argumentation is complete. But there is a problem with this hypothesis. It has been shown some schemes can have critical subquestions under each critical question. For example, the following three critical subquestions have been cited (Walton, 1997, p. 217) as coming under the trustworthiness critical question of the appeal to expert opinion.

Subquestion 1: Is E biased?
Subquestion 2: Is E honest?
Subquestion 3: Is E conscientious?

Bias, meaning failure represent both sides of an issue in a balanced way, is an important factor in evaluating appeal to expert opinion. Honesty is a matter of telling the truth, as the expert sees it. Conscientiousness is different from honesty, and refers to care in collecting sufficient information. Thus here we have three critical subquestions nested under the more general trustworthiness critical question matching version I of the appeal to expert opinion argumentation scheme above.

Suppose a respondent in a given case has asked all six of the basic critical questions corresponding to version I of the appeal to expert opinion scheme and the proponent has answered all of them adequately? Is the respondent now obliged to accept the appeal to expert opinion or can he continue to raise questions about it? We won't try to solve the completeness problem here, but will only suggest that a solution requires recognition of different levels on which critical questioning can take place in a dialogue. At one level, basic critical questions can be asked. At another level, critical subquestions of the basic questions can also be asked. Some authors, such as Gilbert (1991) suggest that this questioning can go on almost indefinitely. Presumptive arguments should always be regarded as open to critical questioning in a dialogue until the dialogue reaches the closing stage. Closure to asking of critical questions thus depends on the stage a dialogue is in.

5 Enthymemes

Invoking the authority of Aristotle, logic has traditonally used the term 'enthymeme' to mean an argument with missing (unstated) premises (or a conclusion). More and more evidence is showing that this meaning of 'enthymeme' is based on a misinterpretation of Aristotle's writings, beginning with the earliest commentators. Burnyeat (1994) has shown that Alexander of Aphrodisias may have been the first to put forward what became traditional view of enthymeme for two millenia. According to Burnyeat, what Aristotle really meant by 'enthymeme' is the plausibilistic type of arguments with a major premise expressing a generalizations that is not absolutely universal, but is defeasible. Such an argument may look like a syllogism with a premise containing what we now call a universal quantifier. But this appearance is misleading. This premise contains a generalization holds only "for the most part", to use Burnyeat's translation of Aristotle's expression. This new interpretation of Aristotle's writings on the enthymeme is quite exciting for those of us studying argumentation schemes. It suggests that the real Aristotelian enthymeme is the defeasible (presumptive) argumentation scheme of the kind described above.

Whatever you call it though, the problem of figuring out how to fill in missing premises or conclusions in a text of discourse is still there. It could be called the problem of incomplete arguments, or the problem of arguments with missing parts. It may seem a simple problem at first, but the many difficulties inherent in it have been shown. Such arguments are expressed in natural language, and a natural language text of discourse can be highly problematic to make sense of. Inserting premises that make an argument valid may misrepresent what the arguer meant to say (Burke, 1985; Gough and Tindale, 1985; Hitchcock, 1985). There is the ever-present danger of the straw man fallacy. This fallacy is the device of exaggerating or distorting an interpretation of an argument in order to make it look more extreme than it is, thereby making it easier to attack or refute it (Scriven, 1976, pp. 85-86). Examining these problems, it may appear the dream of creating an enthymeme machine, a mechanical device that automatically inserts missing premises or conclusions into an argument, is unachievable. Certainly creating such machine is a lot harder than it looks, given the difficulties in dealing with natural language argumentation.

An example taken from an exercise in Copi and Cohen (1994, p. 296) will illustrate some aspects of the problem. The reader is instructed to formulate the missing but understood premise or conclusion in the following enthyememes. One of these enthmemes is quoted below.

> Although these textbooks purport to be a universal guide to learning of great worth and importance - there is a single clue that points to another direction. In the six years I taught in city and country schools, no one ever stole a textbook.

The missing premise seems to be the statement, 'If people thought that these textbooks were a universal guide to learning of great worth and importance, they would steal them if given an opportunity. But the observation stated is that people do not tend to steal these textbooks when given an opportunity. The conclusion is that people do not think that these textbooks are a universal guide to learning of great worth and importance. This example brings out the point that an enthymeme can have an implicit premise that is a defeasible type of conditional. It is a type of conditional that is not absolute or strict. It would not support a deductively valid *modus ponens* argument. It presents us with a defeasible *modus ponens* argument. Of course there are enthymemes that can be reconstructed as *modus ponens* arguments or as syllogisms. But surely there are just as many, or perhaps even more, that can be better reconstructed as defeasible arguments.

REFERENCES

[1] Aristotle, *Topics*, trans. E. S. Forster, Loeb Classical Library, Cambridge, Mass., Harvard University Press, 1939.

[2] Aristotle, *On Sophistical Refutations*, trans. E. S. Forster, Loeb Classical Library, Cambridge, Mass., Harvard University Press, 1928. 11

[3] Aristotle, *The Art of Rhetoric*, trans. John Henry Freese, Loeb Classical Library, Cambridge, Mass., Harvard University Press, 1937.

[4] J. Anthony Blair, 'Review of (Walton, 1996)', *Argumentation*, 13, 1999, 338-343.

[5] Michael Burke, 'Unstated Premises', *Informal Logic*, 7, 1985, 107-118.

[6] Myles F. Burnyeat, 'Enthymeme: Aristotle on the Logic of Persuasion', *Aristotle's Rhetoric: Philosophical Essays*, ed. David J. Furley and Alexander Nehemas, Princeton, N. J., Princeton University Press, 1994, pp. 3-55.

[7] Irving M. Copi and Carl Cohen, *Introduction to Logic*, 10th ed., Upper Saddle River, Prentice Hall, 1998. Also cited is the 9th edition of 1994.

[8] Michael Gilbert, 'The Enthymeme Buster', *Informal Logic* 13(3), 1991, 159-166.

[9] James Gough and Christopher Tindale, 'Hidden or Missing Premises', *Informal Logic*, 7, 1985, 99-106.

[10] Arthur C. Hastings, *A Reformulation of the Modes of Reasoning in Argumentation*, Evanston, Illinois, Ph.D. Dissertation, 1963.

[11] David Hitchcock, 'Enthymematic Arguments', *Informal Logic*, 7, 1985, 83-97.

[12] Manfred Kienpointner, *Alltagslogik : Struktur und Funktion von Argumentationsmustern*, Stuttgart, Fromman-Holzboog, 1992.

[13] Chaim Perelman and Lucie Olbrechts-Tyteca, *The New Rhetoric*, Notre Dame, University of Notre Dame Press, 1969.

[14] Chris Reed and Glenn Rowe, *Araucaria: Software for Puzzles in Argument Diagramming and XML*, Technical Report, Dept. of Applied Computing, University of Dundee, 2001.

[15] Frans H. van Eemeren and Rob Grootendorst, *Speech Acts in Argumentative Discussions*, Dordrecht, Foris, 1984.

[16] Frans H. van Eemeren and Rob Grootendorst, *Argumentation, Communication and Fallacies*, Hillsdale, N. J. Erlbaum, 1992.

[17] Bart Verheij, 'Logic, Context and Valid Inference Or: Can There be a Logic of Law', 2000. Available on `bart.verheij@metajur.unimaas.nl`, `http://www.metajur.unimaas.nl/~bart/`

[18] Douglas Walton, *Argumentation Schemes for Presumptive Reasoning*, Mahwah, N. J., Erlbaum, 1996.

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

[20] Barbara Warnick, 'Two Systems of Invention: The Topics in the Rhetoric and The New Rhetoric', *Rereading Aristotle's Rhetoric*, ed. Alan G. Gross and Arthur E. Walzer, Carbondale, Southern Illinois University Press, 2000, 107-129.

Encoding Schemes for a Discourse Support System for Legal Argument

Henry Prakken and **Gerard Vreeswijk**[1]

Abstract. This paper reports on the ongoing development of a discourse support system for legal argument named PROSUPPORT. A description is given of the system's encoding schemes with which the user can enter his or her analysis of the discourse. These schemes, which are implemented as web browser forms linked to a database, serve to capture support relations of propositions within arguments, and dialectical relations between arguments. In addition, they support the recording of relevant argumentative and procedural speech acts made with respect to these arguments, such as disputing or conceding a claim, and allocating the burden of proof. The main issue in developing these encoding schemes is how expressiveness of the schemes can be reconciled with ease of use, on a suitable theoretical basis.

1 Introduction

In several related areas of computer science there is a growing interest in software support for such discourse processes as discussion, negotiation, dispute resolution and collective decision making. Unlike with 'conventional' decision-support tools (such as knowledge-based systems), the task of such systems is not to produce or suggest solutions to a problem with the help of domain knowledge, but to help the participants in discursive interactions to structure their reasoning and discourse, so that they can make sense of the discourse and interact effectively.

One professional area where such systems are of great potential use is the law. Participants in legal procedures (including alternative procedures such as online dispute resolution) often face the complex task of managing the information they are confronted with and the communication and reasoning they are expected to engage in. Discourse support systems can provide important assistance for these tasks: they could facilitate the structured inputting of a variety of discursive data, such as which claims have been made, conceded or challenged, how the burden of proof was assigned, which grounds and evidence have been adduced and counterattacked, how these grounds and evidence can be assessed, and whether the parties have respected the rules of procedure. The system could then usefully display, combine and restructure this input, and compute the consequences of the user's evaluative decisions (e.g. who wins given a certain allocation of the burden of proof and assessment of evidence?). Such systems could also support the (semi-) automatic generation of case summaries or even verdicts. These functionalities can be put to use in a variety of contexts. Individual users can be supported in making their own analysis of the discourse, invisible for other participants. The joint participants can be supported in their communicative and disputational interactions. Or the supporting staff of a judge or other official can be supported in their task to preprocess an analysis of a case, and to pass on the results to the official. Finally, in online versions of dispute resolution discourse systems could be a principal means of interaction between the participants.

In the field of AI & Law there is a growing body of theoretical research on discourse support for legal argument and legal procedure (e.g. [3, 1, 4, 12]). However, substantial research on architectures for implementation and on user experiences is still sparse. We know of only two systems that have been implemented with practical use in mind, viz. Loui's Room 5 system [8] and Verheij's ArguMed tool [18], and one further system that is currently being implemented, viz. Lodder & Huygen's support tool for online dispute resolution [7].

In other application areas, such as meeting support and intelligent tutoring, more practical experience with discourse support systems has been gained (see e.g. [9, 16, 15, 2]). These experiences raise important issues for legal discourse support systems. One of the main lessons learned is that it is very easy to overestimate the users' ability and willingness to learn a new codification scheme [15, 2]. The PROSUPPORT project, on which this paper reports, intends to take this lesson at heart. Its aim is to develop a discourse support system for legal procedure that provides useful computational power to the user but that is also easy to use.

Naturally, these two goals tend to conflict. The desire to offer useful computational power to the user requires that the user's input is structured as much as possible, in a way that reflects the essential elements of legal discourse. The more these elements are made explicit by the user, the more the system can do with it. However, the desire to make these elements explicit requires complex representation schemes for the user's input, which leads to a tension with the lessons on usability learned in other areas. Put simply, the more expressive a language, the harder it is to learn and use. Resolving this tension in an optimal way is one of the main research themes of the PROSUPPORT project. In other words, the project aims to discover conditions under which "formality" in interactive systems of the studied kind is helpful instead of harmful (cf. [15]).

To elaborate on the desired expressiveness, the following features of legal reasoning are especially relevant. Firstly, legal reasoning is adversarial, which means that arguments pro and con a claim are exchanged and conflicts between arguments must be resolved. Secondly, legal reasoning contains several specialised reasoning forms, such as combining rules and precedents, attacking the application of a rule, using and attack witness or expert evidence, reasoning about causation, and so on. Finally, legal reasoning takes place in a procedural context, where the notions of presumptions and burden of proof are essential, and where not only arguments but also other speech acts are important (such as disputing or conceding a claim and allocating the burden of proof).

There is another tension to be resolved. Being a research project, the system should have a sound theoretical basis, which means that it should be based on plausible theories of the structure and rationality of argumentative discourse. Moreover, since we are dealing with software specification, this theoretical basis should preferably be formal. The latter is particularly important since discourse support

[1] Institute of Information and Computing Sciences, Utrecht University, PO Box 80089, 3508 TB Utrecht, The Netherlands, email: {henry,gv}@cs.uu.nl

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

systems might be expected to compute the 'current state' of a dispute, given the arguments, counterarguments and priority arguments stated thus far. This requires a precise theory of what is to be computed. Now a problem is that most of the available theories are quite complex and subtle, especially when they are formalised. Therefore, directly implementing these theories would again detract from the usability of the system. A user can simply not be expected to master subtle theoretical notions and distinctions, let alone to deal with formal syntax or mathematical notions. Accordingly, a second research challenge of the PROSUPPORT project is to resolve the tension between naturalness and theoretical well-foundedness of the encoding schemes offered to the user.

This paper reports on our current proposals to resolve these two tensions, focusing on the encoding schemes for the user's input. The system is meant for Dutch civil procedure, and will be illustrated with an application to an actual Dutch civil case. It is important to note that in our design the interfaces for entering the user's input and for displaying the system's output are independent. Once information is inputted into the system, it is stored in an internal dataformat, which supports different ways of restructuring and visualising the information. This paper will not discuss interfaces for the latter.

As for the input encodng schemes, we propose a simple generic encoding scheme for argumentative and procedural speech acts. As for arguments, the scheme captures support relations between propositions within arguments and dialectical relations between arguments, but for the rest it imposes a minimum of structure on the user's input. We will show that this encoding scheme can be straightforwardly implemented as web browser forms linked with a database. Furthermore, we will argue that the design can be theoretically based on logics for defeasible argumentation and formal dialogue games for dispute resolution. Finally, we will discuss some limitations and possible extensions of our encoding schemes, and compare our proposals with related research.

2 The application domain

In this section we briefly describe Dutch civil procedure as far as relevant for present purposes. (This description is taken from [12] and inspired by [6]).

A civil law suit is divided into a 'pleadings' phase, where the adversaries plea their case before the judge and provide evidence when assigned the burden of proof by the judge, and a 'decision phase', where the judge withdraws to decide the case. The pleadings phase is separated into a written and an (optional) oral part. In the written part the parties exchange at least two and usually four documents (in fact, the law is about to be changed to make this "usually two"). The first is plaintiff's *Statement of Claim*, which has to contain plaintiff's claim plus his grounds for the claim. These grounds may be purely factual: plaintiff may leave out the legal 'warrant' connecting grounds and claim, as may both parties in all their other arguments. Also, parties do not need to explicitly state common-sense knowledge, and if they state such knowledge, they don't need to prove it. However, the judge decides what is common-sense knowledge. Defendant replies with her *Defence*, which has to contain all of defendant's attacks against plaintiff's claim and grounds. These attacks may also concern issues of procedure, so that the procedural legality of a move can itself become the subject of dispute. The adversaries may then exchange further documents as long as allowed by the judge. Each party may also ask to provide oral pleading. During the pleadings phase, the adversaries may dispute, concede and retract claims, defer to the judge's decision about a claim, support claims with arguments, move counterarguments, and offer to provide evidence for their claims. The judge assigns the burden of proof to a party whenever appropriate, after which that party must provide evidence (usually documents, or witness or expert testimonies). After the pleadings phase has ended, the judge gives his/her verdict, bound by the following rules of evidence.

An important principle of Dutch civil procedure is that the judge is passive with respect to the factual basis of the dispute. For instance, the judge must accept undisputed claims of the adversaries, and s/he must evaluate the evidence and give the verdict on the basis of the facts adduced by the parties, with the exceptions of generally known facts and legal rules. Of course, this does not mean that the judge cannot take factual decisions at all; s/he must still assess whether the facts adduced by the adversaries sufficiently support their claims, which may in turn also be factual.

As for allocating the burden of proof, the general rule is that the parties bear the burden of proving their claims; however, the judge may decide otherwise on the basis of special statutory provisions or on grounds of reasonableness. Among other things, this means that the burden of proof can be distributed over the parties, and that making a claim does not automatically create a burden to prove it; cf. [6, 11].

Given these characteristics of the procedure, our system should allow the following input. As for the adversaries, it should be possible to express which claims the adversaries have made, and which arguments they have stated in support of their claims or by way of counterargument. Furthermore, the system should keep track of which claims have been disputed, conceded, retracted or left to the judge's decision. Finally, the system should capture discussions on the procedural correctness of the adversaries' input (including admissibility of evidence). As for the judge, the system should record his/her decisions about such procedural correctness and about the burden of proof, including the judge's grounds for these decisions (when given). The system should also record the judge's completions of the adversaries' arguments with legal or commonsense knowledge. Finally, the system should allow for the inputting of any other argument moved by the judge, especially his/her assessments of evidence and conflicting arguments.

It is important to note that the PROSUPPORT system is not primarily meant to support the dispute as it actually takes place. Rather, the system is meant to support rational reconstructions of the dispute made by an individual user, either during or after the dispute. For instance, it could be used in the pleadings phase by one of the adversaries in preparing a further procedural document, or in the decision phase by the judge (or his assistants), in preparing the final verdict. It could also be used as an analysis tool by law students in a course on legal argumentation.

3 An example case

Throughout this paper we will use the following example case, concerning a dispute concerning ownership of a large holiday tent. Plaintiff (Nieborg) and his wife were friends of Van de Velde, who owned a large tent at a camp site. At some point van de Velde mentioned that the tent was for sale for dfl. 850. Nieborg replied that he was interested but could not afford the price. Van de Velde still made his tent available to Nieborg, who in return helped van de Velde to paint his house, while Mrs. Nieborg for some period assisted Mrs. van de Velde with her domestic work. At some stage, Nieborg claimed that they had done enough work to pay the sales price for the tent, after which van de Velde became very angry and demanded the tent back

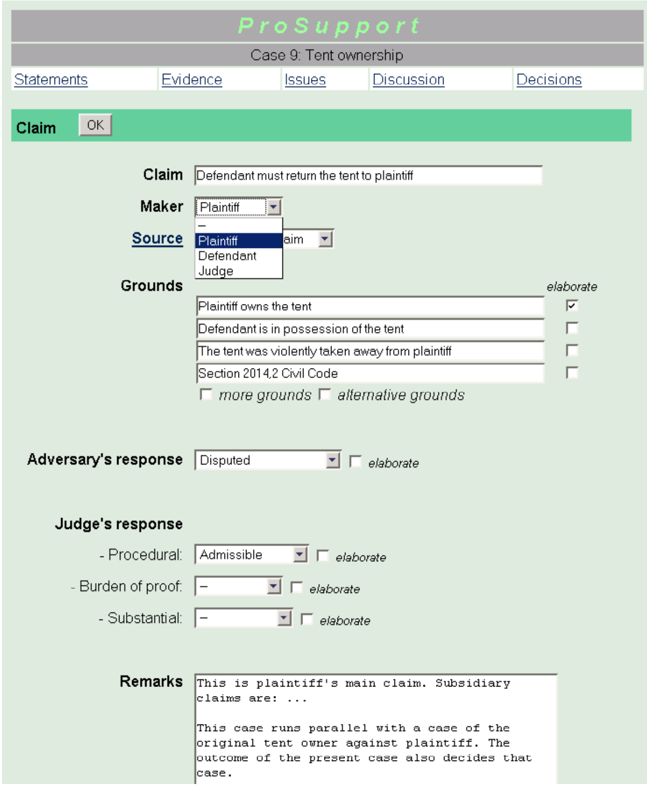


Figure 1. A claim form (expressing an argument).

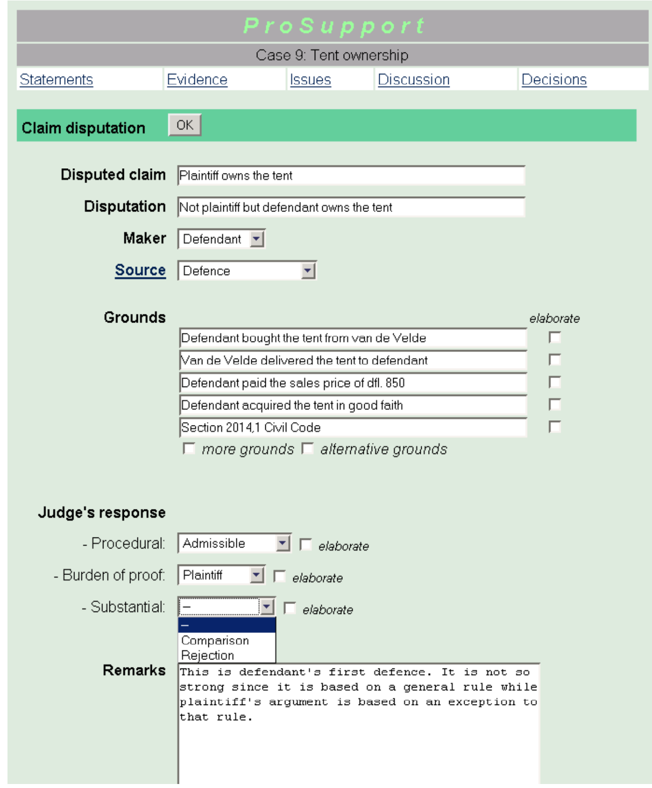


Figure 2. A claim disputation form (expressing a rebuttal).

since, so he argued, he had never sold the tent but only made it available to Nieborg for the period that he himself did not need it. He had done so since Nieborg had told him that he and his wife had never had have enough money to go on holiday. When Nieborg refused to return the tent, van de Velde, assisted by a group of people, threw Nieborg's son (who at that point was the only person present) out of the tent and took it away. A few months later, van de Velde sold the tent to defendant (van de Weg) and his wife. The sales price (dfl. 850) was paid with domestic work by Mrs. van de Weg in assistance of Mrs. van de Velde.

In court, Nieborg (plaintiff) claims return of the tent to him on the basis of his ownership. Van de Weg (defendant) disputes Nieborg's claim on the grounds that van de Velde had not sold the tent to Nieborg but only given it on loan, and that the work done by Nieborg and his wife was not done to pay the sales price but out of gratitude.

The relevant law is quite intricate and will not be explained here. The main issue on which the outcome of the case depended was whether van de Velde had sold the tent to Nieborg, so that Nieborg was owner at the time of the violent events, or whether van de Velde had just given the tent on loan, so that van de Velde had remained the owner.

Nieborg was allocated the burden of proving that Van de Weg had obtained the tent on loan. To meet his burden, he provided three witnesses, Van de Velde and two persons associated to van de Velde, Gjaltema and van der Sluis. Nieborg's main attack on van de Weg's evidence was that the witnesses were not credible: van de Velde had a personal interest in a win by van de Weg, and all three witnesses had declared something that Nieborg claimed was demonstrably false (we will not elaborate the latter point). However, the judge was convinced of their credibility, since their declarations supported each other and since Van de Weg had failed to find counterwitnesses. Nieborg therefore lost the case.

4 The discourse encoding schemes

We now turn to a description of the system's input encoding schemes, all based on the same generic scheme. In the present section we discuss their expressiveness and naturalness, while in the following section we describe them from a software-architecture point of view.

4.1 The schemes

In the present phase of the project, we have chosen for a simple format of arguments. Essentially, arguments are 'and trees' where the nodes are propositional atoms and the links are inference rules. The tree's root is the conclusion and its leafs are the premises of the argument. This setup enables us to let the user input elementary arguments with a web form with a list of fields, as is illustrated by Figure 1[2], which displays a **Claim** form expressing an argument for plaintiff's main claim. The top field is the argument's conclusion and the fields under **Grounds** are its premises. If more than four grounds are needed, the user can tick the *more grounds* box and push the **OK** button. This scheme for arguments is recursive: elementary arguments can be extended by replacing one of its grounds with a subargument for that ground. This is achieved by ticking the *elaborate* box next to the ground to be elaborated and pushing the **OK** button,

[2] The actual system is in Dutch; the English screens in this paper are created by manually editing the original HTML files.

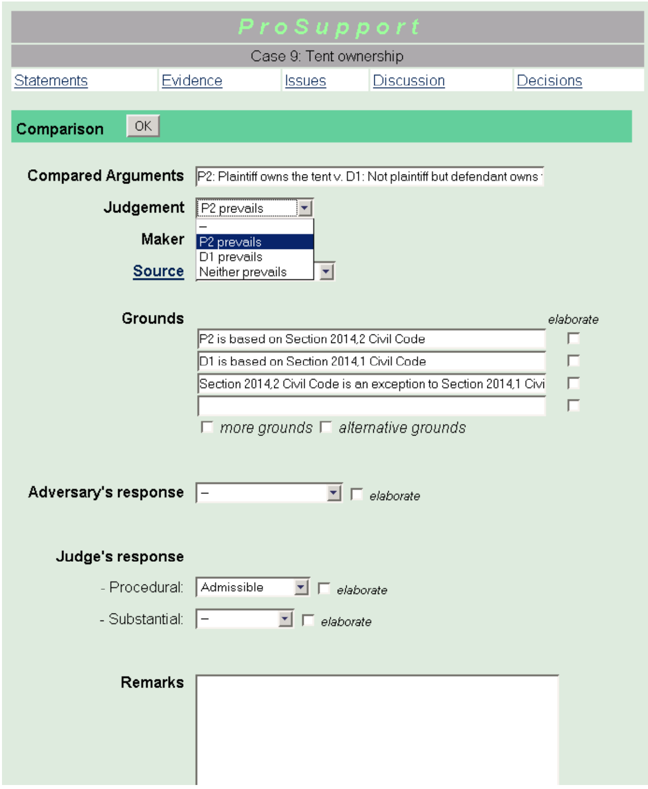


Figure 3. An argument comparison form (expressing a priority argument).

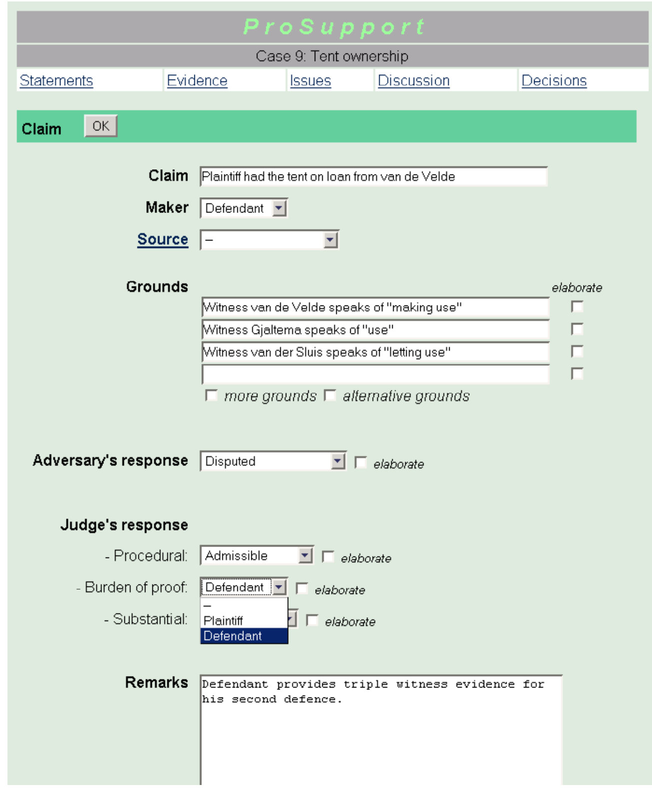


Figure 4. Another claim form (with an argument based on witness evidence).

which returns another instance of the claim form, with the top field filled by the to-be-elaborated ground. This box can also be used if any other information about the ground is to be entered, such as that it was disputed, or that a certain burden of proof was attached to it.

To describe the further setup of the claim form, the top row hyperlinks are links to various overviews of the discourse generated by the system on the basis of previous input. Of these, as yet only the **Statements** and **Discussion** links have been implemented. The **Statements** link returns a table with all statements made so far by any of the participants, including useful ‘metadata’, such as who made the statement, how the other parties responded, and so on. The **Discussion** link returns a visualisation of the discussion so far.

With the choice menu **Maker**, the user can enter who made the claim, by choosing from the options *Plaintiff*, *Defendant* and *Judge*. With the choice menu **Source** the user can enter the case file document in which the claim can be found and, if desired, make a hyperlink to the relevant fragment in the document (this hyperlink feature is not yet implemented). Under **Adversary’s response** and **Judge’s response** the user can enter the eventual responses of the adversary, respectively the judge to the claim. These options will be explained in more detail below. Finally, at the bottom of the form there is a large **Remarks** field, for entering anything of interest that cannot be entered in the other fields or menus.

To return to arguments, they can, depending on their role in the dispute, take on several (non-exclusive) dialectical roles: they can be initial arguments, counterarguments, priority arguments, and procedural arguments. (Unless indicated otherwise, we below mean with ‘argument’ an elementary argument as expressed in a single form).

Counterarguments can in turn be of two types. *Rebutting* counterarguments deny the conclusion of the attacked argument, while *undercutting* arguments deny that the premises of the attacked argument support its conclusion. An example of a rebuttal is that not plaintiff but defendant owns the tent, since defendant bought and acquired the tent from the previous owner (see Figure 2, which contains a rebuttal of a (not shown) subargument for the first ground in Figure 1). An example of an undercutter is an attack on the credibility of a witness whose testimony was used in the attacked argument. Figure 5 displays an undercutter moved by plaintiff in attack of defendant’s argument displayed in Figure 4. In legal disputes undercutters are very common, which is why we want to make the distinction between rebuttals and undercutters explicit, even though we are aware that this complicates the encoding schemes and therefore might detract from their usability.

The system cannot automatically recognise from an argument’s syntax whether it is a counterargument, since its input forms do not make negation explicit. Instead, the user must explicitly move a counterargument as an attack on another argument.

For counterarguments moved by an adversary this happens as follows. First from the **Adversary’s response** choice menu the ‘disputed’ option must be chosen (as in Figure 1). This returns another choice menu, this time non-exclusive, with the options ‘dispute claim’ and ‘dispute support’ (not shown). The first choice makes the system return a **Claim disputation** form (See Figure 2, but note that that form was not the result of disputing plaintiff’s main claim in Figure 1 but of disputing plaintiff’s first ground. This disputation was entered in the subform (not shown) that elaborates this ground). The top field of a claim disputation form contains the disputed proposition, the second field is for the formulation of the disputation, and the remaining fields are for the grounds for the disputation. The sys-

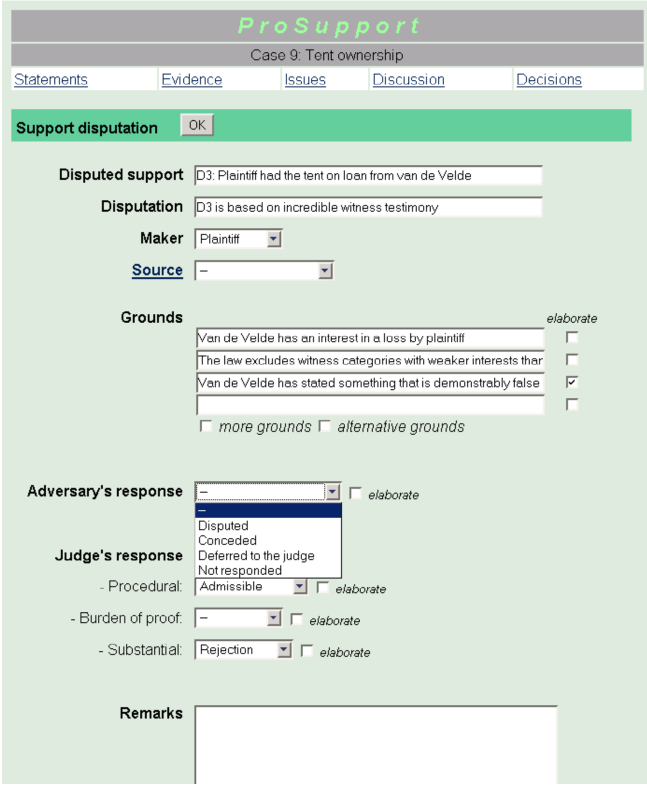


Figure 5. A support disputation form (expressing an undercutter).

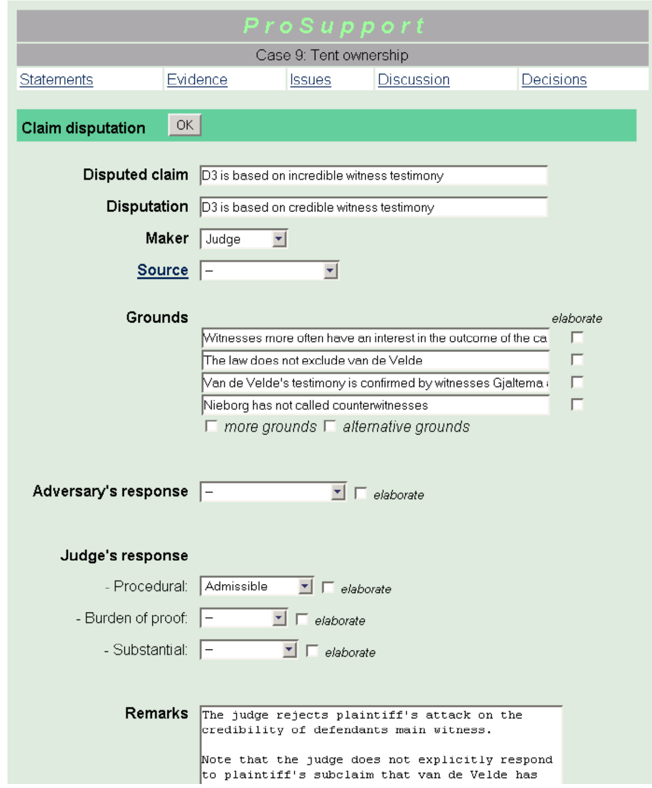


Figure 6. An implicit argument comparison by the judge

tem then treats the conclusions of an argument and its rebuttal as logical contraries. A choice for 'dispute support' makes instead the system return a **Support disputation** form (as in Figure 5, which resulted from disputing plaintiff's claim in Figure 4). Its top level field contains a system-generated description of the undercut support (in the current version an identifier plus the supported claim), its second field can be used to fill in the formulation of the undercutter, and the remaining fields can be used to enter the grounds for the undercutter.

A counterargument moved by the judge can be entered via the choice menu **Judge's response — substantial**, by choosing the option *rejection* (as in Figure 5). This makes the system return the same menu as with a 'disputed' choice for the adversary's response.

A *priority argument* is an argument that adjudicates a conflict between a rebuttal and its target argument. A priority argument of the judge can also be entered via the choice menu **judge's response — substantial**, by choosing the option *comparison* (see Figure 2). This returns a list of all rebuttals moved against the argument expressed on the form (not shown). The user can choose one of them, after which the system returns an **argument comparison** form (Figure 3). The top field mentions the identifiers and conclusions of the two arguments to be compared, the second field contains a choice menu for stating a preference between the arguments (a special form of a claim), and the rest of the form is as in the claim form. Note that thus we have slightly enriched our propositional language with the means to express preferences between arguments. In Figure 3 the judge adjudicates between two conflicting arguments concerning ownership of the tent. The judge prefers plaintiff's argument on the grounds that it is based on a legal rule which is an exception to the rule used by defendant's argument.

We do not allow priority arguments to adjudicate between an argument and its undercutter: if an undercutter is regarded as inconclusive, this should be expressed with a counterargument against the undercutter (as is done by the judge in Figure 6 with a rebuttal of plaintiff's undercutter in Figure 5). Such a counterargument can be a rebuttal (e.g. "no, the witness is credible, since ...") and then a priority argument can be moved on whether the undercutting argument or its rebuttal prevails (in fact, we regard a rebuttal moved by the judge as implicitly preferred over its target).

The last dialectical argument type is that of *procedural arguments*. They are subdivided into arguments on procedural correctness and arguments on allocating the burden of proof. A decision on procedural correctness can be entered with the choice menu **Judge's response — procedural** with the default *admissible* and a second option *inadmissible*. To enter an argument for an inadmissibility decision (which is optional), the box *elaborate* can be ticked, which makes the system return a form named **Violation**. Likewise for a decision on the burden of proof, via the choice menu **Judge's response — burden of proof**, which, when elaborated, returns a **Proof burden** form.

Finally, we must allow for alternative arguments for the same claim. Note that in a defeasible setting alternative arguments are not equivalent to a single argument with a disjunctive premise, since such a single argument does not capture that alternative arguments might be based on different kinds of inference schemes. For instance, one argument might be based on a statutory rule, while another argument might be based on legal policy considerations. Accordingly, below the list of grounds a box *alternative grounds* can be ticked, which returns an alternative claim form for the same claim. The alternative argument is assigned a different identifier than the original one.

4.2 How logical syntax is avoided

In our encoding scheme the user does not have to manipulate logical syntax, since logical operators are either implicit or not available. Above we already explained how negation is left implicit in the way rebuttals and undercutters are moved. Conjunction is, of course, implicit in the list of grounds. Furthermore, conditional operators are avoided since arguments do not have to be propositionally valid, so that conditional premises can be left implicit, paraphrased or named (e.g. with the name of a statutory rule as in Figures 1 and 2). Also, we think that there is no stong need for making disjunctions explicit. Firstly, as we explained above, alternative arguments for a claim (which are quite frequent) are not the same as an argument with disjunctive premises. Secondly, when a rule contains a disjunctive antecedent, we expect that in the great majority of cases to which the rule is applied, one of the disjuncts will hold. Consider, for instance, a social benefit law stating that being unemployed, ill or disabled entitles to a certain supplementary benefit. Finally, we expect that arguments that crucially depend on quantifiers or modal (such as deontic) operators will in practice be rare.

Of course, it is very likely that cases are found where our schemes are too limited. However, we think a discourse support system should not aim at 100% expressiveness, since that would conflict with the goal of usability.

4.3 How Dutch civil procedure has been modelled

In Section 2 we listed the features that our encoding schemes should capture. As can be seen from the above description, our schemes support the entering of all relevant dialectical types of arguments, as well as of all propositional attitudes (except retraction) that can be expressed by the adversaries and procedural decisions that can be taken by the judge.

We next recapitulate how the judge's substantial decisions can be entered. Completing the grounds of an adversary's argument can be simply done by adding a ground to an argument, ticking the corresponding *elaborate* box, and indicating in the elaboration form that the ground was moved by the judge. If the judge accepts an adversary's claim on alternative grounds, the user can simply check the box 'alternative grounds', enter such grounds and again indicate that they were moved by the judge. If a judge has rejected a claim or a claim's support on certain grounds, the user must choose the *rejection* option in the **Judge's response — substantial** menu, after which the claim or support can be disputed in the way explained above. Finally, the judge's comparative decisions can also be entered in a way explained above, by choosing the *comparison* option in the same menu. Note that the forms do not contain an explicit way to enter that the judge has accepted a certain claim. Such acceptance can be expressed either implicitly by doing nothing or, if the opponent had moved a counterargument, by attacking that argument in one of the available ways.

5 System architecture

We now describe the encoding schemes from a software-architecture point of view.

5.1 Design philosophy

The system architecture is based on the idea that all aspects of a case (issues, speech acts, source documents) are nodes in a network. The basic component (node) of the system's internal datastructure is called a *form*. Each form is intended to express a speech act. A form possesses several fields (or attributes), such as an ID, type, target, statement, maker, source, remarks, and typed pointers to other forms, such as grounds, adversary's response and judge's responses. Typically, each form uses only some of these attributes. For example, the main claim will have no value for the attribute 'target' because the main claim is the initial claim and by definition does not dispute other claims (see Figure 1). And a claim disputation form will have no adversary's responses, since a disputation is itself such a response (see Figure 2). When a form is presented to the user, undefined attributes are not shown, and the form takes its own "shape" depending on its type. Furthermore, depending on the type of form, its various attributes might be named in different ways. For instance, the attribute 'target', which links the form to a preceding form, is in a claim disputation form (Figure 2) called "disputed claim" and in a violation form (not shown) called "inadmissible speech act". And the attribute 'statement', which indicates the proposition a form is about, is in a claim form (Figure 1) called "claim" and in a 'comparison' form (Figure 3) called "judgement".

To prevent redundancy and preserve the logical structure of a case, every form is unique, which means that the same thing is always expressed in the same way. For example, if the statement field of a certain form is changed, and this form is used by forms A, B, and C, (e.g. as ground for their statement) then this change will be reflected if A, B or C are retrieved and presented on screen. Further, the system suggests the user to reuse forms by presenting ID's of existing forms. If the user enters a form-ID rather than plain text, the system will recognise this and will establish a link rather than create a new form. This feature can be used, for instance, to reuse old statements as grounds of a new argument.

As said above, form types are meant to stand for speech acts. We currently distinguish *Claim*, *Claim disputation*, *Support Disputation*, *Comparison*, *Violation*, and *Proof burden*. For instance, *Claim* stands for making a claim, *Claim disputation* for disputing a claim, and *Violation* for deciding a speech act procedurally inadmissible. For some types of speech acts we do not want to allow for elaboration; such speech acts are not captured by their own form, but simply as an attribute of another form. For instance, conceding a claim is an attribute of a claim form. Finally, the speech act of moving an argument, i.e., of stating grounds in support of a claim or disputation is left implicit in the forms and how they are linked.

5.2 Aspects of human-computer interaction

Forms can be presented to the user in various formats. Currently, it is possible to view forms in isolation, and to view them all together. When viewed in isolation, all relevant attributes of a form are shown, including the contents of the statement fields of connected forms, and links to them. Showing the statement fields of connected forms increases the cohesion of the network and enables to user to quickly navigate through a case.

Viewing forms together enables a bird's-eye perspective on a case. Currently, the following global views are possible. The most obvious presentation consists of a table of all statements, accessible via the **Statements** hyperlink. This table can be sorted among various dimensions (e.g. ID number, type, time of input, time of modification).

or filtered through various criteria (e.g. "show all disputed statements made by plaintiff for no burden of proof has yet been allocated"). Further, it is possible to view a tex-based summary of the case (via the **Discussion** hyperlink) and to view the case as a directed graph (not yet incorporated in the above screens). It should be noted that our architecture does not commit to a particular visualisation style of the discussion; it equally supports text-based and graph-based styles.

One of the greatest challenges of our project is to keep the layout of the input forms as simple as possible, while respecting the complexity of the case. The approach that PROSUPPORT follows is that it is kept simple and fixed for beginners, while advanced users may opt for more features and flexibility.

5.3 Current state of the implementation

The current version of our system is implemented in Mason (`http://www.masonhq.com`). Mason is a Perl-based web site development and delivery engine. With Mason it is possible to embed Perl code in HTML and construct pages from shared, reusable components. Mason requires an Apache HTTP server with a software package that embeds a Perl interpreter into the webserver (typically `mod_perl`). Forms are written to and retrieved from a Berkeley type data base, where forms are accessed by their ID.

As for the current state of implementation, the above-described form-based datastructures have been implemented, as well as a first method to navigate between the encoding screens. Of the overview facilities, only the **Statements** and **Discussion** features have been implemented. We have not yet implemented the function that is meant to compute the 'current outcome' of a case.

Some elements of our implementation are still provisional. Firstly, as for navigating between the forms, some problems still have to be solved. One problem is that the user can mark more than one text field for further elaboration. In such cases, more than one form needs to be filled out and it is not immediately clear which of these forms that should be, i.e., which of these forms must be presented next to the user. One solution is to work with a prioritised agenda, called "forms to be processed," and then to enable the user to process these forms as he sees fit. Secondly, our current way to visualise the discussion is also still provisional; in fact, a full implementation of this feature is an important research issue of the PROSUPPORT project, which will touch upon cognitive as well as technical issues.

6 Theoretical foundations

As said above, one goal of the PROSUPPORT project is to investigate how a natural encoding scheme for argumentative discourse support can be developed on a sound formal basis. We think that such a basis can be provided by combining two recent developments, viz. logics for defeasible argumentation and formal dialogue systems for critical discussion.

6.1 Logics for defeasible argumentation

Logics for defeasible argumentation (see [14] for an overview) are one approach to the formalisation of so-called defeasible, or nonmonotonic reasoning. This is reasoning where tentative conclusions are drawn on the basis of uncertain or incomplete information, which might have to be withdrawn if more information becomes available. Logical argumentation systems formalise this kind of reasoning in terms of the interactions between arguments for alternative conclusions. Nonmonotonicity arises since arguments can be defeated by stronger counterarguments.

There are several reasons why argumentation systems are a promising formal basis for argumentative discourse support systems. Clearly, modelling inference as comparing arguments and counterarguments fits very well with the dialectical nature of argumentative discourse. Moreover, argumentation systems often abstract to a large degree from the logical language in which arguments are expressed and from the rules according to which they are constructed. This makes such systems particularly suitable for dealing with natural-language input. For instance, above we saw how logical syntax can be avoided and how hidden premises can remain implicit. Finally, argumentation logics have been applied to a number of phenomena that we think are important in argumentative discourse support, such as the format of arguments as trees of inference rules (e.g. [10, 19]), the distinction between rebuttals and undercutters (due to Pollock, e.g. [10]), and priority arguments (e.g. [5, 13]). Note that all these three phenomena are captured by our encoding schemes.

6.2 Dialogue games for dispute resolution

In the introduction we said that one use of formal foundations is as a basis for computing the 'current outcome' of a dispute. Now it is important to note that the outcome of a dispute depends not only on the arguments that are stated but also on the various argumentative speech acts and procedural decisions. For instance, if a premise of an argument is disputed and no further argument for it is given, the argument does not count in determining the outcome of the dispute; likewise for an argument of which one premise was ruled to contain inadmissible evidence. And for computing the effect of priority arguments on the outcome of a dispute, it is important to know who has the burden of proof: if two conflicting arguments are decided to be equally strong, this benefits the adversary who does not have the burden of proof.

So argumentative speech acts of various kinds interact in subtle ways in determining the outcome of a dispute. Therefore, the formal basis of a discourse support system cannot be confined to argumentation logics; they need to be embedded in formal dialogue systems for dispute, for instance, in the dialogue systems of [21]. For two examples of work of this kind see [3] and [12].

Accordingly, we have set up PROSUPPORT such that each input in the system can be formally translated as a move in such a dialogue system (although we have not yet fully carried out this translation). On the other hand, we have also designed the system such that the user needs not be aware of this translation. The reason is that we expect the intended users will find a WEB-form interface more natural than an explicit dialogue game style interface, which still seems somewhat artificial.

7 Discussion of alternatives and remaining issues

As for arguments, the expressiveness of our system lies mainly in two aspects: it can keep track of (often nested) support relations between statements, and it can identify the main dialectical relations between arguments. However, our language for expressing arguments is (deliberately) very simple. We now discuss some possible enhancements.

As explained, our system allows to distinguish three parts of (elementary) arguments: their premises, their conclusion, and their inference rule. (Actually, the nature of the inference rule is not made explicit; instead it is only named). We could, of course, have imposed more structure. One scheme that comes to mind is Toulmin's well-known generic argument scheme [17]. However, we fear that this scheme might be too rigid and too complex for practical use, since it requires that for every argument a uniform distinction between data, warrant and backing is made explicit. Especially when combined with the practical need to make the scheme recursive, this often leads to quite complex encodings of legal arguments, as was shown by [9].

In our opinion, a more promising refinement is the inclusion of a set of optional specialised argument schemes. ("Optional" means that such schemes could be offered as an advanced option to experienced users of the system.) Specialised argument schemes are an important research topic within argumentation theory (see e.g. [20]). For present purposes, some useful schemes are the use of types of evidence (such as witness testimonies, expert reports, and documents). Such specialised argument schemes are less rigid and abstract than Toulmin's scheme. Moreover, they come with specific sets of 'critical questions', which can focus a discussion. Finally, the logical interpretation of argument schemes is rather straightforward: they naturally map onto Pollock's well-known notions of defeasible reasons and defeaters. Note that a negative answer to a critical question attached to an argument scheme will in fact be a counterargument, often of the undercutting type. For instance, Walton in [20] lists as one of the critical questions of arguments from testimony, the question whether the witness is credible. Above in Figure 5 we formulated a negative answer to this question as an undercutting counterargument.

An important restriction of our generic scheme is that, as for support relations between propositions, it can only capture and-tree relations between propositions. For certain types of reasoning, such as abductive-causal reasoning or probabilistic reasoning, this may not be suitable.

Finally, we have chosen not to model the concept of propositional commitments in our system. Although this is a very important theoretical concept (cf. [21]), we think that violation of commitments will in practice not often be an issue, while modelling them makes the system more complex and thus detracts from the goal of usability.

8 Related research

In the legal field, so far been two implemented architectures for practical use have been described, viz. Loui's Room 5 system [8] and Verheij's ArguMed [18]. A related system outside the legal field is Belvedere [16], a system for teaching scientific argumentation. Furthermore, Lodder & Huygen [7] report on the ongoing development of their support tool $eADR$ for simple procedures for online dispute resolution.

All four systems support the user in drafting arguments and counterarguments (Room 5 also supports the search of legal case databases and the incorporation of retrieved case citations in arguments). ArguMed is the only system that, besides rebuttals, also supports undercutters; none of the systems supports priority arguments. Unlike PROSUPPORT, these systems do not support the entering of other relevant speech acts. Room 5 and ArguMed are, like PROSUPPORT based on logics for defeasible argumentation, and have an implemented 'current outcome function' based on such a logic. Belvedere and $eADR$ are not based on formal foundations. As for the appearance of the input forms, ArguMed and Belvedere are graph-based, while Room 5 uses encapsulated text frames and $eADR$ uses a format similar to threaded discussion boards, where replying messages can be either supporting or attacking replies (the authors do not specify whether multiple supporting replies are meant to be cumulative or alternative grounds). Neither of these projects addresses the issue of the generation of discussion overviews in formats different from their encoding schemes. Finally, Belvedere is the only of these four systems that has been subjected to systematic field studies.

Summarising, we think that, compared to these systems, our main contributions are a separation of the layouts of the input and output interfaces, an alternative, web-browser-based interface for input encoding schemes, and the modelling not only of arguments and their dialectical relations, but also of argumentative and procedural speech acts. The latter feature especially allows for an adequate modelling of reasoning under burden of proof, which in legal applications is very important. It remains to be seen whether this extra expressiveness makes the resulting extra computational power outweigh the increased complexity of use.

9 Conclusion

In this paper we have investigated to which extent a theoretically well-founded account of argumentative discourse can be implemented as an argumentative discourse support system. We have especially focused on the encoding schemes with which the user can enter his or her analysis of a dispute. The main question was how such encoding schemes can, on the one hand, be natural and easy to use and, on the other hand, support useful computational power of the system. With respect to the latter, we have especially kept in mind a feature that computes the 'current outcome' of a dispute.

We have argued that, if the expressiveness of the encoding schemes is sufficiently restricted, a natural and useful implementation is possible with a world-wide popular software tool, viz. web browsers, linked to a database. We have also argued that, with respect to expressing arguments, a suitable restriction is to encode no more than support relations between statements within arguments, and dialectical relations between arguments. Moreover, we have argued that our encoding schemes can be given a formal basis in terms of logics for defeasible argumentation and formal dialogue systems for critical discussion.

Of course, our findings are still preliminary. For one thing, we have so far tested our designs on the case files of only one case. More importantly, so far we have not obtained any substantial user experience, which yet is essential for testing usability and usefulness. Nevertheless, we think the results so far are promising enough to further develop our approach and conduct realistic field tests.

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**, 233–254, (2000).

[2] J. Conklin, A. Selvin, S. Buckingham Shum, and M. Sierhuis, 'Facilitating hypertext for collective sensemaking: 15 years on from gIBIS', in *Proceedings of the The Twelfth ACM Conference on Hypertext and Hypermedia (Hypertext 2001)*, New York, (2001). ACM Press. In Press. Also available as Technical Report KMI-TR-112, Knowledge Media Institute, The Open University, UK.

[3] T.F. Gordon, *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[4] J.C. Hage, 'Dialectical models in artificial intelligence and law', *Artificial Intelligence and Law*, **8**, 137–172, (2000).

[5] R.A. Kowalski and F. Toni, 'Abstract argumentation', *Artificial Intelligence and Law*, **4**, 275–296, (1996).

[6] R.E. Leenes, 'Burden of proof in dialogue games and Dutch civil procedure', in *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, pp. 109–118, New York, (2001). ACM Press.

[7] A.R. Lodder and P.E.M. Huygen, 'eADR: a simple tool to structure the information exchange between parties in online alternative dispute resolution', in *Legal Knowledge and Information Systems. JURIX 2001: The Fourteenth Annual Conference*, pp. 117–129, Amsterdam etc, (2001). IOS Press.

[8] R.P. Loui, J. Norman, J. Alpeter, D. Pinkard, D. Craven, J. Linsday, and M. Foltz, 'Progress on Room 5: A testbed for public interactive semi-formal legal argumentation', in *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, pp. 207–214, New York, (1997). ACM Press.

[9] S.E. Newman and C.C. Marshall, 'Pushing Toulmin too far: Learning from an argument representation scheme', Technical Report SSL-92-45, Xerox Palo Alto Research Center, Palo Alto, CA, (1992).

[10] J.L. Pollock, *Cognitive Carpentry. A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA, 1995.

[11] H. Prakken, 'Modelling defeasibility in law: logic or procedure?', *Fundamenta Informaticae*, **48**, 253–271, (2001).

[12] H. Prakken, 'Modelling reasoning about evidence in legal procedure', in *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, pp. 119–128, New York, (2001). ACM Press.

[13] H. Prakken and G. Sartor, 'Argument-based extended logic programming with defeasible priorities', *Journal of Applied Non-classical Logics*, **7**, 25–75, (1997).

[14] H. Prakken and G.A.W. Vreeswijk, 'Logics for defeasible argumentation', in *Handbook of Philosophical Logic*, eds., D. Gabbay and F. Günthner, volume 4, 219–318, Kluwer Academic Publishers, Dordrecht/Boston/London, second edn., (2002).

[15] F.M. Shipman and C.C. Marshall, 'Formality considered harmful: Experiences, emerging themes, and directions on the use of formal representations in interactive systems', *Computer Supported Cooperative Work*, **8**, 333–352, (1999).

[16] D. Suthers, A. Weiner, J. Connelly, and M. Paolucci, 'Belvedere: engaging students in critical discussion of science and public policy issues', in *Proceedings of the Seventh World Conference on Artificial Intelligence in Education*, pp. 266–273, (1995).

[17] S.E. Toulmin, *The Uses of Argument*, Cambridge University Press, Cambridge, 1958.

[18] B. Verheij, 'Automated argument assistance for lawyers', in *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, pp. 43–52, New York, (1999). ACM Press.

[19] G.A.W. Vreeswijk, 'Abstract argumentation systems', *Artificial Intelligence*, **90**, 225–279, (1997).

[20] D.N. Walton, *Argumentation Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[21] D.N. Walton and E.C.W. Krabbe, *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*, State University of New York Press, Albany, NY, 1995.

Cues for Reconstructing Symptomatic Argumentation

Francisca Snoeck Henkemans[1]

1 Argumentative indicators

Every argument can be characterized by an argumentation scheme which defines the justificatory relation between the argument and the standpoint to which the argumentation relates. In the pragma-dialectical approach, a distinction is made between three main categories of argumentation schemes: argumentation based on a causal relation, argumentation based on a relation of analogy and argumentation based on a symptomatic relation [2]. A similar division of types of schemes can be found in the classical rhetorical literature, in the traditional American debate textbooks and in the work of modern rhetoricians such as Weaver [7].

In a research project on argumentative indicators Frans van Eemeren, Peter Houtlosser and I are carrying out, we investigate which clues in the verbal presentation can be used to reconstruct the relationship on which an argumentation is based and to determine what type of argument is used. The project is embedded in the theoretical framework of the pragma-dialectical approach to argumentation. Its aim is to make a systematic inventory of the verbal means used in the Dutch language to express an argumentative function of language use, to classify these means in terms of the ideal model of a critical discussion and to identify the conditions under which they can fulfil a specific argumentative function.

In our project we pay attention to all elements that are crucial to the evaluation of the argument and need to be represented in an analytic overview of an argumentative text or discussion, such as the type of dispute, the argumentation structure and the argumentation schemes. For each discussion stage we establish which words and expressions can function as indicators of the relevant moves in that particular stage and as indicators of the relations between these moves. Each type of argumentation has its own assessment criteria: for each type of justificatory relation different critical questions are relevant. Someone who makes use of a particular argumentation scheme, thereby takes the first step in a dialectical testing procedure that requires the arguer to deal with specific forms of criticism in order to defend the standpoint successfully (see van Eemeren, 'The importance of being understood') . In anticipation of possible criticism, the protagonist of a standpoint can follow up his argument with further arguments dealing with relevant objections. In a fully externalized discussion, the reactions of the opponent will relate to the evaluation issues that are relevant to the argumentation scheme concerned. It is therefore not only in the presentation of the argumentation itself, but also in the critical reactions of the opponent, and in the speaker's follow-up to his argument, that clues can be found as to the type of relation between argument and standpoint.

In this paper, I shall illustrate our approach to argumentative indicators by discussing various types of indicators of symptomatic argumentation. I shall make a distinction between 1) clues in the presentation of the argumentative relation, 2) clues in the critical reactions of the opponent, and 3) clues in the speaker's follow-up to his argument. I shall first explain why the expressions concerned can be seen as indicators. Then I shall specify to which elements of the symptomatic argumentation scheme the expressions concerned refer.

[1] Faculty of Humanities University of Amsterdam

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

2 The symptomatic relationship

In argumentation that is based on a symptomatic relation, a property, class membership, distinctive characteristic, or essence of a particular thing, person, or situation is mentioned which implies that this thing, person or situation also has the characteristic property that is ascribed to it in the standpoint. The following example is an instantiation of the symptomatic argumentation scheme:

(1) Bill is very egocentric
because Bill is an only child
and Egocentrism is characteristic of people who are an only child

In this example, the fact that Bill belongs to the class of people who are an only child is used as a basis for concluding that he also has the characteristic of being egocentric. Such a symptomatic relation can also be used in the opposite direction. The fact that Bill is egocentric is then used as an argument for the conclusion that he must be an only child:

(2) Bill must be an only child
because He is very egocentric
and Egocentrism is characteristic of people who are an only child

According to their definition of symptomatic argumentation, van Eemeren and Grootendorst consider this variant as the prototypical form of symptomatic argumentation:

> The argumentation is presented as if it is an expression, a phenomenon, a sign or some other kind of symptom of what is stated in the standpoint [2, : 97].

By this definition, the argument that is advanced can be seen as an indication or a sign that something is the case, or that a particular qualification is justified. For Perelman [5], the distinction between the sign (or the manifestation of a particular phenomenon) and the phenomenon itself is a hierarchical distinction. In relations of *co-existence* (Perelman's term for symptomatic relations), the elements that are connected are always on an unequal level:

> Liaisons of coexistence establish a tie between realities on unequal levels; one is shown to be the expression or manifestation of the other [5, : 89-90].

A prototypical example given by Perelman of the relation of co-existence is the relation between a person and his actions, opinions or works. There is a continual interaction between the person and his actions. The relationship can therefore be used in two ways: the image one has of the person makes it possible to arrive at conclusions concerning his acts (or other manifestations of the person) and vice versa [5, : 90].

The general argumentation scheme for the symptomatic relation is, in the pragma-dialectical theory, as follows:

	Y is true of X,	
because:	Z is true of X	
and:	Z is typical (characteristic/symptomatic) of Y.	

According to van Eemeren and Grootendorst [2, : 101]the following critical questions are to be asked about a symptomatic argument:

- Is Z indeed typical of Y?
- Is Z not also typical of something else (Y')?

3 Clues in the presentation

3.1 Expressions referring to a symptomatic relation

In characterizations of the symptomatic relationship the notions 'characteristic' and 'sign' play a crucial role. I shall take these two notions therefore as the starting point in my search for examples of expressions that are indicative of the symptomatic relation. I make a distinction between (1) expressions that can indicate relations in two directions: the characteristic can be mentioned either in the argument or in the standpoint, and (2) expressions that can only indicate relations in one direction and the characteristic or sign can only be mentioned in the argument.

In order to determine which type of expressions can serve as indicators of the symptomatic relation, I start by looking at the definitions of these two key notions that are given in the *Oxford English Dictionary* [6]. The following uses are, among others, mentioned of the words 'characteristic' and 'sign':

Characteristic

- a distinctive mark, trait, or feature; a distinguishing or essential peculiarity or quality
- (adj.) that seems to indicate the essential quality or nature of persons or things; displaying character; distinctive; typical

Sign

- a mark or device having some special meaning or import attached to it, or serving to distinguish the thing on which it is put
- a token or indication (visible or otherwise) of some fact, quality etc.
- an objective evidence or indication of disease
- a trace or indication of something
- a mere semblance of something
- an indication of some coming event

According to these definitions, the notion 'characteristic' can both refer to the characteristic properties of a person or thing and to a sign of something being the case or something or someone being of a particular type. 'Sign' is used as a synonym of 'proof' or 'evidence' for the existence or the nature of something or someone. An important aspect of the meaning of a characteristic as well as a sign is that they make something perceptible - or at any rate knowable.

By also taking into account the synonyms of the terms that are used in these definitions, a non-exhaustive list can be made of expressions that may be indicative of the symptomatic relation[2]. In these expressions it is more or less explicitly stated that the relation is symptomatic. Most of these expressions (with the exception of the last four expressions under b.) will generally be found in the major premiss of the argument, since this is the premiss in which the relationship between standpoint and argument becomes apparent.

a. **Indications of symptomatic relations in two directions**
X is characteristic of Y
X is typical of Y
X is illustrative of Y
X marks Y

b. **Indications of symptomatic relations in one direction**
X is a sign of Y
X is evidence of Y
X shows Y
X implies Y
X means Y
X proves that Y
X indicates Y
X testifies to Y
X is a token of Y
X tells us something about Y

X, (so) apparently Y
X, (so) obviously Y
X, (so) it is clear that Y
X, (so) it turns out that Y

In the examples (3) to (6), various indicators of symptomatic argumentation are used:

(3) The woman had requested her family to let the cats be put to sleep and to bury them with her in the position in which they would normally sleep in her bed at night: one at the head of the bed, one on her belly and one at the foot of the bed. This development *tells us something about* our society, that *apparently* sees an animal as the substitute of a fellow creature (*de Volkskrant*, May 10, 1996).

(4) The truth is, sex and violence have never been bad business for advertisers, *proven by the fact that* one of the world's biggest sponsors, Procter & Gamble, has for years produced daytime soaps - including CBS' "The Guiding Light" and "As the World Turns" - that contain as much sexuality ounce for ounce as any other programming on television. (*Los Angeles Times*, September 19, 2000).

(5) Only a few thousand curious fans stopped by the Arrowhead Pond to check out Pierre Gauthier's summer remodeling job. What they witnessed Monday was hardly worth the trip. One lackluster offensive showing would be forgivable as typical of early exhibition games. Two wouldn't be anything to fret about. But three in a row *means* a disturbing trend has developed, which is where the Ducks stand today after a 2-0 loss to the Phoenix Coyotes left them winless in three exhibitions. (*Los Angeles Times*, September 19, 2000).

(6) Cadans has never done anything to rehabilitate me or support me, never have I received a benefit or sickpay, nor have I ever been medically examined. *It's clear that* there is something wrong with the organization of this institution for social security (*de Volkskrant*, CD-Rom 1998).

A difference between the indicators of symptomatic relations in one direction and those in two directions is that the former, unlike the

[2] In van Eemeren and Grootendorst [2, : 98-99] a list of more or less standardized expressions for indicating a particular argumentation scheme is provided in which many of the expressions I deal with are mentioned.

latter, always establish an argumentative connection between the two connected elements: they also indicate that the first element (X) is evidence for or proof of the other (Y). The indicators of symptomatic relations in two directions, on the other hand, can also be used to argue for the opposite, i.e. that Y is evidence for X, as in example (1). They may also be used non-argumentatively, for instance when giving a description of something or someone. Within the group of expressions indicative of the symptomatic relationship in one direction, the expressions 'X, apparently Y,' 'X, obviously Y,' 'X, it is clear that Y,' and 'X, it turns out that Y' form a separate group, because they can be combined with 'so', while this is not the case with the other expressions. 'Apparently', 'obviously' etc. can only occur in the standpoint of the argument, not in the major premiss.

3.2 Expressions referring to aspects of the symptomatic relation

There are also expressions that do not express the whole relationship between argument and standpoint but that can be indicative of specific aspects of the symptomatic relation. In particular, there are a number of expressions that refer to aspects connected with what Perelman calls a relation between the person and his manifestations. The expressions mentioned below, for example, are an indication that a particular quality or trait is inherent in a particular person, animal or thing, that it is an essential characteristic, or that someone or something constantly has a certain quality or repeatedly shows a particular kind of behavior.

Only if these expressions occur in the major premise of the argument they are a direct indication of the symptomatic relation. In that case, they provide just as strong evidence as the expressions already mentioned, which make the symptomatic relation explicit. All the expressions indicative of certain aspects of a symptomatic relation can not only occur in the major premise, but also in the minor premise and in the standpoint. If they occur in the minor premise or the standpoint, they offer an indirect clue that the relation in question may be symptomatic. Then the use of these expressions shows at least that the presence of certain inherent or permanent qualities plays an important role in the argument, so that there is reason to believe that we could be dealing with a symptomatic argument.

Expressions indicative of aspects of a symptomatic relation
is by nature
is in his blood
is a seasoned/experienced

is a true, real, regular, veritable, first-rate
is essentially, basically, at bottom, at heart, fundamentally
is simply/just
is by definition
is known as/reputed to be
is by tradition

will (always) be
remains
always/all his (or her) life

In example (7) to (9) such expressions are used. To show clearly which statement contains the indicator, I give a reconstruction of the argumentation in these examples.

(7) [It has turned out that a Scottish bishop has a son and is living together with a divorced woman]
Fortunately there was in Kendal also Mrs. Mitchell, the neighbor of the sinful Scottish bishop. She told the paper that she could easily understand all this. "Men will be men" (*de Volkskrant*, September 23, 1996).
Reconstruction example 7 (indicator of relation in major premise)
It is understandable that the bishop has violated the rules of celibacy (because he is a man)
and *men will be men* [= it is characteristic of men that they find it difficult to remain celibate]

(8) "Do you really believe that businessmen in the West set light to each others shops?", I asked. "It has to be so," he said. "Because actually, Russians are good by nature" (*de Volkskrant*, August 29, 1996).
Reconstruction example 8 (indicator of relation in minor premise)
It can't be Russians who set light to the shops
because Russians are good *by nature*
(and it is characteristic of people who are good by nature that they do not set light to shops)

(9) Brinkman has become a real Italian. She lives from one day to the next, carpe diem (*de Volkskrant*, September 23, 1999).
Reconstruction example 9 (indicator of relation in standpoint)
Brinkman has become a *real* Italian
since she lives from one day to the next
(and living from one day to the next is characteristic of Italians)

3.3 Clues for the symptomatic relation in the sentence structure

Apart from the expressions that can be indicative of the symptomatic relation or aspects of it, there is a sentence structure that is preeminently suitable for constituting the standpoint or minor premise of a symptomatic argument. Some of the expressions that point to aspects of the symptomatic relation can be combined with this sentence structure. The structure in question is the 'subject - copula - complement' sentence structure, in which the complement consists of an adjective or a noun. Examples of this structure are the following:

X is (a) Y
X seems (to be) (a) Y
X appears to be (a) Y

This sentence structure has a number of properties which seem to make it suitable for presenting the standpoint or the minor premise of a symptomatic argument. According to Greenbaum [4], predicatives typically characterize the subject, and the verb 'to be,' when used in such a construction, is a stative verb, that is, a verb used in referring to a state of affairs (1996: 73-74). Since symptomatic argumentation is generally speaking about qualities and features and not about events or processes, it is plausible to assume that when an argument or standpoint has the sentence structure subject - copula - complement, this is already an indication that the argumentation might be based on a symptomatic relation. The similarity of the properties of this sentence structure to that of the symptomatic relation becomes even more apparent when variants of the symptomatic argumentation scheme are taken into account. In his comparison of various approaches to argumentation schemes, Garssen [3] considers the following types of argument that are mentioned in the literature as variants of what pragma-dialecticians call the symptomatic argumentation scheme:

- Argumentation based on a classification

- genus-species argumentation
- argumentation based on evaluation criteria
- argumentation based on a definition
- identity relations [3, : 77, 120, translation FSH]

When we compare these variants with the functions the *Collins Cobuild English Grammar* [1] lists of the sentence structure subject - copula - complement, there appears to be a close parallel between the purposes for which this sentence structure is used and the types of relation that are considered to be symptomatic:

- to say what type of person or thing someone or something is
- to describe or identify the subject
- to indicate what qualities someone or something has
- to indicate exactly who or what someone or something is ('indicating identity') [1, : 173-176]

The copulas 'to seem' and 'to appear' can fulfil similar functions as 'to be' when they are combined with a complement, but lend a specific modal shade to the sentence: 'to seem' and 'to appear' are both used when the speaker is making a statement of which he is not completely certain or that he knows from hearsay.

4 Clues in the way the argumentation is criticized and the arguer deals with criticism

Since the reactions of the opponent may be expected to relate to the evaluation issues that are relevant to the argumentation scheme concerned, it is not only in the presentation of the argumentation itself, but also in the critical reactions of the opponent, and in the speaker's follow-up to his argument, in which he comes up with further supporting arguments to deal with anticipated or real criticism against his orginal argument, that clues can be found as to the type of relation between argument and standpoint. The wording of the criticism may given an indication of the type of critical question the opponent is raising. And the arguer's follow-up to his argument may provide clues as to the type of criticism the is anticipating. I shall illustrate this by discussing some examples.

In example (10), Mr. Moghraby suggests that the warm reception he and his fellow passengers received in Iraq might be seen as a sign that the hijack he was involved in had been planned, or at the very least, that treating the stranded passengers so well suited the purposes of the Iraqi government. This argumentation is subsequently criticized in a letter to the editor: the letter writer claims that the good treatment that was given to the passengers cannot be seen as an indication of any ulterior motive (first critical question), since it is characteristic of Iraqis that they always treat foreigners well. One should therefore "not read something into this situation that is not really there."

(10) Britons taken to Baghdad by hijackers aboard their Saudi plane were astonished to discover that their detour coincided with the start of "Iraqi Tourism Week". [...] The 86 passengers, 40 of them Britons, aboard the Jeddah-London flight hijacked on Saturday, were "treated like royalty", said Omer Moghraby [...] Mr. Moghraby said: "I don't know if the warm reception was a set-up, but it did all *seem* convenient. It didn't feel like the hijack was planned, but they were *obviously* very happy to see us and made full use of our being there" (*The Daily Telegraph*, October 17, 2000).
Reaction (letter to the editor):
SIR - I can easily believe that the hijacked passengers taken to Baghdad were treated like royalty (report, Oct 17). Iraqis *have always* treated foreigners, whether they are British or not, as VIPs. It is a shame that the "world" *is reading something into this situation that really isn't there* (*Daily Telegraph*, October 18, 2000).

In example (11), Smoak-Bartolo reacts to the argument that the fact that Latin American women spend much time in front of the mirror proves that they are vain. She accuses people who think this of not understanding that the behavior of Latin American women is in fact a sign of something else (second critical question): it is a way of honoring their tradition - or in Smoak-Bartolo's words: it is a reflection of our grandmothers, our homeland and our pride:

(11) Why is it that Latinas catch so much flack over the time we spend in front of the mirror? "It can *seem* like vanity, but I think those who think that about us do not understand it's part of our heritage," says Smoak-Bartolo. "It's deeply rooted. It's a *reflection* of our grandmothers, our homeland and our pride." (*Los Angeles Times*, October 10, 2000)

The way in which a protagonist follows up his argument in an attempt to silence possible opponents by showing that a possible criticism does not apply can also provide a further indication of the type of relation on which the argument was based. In example (12), Lamar Alexander's leaving the presidential race and Warren Beatty's entering it are presented as a sign of new developments in the presidential race. To make it clear that these two actions are indeed a sign of new developments (first critical question), the arguer supplies further argumentation: Beatty's entering the presidential race and Alexander's leaving it show that this race is growing more attractive for message candidates and less attractive for conventional contenders.

(12) Lamar Alexander - two-term governor of Tennessee, former Education secretary - has left the presidential race. And Warren Beatty - actor, director and behind-the-scenes Democratic activist - might enter it. That's a sure *sign* some new curves are emerging on the road to the White House. [...] As Beatty's flirtation *suggests*, the presidential race is growing more attractive for message candidates, even as it becomes more daunting for conventional contenders like Alexander (*Los Angeles Times*, August 23, 1999).

5 Making use of indicators in reconstructing the argumentative relation

To arrive at a well-founded reconstruction of symptomatic argumentation, one cannot restrict oneself to merely pointing out there is an indicator of symptomatic argumentation. In the first place, it has to be established that the indicator is really used in an argument. A lot of the indicators of symptomatic argumentation also occur in non-argumentative discourse. An example of this is the expression 'is characteristic of,' which can be an indicator of the symptomatic relation 'in two directions'. The presence of an expression such as 'is characteristic of' is by itself not sufficient evidence of an argumentative relation, since indicators of symptomatic relations in two directions do not establish an argumentative connection between the connected elements. That is exactly why they can be used in two directions when they *are* used to connect the minor premiss of an argument to the standpoint. The expression 'is characteristic of' can also be used merely descriptively, as in example (13):

(13) [From a book review]
This over-consciousness, of usage, but also of emotions, gestures and minimal changes in behavior *is characteristic of* this novel (*de Volkskrant*, 22 January 1999).

Even if a text is clearly argumentative, the indicators that have been mentioned here are not always decisive. Some of the weaker indicators can be used in more than one type of argument. Whether they really are an indication of symptomatic argumentation or of a different type of argument, may depend on their position in the argument, but in the analysis other conditions may also need to be taken into account.

In this paper, I have only discussed indicators of symptomatic argumentation. In our research project, we have also looked at clues in the verbal presentation for the two other types of argumentation schemes, causal argumentation and argumentation by analogy, and their subtypes. From Garssen's [3] empirical research on the recognition of argumentation schemes by ordinary language users, it has emerged that in particular distinguishing symptomatic argumentation from causal argumentation proves to be difficult in practice. By comparing the various clues for the different argumentation schemes, we argue that, especially in cases where there is room for doubt, it is possible to arrive at a more well-founded analysis of the type of argumentation at issue,

Let me illustrate some of the problems of analysis by taking one of the less strong indicators of symptomatic argumentation, 'it is clear that' as an example. A first condition for this expression to be indicative of symptomatic argumentation is that it should occur in the standpoint, not in the argumentation. If 'it is clear that' is part of the reasons, the argumentation may also be based on a causal relation or a relation of analogy. In example (14), for example, the argumentation is a pragmatic argument based on a causal relation:

(14) '*It is clear that* our economy suffers from the lack of confidence on the part of national and international investors, said Minister of Finance Thanong Bidaya. 'It should therefore be the first priority of this government to restore that confidence' (*de Volkskrant*, August 6, 1997).

As we have seen, some expressions only function as indicators of a particular relation if they occur in a specific part of the argumentation scheme (the standpoint, the major premise or the minor premise). But the position of the indicating device is also not always decisive. Even if the expression 'it is clear that' is part of the standpoint, the argumentation may still be causal, as in example (15):

(15) *It is clear that* the present system of schools with different denominations is going to founder. The number of types of schools keeps growing. You can already see it now: next year an evangelical school will open its doors, and the year after that probably an Islamic school (*de Volkskrant*, October 1, 1998).

In this example the standpoint consists of a prediction ('the present system of schools with different denominations is going to founder'), which is by itself an indication that the argumentation might be causal. The arguer supports this prediction by pointing at present and future developments that will lead to the failure of the present system. A further indication that the argumentation in this example is causal, and not symptomatic, is the fact that both the standpoint and the argument refer to processes or events, not to states of affairs. This is different in example (16), where 'it is clear that' functions as an indicator of a symptomatic relation:

(16) *It is clear that* the boy's behavior was very difficult indeed. One neighbor was reported as saying that he had threatened her with a knife when she tried to stop him throwing stones at dumped cars. (*The Sunday Times*, September 24, 2000)

In this example, the standpoint qualifies the boy's behaviour as 'very difficult', thereby referring not so much to a particular event but to a repeated pattern of behaviour or disposition, in other words to a static situation or state of affairs rather than an event. Next, a particular instance of the boy's behavior is mentioned as evidence of the fact that he has been behaving badly. So, 'it is clear that' can only be an indication of a symptomatic relation if the expression occurs in the standpoint and either the standpoint or the argument (or both) refers to a state of affairs.

6 Conclusion

Starting from an analysis of the main characteristics of the symptomatic relationship, I have discussed various types of clues for symptomatic argumentation. These clues are to be found in the presentation of the reasons and the standpoint, in the critical reactions and in the speaker's follow-up to his argument. Each of these verbal devices may provide a strong or a less strong indication that the argumentation may have to be reconstructed as symptomatic. As an illustration of the use of these presentational clues for symptomatic argumentation, I have given a number of examples, taken from various journals, in which these clues are present.

Some of the indicators of the symptomatic relation I discussed have already been mentioned in earlier pragma-dialectical publications. I have made an attempt to provide an explanation for the fact that these expressions can be a clue for identifying symptomatic argumentation and to specify the conditions that need to be fulfilled in order for the expressions to fulfill their indicative function. The clues that I have discussed range from expressions by means of which it is stated explicitly that the relation is symptomatic ('X is a sign of Y') to less unambiguous indications of the symptomatic relationship ('apparently') or expressions associated with aspects of the symptomatic relationship between the person and its manifestations ('is by nature'). The list of expressions I have discussed is, of course, by no means exhaustive.

As I hope to have made clear, for a well-founded reconstruction, apart from the indicating device, a number of factors need to be considered, among which the main characteristics of the argumentation scheme at issue and those of the alternative schemes, the part of the argumentation scheme in which the potential indicator occurs and the type of propositions that constitute the premisses and the standpoint. It is only by looking at the combination of these factors that the analysis of the relationship between argumentation and standpoint can be justified.

REFERENCES

[1] *Collins Cobuild English Grammar*, London/Glasgow: Collins, 1991.
[2] F.H. van Eemeren, and R. Grootendorst, *Argumentation, Communication and Fallacies. A Pragma-dialectical Perspective*, Hillsdale: Lawrence Erlbaum, 1992.
[3] B. Garssen, *Argumentatieschema's in pragma-dialectisch perspectief. Een theoretisch en empirisch onderzoek*, Amsterdam: IFOTT, 1997.
[4] S. Greenbaum, *The Oxford English Grammar*, Oxford: Oxford University Press, 1996.
[5] C. Perelman, *The Realm of Rhetoric*, Notre Dame/London: University of Notre Dame Press, 1982.
[6] *The Oxford English Dictionary*, 2nd ed. Oxford: Clarendon Press, 1989.
[7] R. Weaver, *The Ethics of Rhetoric*, Chicago: Henry Regnery, 1953.

Session III

Counterexamples and Degrees of Support

Claude Gratton[1]

Abstract. My goal is to present recent work in the logic of counterexamples that could be of value to experts working to create computer models of arguments in natural language.

A very crucial skill in the evaluation of an argument in natural language (which I will also refer to as a "natural argument") is the construction of counterexamples to assess the support of its premises for its conclusion. So, if a computational model of natural argument neglected the construction and evaluation of counterexamples, then it would be very seriously deficient. To my knowledge there have not been any publications, besides my own single publication (see [3]), on the logic of counterexamples in natural language. Of course some argumentation and critical thinking textbooks mention counterexamples, but they offer superficial suggestions. Argumentation can be an odd discipline because it sometimes discovers what needs to be investigated *after* critical thinking textbooks have been published. This paper represents my attempt to further explore the logic of counterexamples in natural language. I will first contrast two different kinds of counterexamples, and then use one of them to assess the degrees of support of premises for their conclusion. Since I know nothing about computational models or artificial intelligence, and most of the members of my audience work in at least one of these areas, I will not be able to present my ideas in a way that is familiar to you. However, I will attempt to present my work as clearly as possible and make occasional references in the paper where I suspect that particular challenges would arise for those who would venture to construct computer models of counterexamples.

It is very easy to assess the *validity* of many everyday arguments: we simply construct a counterexample by imagining a situation where all the premises are true and the conclusion false. However, the standard use of this technique is inadequate against arguments that are not intended to be valid. Most everyday arguments are not intended to provide conclusive support. In other words, for most everyday arguments, if all their premises were true, their conclusions would be intended to be probably true, but not necessarily true. Given the general ease of inventing counterexamples against the validity of an argument, I will explore the logic of such counterexamples in order to find a way of using them to assess degrees of support that are less than conclusive.

Since there are two basic kinds of counterexamples against the validity of arguments, and my investigation will apply to only one of them, I will first clarify the distinction between them. An accurate computational model of natural argument would need to take these distinctions into account. The counterexamples whose logic I will be examining are very different from counterexamples by analogy. No textbook author describes in any detail how they differ, but only a few do present them as being different (see [1, 2, 4, 5]). We can see their differences by comparing and contrasting them when they are advanced against the same invalid argument. Let that argument be:

(A) (1) Derrida will pass the logic course only if he registers for the course.
(2) He has registered for the course.
So, (3) Derrida will pass the logic course.

This argument has the form, (1) P only if Q. (2) Q. So, (3) P. The fact that this is an example of the formal fallacy of affirming a consequent, and that we would typically quickly reject the argument without using any kind of counterexample, is irrelevant. I am just using it as an example against which both kinds of counterexamples can be advanced. Once we have identified the logical form of an argument, a counterexample by analogy against that argument must have the same form, but have true premises and a false conclusion. The more obviously true the premises and obviously false the conclusion, the more effective is the counterexample by analogy in showing the invalidity of a particular form. I suspect that this would be a challenge for computer models of natural arguments because what is obviously true and obviously false will vary according to the knowledge, intelligence, and experience of one's audience. In this particular case we can advance the following counterexample by analogy against argument (A):

CE1[2] against argument (A):
(1) There's a fire in this room only if there's oxygen in this room.
(2) There's oxygen in this room.
So, (3) there's fire in this room.

Let us now contrast it to the next counterexample:

CE2 against argument (A)
It is possible that:
(1) Derrida will pass the logic course only if he registers for the course. **AND**
(2) He has registered for the course. **AND**
What if Derrida does not do adequate studying. **AND**
Not-(3): *It is not the case* that Derrida will pass the logic course.

Differences

Both counterexamples successfully show that argument (A) is invalid, in other words, they both show that its premises are not sufficient for its conclusion. However, there are some logically significant differences between them[3]

1. A counterexample by analogy is an *argument* analogous in form to the argument against which it is advanced. But a counterexample such as CE2 is *not* an argument, and so such a counterexample

---

[1] Department of Philosophy, University of Nevada, Las Vegas

[2] I will be using special notation to distinguish arguments and their counterexamples: "CE1 against argument (A)" simply means "counterexample number 1 against argument (A)", and "CE2 against argument (A)" means "counterexample number 2 against argument (A)".

[3] The following seven points were presented at the Eleventh NCA/AFA Conference on Argumentation in August 1999, and published in the refereed proceedings of that conference (see [3]).

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

cannot have the form of the argument against which they are advanced. A counterexample by analogy is *not* a mere conjunction of propositions. However, the kind of counterexample illustrated by CE2 is a mere possible conjunction of propositions. Accordingly, I propose that we name it a **"counterexample by possible conjunction"**. I invite anyone to propose a better descriptive label that will clearly differentiate this kind of counterexample from counterexamples by analogy.

2. In a counterexample by possible conjunction each premise of an argument is granted and unchanged (all the given reasons *as stated* are assumed to be true), and the argument's conclusion is negated. These two characteristics are necessary because the goal of a counterexample by possible conjunction is to show that all the given premises are not jointly sufficient for the truth of their conclusion. In contrast to these two characteristics, counterexamples by analogy, as illustrated by CE1, alter some of the content of the premises and conclusion, and they do not negate the conclusion.

3. In counterexamples by possible conjunction all the given premises of an argument and the negation of its conclusion are conjoined to a finite number of other statements, e.g. "What if Derrida does not do adequate studying" in CE2. These statements play the very important role of *making us understand how it is possible for all the given premises to be true and the conclusion false.*
   Why is this understanding so important? Though a counterexample by possible conjunction is not *in itself* an argument, it is evidence advanced to show to someone who has presented an argument that his/her premises are not sufficient. If a counterexample is not understood by the person presenting the argument, then s/he will not be convinced that the premises are not sufficient, in other words, s/he will not be convinced that his/her argument is invalid. Thus, understanding the counterexample, which involves understanding how it is possible for the argument's premises to be true and its conclusion false, is a necessary condition to show to an arguer that his/her argument is invalid. This is analogous to the construction of any argument: if the argument is not understood by its intended audience, then the it will not be convincing, even if it is impeccably logical and has necessarily true premises. This aspect of the construction of a counterexample is very context dependent: it will be effective generally only when it is sensitive to the level of knowledge, intelligence, and imagination of the person to whom the counterexample is presented. And these three factors affect one's level of understanding. It appears that computer models of natural arguments encounter again the challenge of context and audience dependence, but now there is the additional challenge of adequately representing the nebulous concept of understanding in computer models.
   Given this crucial role of the statements conjoined to the granted premises and negated conclusion to form a counterexample by possible conjunction, I need a convenient way to distinguish them from the premises and negated conclusion. I will thus sometimes label them by means of the letter "X". Since CE2 would typically be succinctly presented as "*What if* Derrida does not do adequate studying", and this common way of communicating this kind of counterexample focuses *exclusively* on the statements that make us understand how it is possible for all the given premises to be true and the conclusion false, I propose to name them the **"what-if-statements"** of the counterexample. Again, I invite anyone to propose a better label. In contrast to these counterexamples, no new statement is added to a counterexample by analogy.

4. The conjunction constituting the counterexample, $P_1 \& P_2 \ldots \& P_n \& X_1 \& X_2 .... \& X_n \& \tilde{} C$, is just presented as a logical possibility. However, as illustrated by CE1, a counterexample by analogy can have actually true premises and an actually false conclusion. I am wondering whether the notion of possibility can be easily represented in computer models of natural arguments. If not, there is another challenge here.

5. Counterexample CE2 has the specific form, *it is possible that* $P \& X \& \tilde{} C$. The general form of a counterexample by possible conjunction is, *it is possible that* $P_1 \& P_2 \ldots \& P_n \& X_1 \& X_2 \ldots \& X_n \& \tilde{} C$. Of course these conjuncts could be in any order, but I present them in this order because it is clearer, and because this order closely parallels the general structure of the argument against which it is advanced.
   In contrast, counterexamples by analogy do not have a common general logical form. For as illustrated by CE1, the form of a counterexample by analogy must correspond precisely to the form of the specific argument against which it is advanced, and of course there is no specific form common to all arguments. For example, not all arguments correspond in form to argument (A).

6. Counterexamples by possible conjunction help us to identify implicit assumptions of an argument. For example CE2 shows that argument (A) rests on the assumption that Derrida does or will do sufficient amount of studying. In other words, argument (A) assumes the *contradictory of the what-if-statement* in counterexample CE2. It must assume it in order to block counterexamples that use that specific what-if-statement. Such counterexamples are blocked because they must grant all the premises of the argument against which they are advanced; and if a reconstructed argument contains the negation of a what-if-statement as a premise, no counterexample can use that what-if-statement, and so such counterexamples are automatically eliminated. Counterexamples by analogy, on the other hand, do not identify any implicit assumptions. It seems that if a computer model could effectively construct counterexamples by possible conjunctions, it would be very easy to identify this kind of implicit assumption: it's simply the negation of the what-if statement.

7. The consequences of these two types of counterexamples are different. A successful counterexample by possible conjunction shows that the *specific premises*, $P_1 \& P_2 \ldots \& P_n$, are not sufficient for the truth of a *specific conclusion C*: these specific premises do not guarantee the truth of that specific conclusion. However, a successful counterexample by analogy shows that *the specific form* it expresses is invalid, and consequently, it proves that *any argument having its form* (and no other form that is valid)[4] is invalid. So, *no* premises of any argument having this form (and no other form that is valid) are sufficient for the truth of conclusion C.

[4] I include this parenthetical phrase in order to take into account the fact that an argument can have more than one form, and is usually considered valid if it has at least one valid form. For example, the argument, "All philosophers are human. All humans are mortal. So all philosophers are mortal." has at least two forms. If we consider only the propositions, there is the invalid form, "P. Q. So, R". But if we consider the quantifiers within the propositions, there is the valid form "All A are B. All B are C. So all A are C." This argument is valid even though it also has an invalid form.

From the preceding differences it follows that counterexamples by possible conjunction and by analogy are two very different kinds of counterexamples.

Consistency in Counterexamples by Possible Conjunction

I will next show that the *mere consistency* among the granted premise(s), the what-if-statement(s), and the negated conclusion in a counterexample by possible conjunction is not enough for the counterexample to show us that those premises are not sufficient for their conclusion. Consider the following counterexample against argument (B):

(B) (1) Winds are blowing a rain storm in our direction.
So, (C) it's going to rain here tomorrow.

CE3 against argument (B)
It is possible that:
(1) Winds are blowing a rainstorm in our direction. **AND**
What-if-statement: "Sirius" is the name of the closest star to our solar system. **AND**
Not-(C): It is not the case that it's going to rain here tomorrow.

The counterexample has the correct form, *it is possible that* $P\&X\&\tilde{}C$, and all the propositions are consistent, yet the counterexample fails to show us that the premise is not sufficient for the conclusion.

Contrast it to the next example:

CE4 against argument (B)
It is possible that:
(1) Winds are blowing a rainstorm in our direction. **AND**
What-if-statement: Strong winds from another direction are going to divert the storm away from us. **AND**
Not-(C) It is not the case that it's going to rain here tomorrow.

This counterexample is effective in proving to us that the premise is not sufficient for its conclusion. Since the only difference between counterexamples CE3 and CE4 is that it is only in the latter case that the what-if-statement makes us understand how it possible for the premise to be true and its conclusion false, then that understanding is a necessary condition for a counterexample to show us that premises are not sufficient for their conclusion. A discussion of the logic involved in making us understand how it is possible for premises to be true and their conclusion false is beyond the scope of this paper, and is not necessary in order to grasp the practical rudiments of this kind of counterexample. This particular logic will probably have to be well investigated if computer models of counterexamples are to be effective.

Counterexamples by possible conjunction and degrees of support

We have been examining some of the logic of the typical use of counterexamples by possible conjunction: to determine whether an argument is valid. Whenever a counterexample is successful, it proves that an argument's premises are not sufficient for (do not guarantee/necessitate) its conclusion. The serious limitation of this standard use is that the premises of most everyday arguments are not intended to be provide conclusive support, but rather only significant support. We will now explore a way to use these counterexamples to estimate the degree of support that is less than conclusive.

Elementary probability theory suggests a way to begin examining the logic of this additional role. I hope that my use of probability will also help you in your computer modeling of counterexamples.

(1) Pr($\tilde{}P$ or P) = 1.
(2) Pr($\tilde{}P$) + Pr(P) = 1.

We are looking for a substitution of "$\tilde{}P$" and "P" that will allow us to assess the degree of support of any argument, $P_1\&P_2\ldots\&P_n$, so C. Let the degree of support be expressed by the probability of C given $P_1\&P_2\ldots\&P_n$: $Pr(C \mid P_1\&P_2\ldots\&P_n)$. Replace both "$P$'s" in (2) by, $P_1\&P_2\ldots\&P_n\&\tilde{}C$:

(3) $Pr(\tilde{}(P_1\&P_2\ldots\&P_n\&\tilde{}C))+Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)=1.$

Subtract $Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)$ from both sides of the equation:

(4) $Pr(\tilde{}(P_1\&P_2\ldots\&P_n\&\tilde{}C)=1-Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C).$

Replace $(P_1\&P_2\ldots\&P_n\&\tilde{}C)$ in (4) by the logically equivalent expression, $(P_1\&P_2\ldots\&P_n\Rightarrow C)$, which stands for the relation of support that the premises bring to the conclusion:

(5) $Pr(P_1\&P_2\ldots\&P_n\Rightarrow C)=1-Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C).$

The standard way of expressing the premises' support for the conclusion is rather:

(6) $Pr(C\mid P_1\&P_2\ldots\&P_n)=1-Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C).$

On the right side of this equation $Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)$ represents the probability of all the counterexamples by possible conjunction against the argument $P_1\&P_2\ldots\&P_n$, so C, whose support is represented on the left side of the equation.

Formula (6) coincides with our intuitions. First, when there are no counterexamples, the formula derives what we would expect with a deductively valid argument, for when $Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)=0$, then $Pr(C\mid P_1\&P_2\ldots\&P_n)=1$: if the premises were true, the conclusion would also be true. Secondly, it entails that the greater the probability of all those counterexamples, the weaker the support (i.e. the smaller the probability of the conclusion given that all its premises are true), and the smaller the probability of all the counterexamples, the stronger the support for the conclusion. There is thus an inverse relation between the probability of the counterexamples and the strength of the support (the probability of the conclusion when all its premises are true). This inverse relation seems to be an aspect of this extended use of counterexamples that could be easily implemented in a computer model of natural arguments. But now we move to greater challenges.

How do we *estimate* the probability of $Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)$? Let us examine an everyday argument and various counterexamples against it.

C (1) Each student beginning my course is sufficiently intelligent to pass the course.
(2) So, each student beginning my course will pass it.

CE5 against argument (C)
It is possible that:

P: Each student beginning my course is sufficiently intelligent to pass the course. **AND**
X: What if at least one student will be sick too often to do all the necessary work to pass. **AND**
$\tilde{}C$: It is not the case that each student beginning my course will pass it.

Consider the following condensed counterexamples against argument (C). Assume that their what-if-statements, represented by "X", are conjoined to $P\&\tilde{}C$, and that the conjunction of all these statements forming each counterexample falls within the scope of the operator, "it is possible that", just as in CE5 against argument (C).

CE6(C) X: What if at least one students will not study material that must be studied to pass it.

CE7(C) X: What if at least one student has family responsibilities that very seriously interfere with his/her academic performance.

CE8(C) X: At least one student has personal problems that very seriously interfere with his/her academic performance.

CE9(C) X: What if the teacher will grade unfairly.

CE10(C) X_1: What if there is a personality conflict between the teacher and at least one student. **AND**
X_2: What if that student drops the course.

Regardless of the actual probability of any specific counterexample by possible conjunction, it is significantly smaller than the Pr(CE5(C) **or** CE6(C) **or** CE7(C) **or** CE8(C) **or** CE9(C) **or** CE10(C)). So, if we were to use the probability of only one counterexample to estimate the degree of support, and discard the probability of this disjunction of counterexamples, then we would significantly overestimate the degree of support of the premise - even if the chosen counterexample had the highest probability. Each counterexample must be included in our estimation of the degree of support because each one exposes other serious weaknesses in the support that would be overlooked even by the most probable counterexample. Since most everyday arguments are vulnerable to more than one counterexample with probabilities worth considering, we must take into account not just the most probable counterexample but also other probable counterexamples.

Hence, formula (6) can be restated more precisely as:

(7) $Pr(C \mid P_1\&P_2\ldots\&P_n) = 1 - Pr(\text{CE1orCE2orCE3...orCEn})$.

(I will address one of the challenges of estimating such a disjunction of probabilities later.) However, this added formulaic precision does not necessarily give us an accurate degree of the support of premises, for we very rarely have all the counterexamples against the support an argument, and consequently our estimation of the support is very rarely final and complete. This is a challenge not just for computer models of natural argument, but for anyone who wants a rough estimation of the degree of support.

Formula (7) can be further simplified. In any counterexample by possible conjunction all the given premises and the negation of the conclusion are assumed to be true:

$$Pr(P_1\&P_2\ldots\&P_n) = Pr(\tilde{}C) = 1.$$

Since the probability of a typical counterexample is,

$$Pr(P_1\&P_2\ldots\&P_n\&X_1\&X_2\ldots X_n\&\tilde{}C) = Pr(P_1\&P_2\ldots\&P_n)xPr(X_1\&X_2\ldots X_n)xPr(\tilde{}C),$$

then

$$Pr(P_1\&P_2\ldots\&P_n\&X_1\&X_2\ldots X_n\&\tilde{}C) = Pr(X_1\&X_2\ldots X_n).$$

Hence, when talking about the probability of a counterexample by possible conjunction, we are talking about the probability of the conjunction of its what-if-statements. Therefore, (7), which includes more than one counterexample against a the support of an argument, can be more simply formulated as:

(8) $Pr(C \mid P_1\&P_2\ldots\&P_n) = 1 - Pr(X1_1\&X1_2\ldots\&X1_n or X2_1\&X2_2\ldots\&X2_n\ldots or Xn_1\&Xn_2\ldots\&Xn_n)$

There is a further challenge for natural language users and computer experts to meet when using counterexamples by possible conjunction to estimate the degree of support of premises: they must determine when to stop constructing counterexamples. For instance, we could have continued inventing more counterexamples against the support of argument (C). If we wanted to have a reliable estimate of the degree of support that (C)'s premises give to its conclusion, where should we stop? Assuming that time is not an obstacle, we stop when we can only invent extremely unlikely counterexamples, and we have reason to believe that we would continue inventing only such unlikely ones. Here is an example of an extremely improbable counterexample:

CE11 against argument (C)
It is possible that:
P: Each student beginning my course is sufficiently intelligent to pass the course. **AND**
X: What if at least one student is abducted by an extraterrestrial at the beginning of the course. **AND**
$\tilde{}C$: It is *not* the case that each student beginning my course will pass it.

We stop when we can construct only very unlikely counterexamples because they add nothing significant to the probability of the disjunction of all the realistic counterexamples we have already constructed. It is important to bear in mind that wherever we stop, it will be due to our limited knowledge and imagination. So, we can *never* be sure that we have taken into consideration *all* the counterexamples that are represented by $Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)$ in formula (6) $Pr(C \mid P_1\&P_2\ldots\&P_n) = 1 - Pr(P_1\&P_2\ldots\&P_n\&\tilde{}C)$. For this reason, it is sometimes important to persist inventing a few the wildly imaginative counterexamples because sometimes that process can help us to discover more realistic ones.

There is a further practical challenge in determining $Pr(X1_1\&X1_2\ldots\&X1_n or X2_1\&X2_2\ldots\&X2_n\ldots or Xn_1\& Xn_2\ldots\&Xn_n)$: not all counterexamples (or more simply, not all what-it-statements) are independent of one another. Event M is independent of event N if and only if N does not affect the probability of M: if and only if Pr(M, given N) = Pr(M). For instance, if I am boarding a taxi for a destination that is five miles away, and I infer from my taking the taxi that I will arrive at my destination in less than an hour, there are many interdependent counterexamples against the inference: what if there is an accident;

what if there is a flat tires; what if the driver becomes sick. These are different physical possibilities that could prevent me from reaching my destination on time, and they are partly interdependent: some accidents are caused by flat tires, and some accidents are caused by a driver's illness. If we are to continue using probability theory, then matters would get complicated. For if events M and N are not independent, then Pr(M or N) = Pr(M) + Pr(N) - Pr(M & N), and so the estimation of the probability of my arriving on time would need to include the probability of a flat or an accident, which equals the sum of the probability of a flat and the probability of an accident, minus the probability of the conjunction of a tire having a flat and the taxi driver having an accident. Given the interdependence of many daily counterexamples, the costs, in terms of time, mental energy, and possibly even money, of this further application of probability theory would seem to outweigh the benefits.

How would we estimate the probability of the counterexamples against argument (C) if we also estimated their interdependence? My estimation of the probability of the disjunction of those six counterexamples, without considering their interdependence, is that it is *at least* moderately probable. Consequently, the probability of the conclusion that each student beginning my course will pass it is *at most* moderately improbable. What would I change if I now take into consideration the interdependence of the counterexamples? The overall combined probability of all the counterexamples would have to diminish, and there would be a corresponding increase in the strength of the support. What would be the amount of that change? My estimation is that the probability of the conjunction of all those counterexamples would still be roughly at least moderately probable. So my consideration of the interdependence makes me only qualify my estimation with "roughly".

If this ordinary example is representative of most everyday examples, then for practical everyday purposes, will considerations of the interdependence of counterexamples be useful?

In most situations we don't have the information or the time to figure out Pr(M & N), it is challenging enough just to estimate Pr(M) and Pr(N). However, knowing that some counterexamples against the support of an argument are interdependent makes us aware that the disjunction of the counterexamples' probabilities is in fact less than the sum of their individual probabilities, thereby indicating from the inverse relation that the support of the premises is stronger than initially estimated. The greater the interdependence between counterexamples (i.e. the greater the probability of one given the other) against the support of an argument, and the greater the number of interdependent counterexamples, the smaller the sum of the counterexamples' individual probabilities; and consequently, the stronger the support of the argument's premises. It is possible that in some cases, the interdependence might be significant and easy to estimate, thus we might easily realize the significant decrease of the probability of a disjunction of counterexamples against an argument. Though for most everyday purposes these consideration will be beyond our knowledge and available time, it might be prudent in some cases to raise the questions, "Are there any interdependent counterexamples? To what degree are they interdependent?". By realizing the extent of interdependence and the number of the counterexamples, we come to see that the degree of support of premises is stronger than what we might have initially estimated. In order not to underestimate the degree of support that premises bring to their conclusion, it might be prudent in some cases to raise the question, "How many counterexamples are interdependent? To what degree are they interdependent?".

In this paper I identified two kinds of counterexamples: counterexamples by possible conjunction and counterexamples by analogy; described the logical differences between them; examined some of the logic of counterexamples by possible conjunction. It is a logically possible conjunction of all the premises of an argument (all are assumed true), the conclusion is negated, and one or more statements, named "what-if-statements". I showed that the latter have the very special function of making the proponent of an argument understand how it is possible for his/her premises to be true and conclusion false; argued that the mere consistency of all the statements constituting these counterexamples is not sufficient for the success of these counterexamples. I used elementary probability theory to justify extending the use of these counterexample to estimate the degree of premise support that is less than conclusive; showed that the strength of support (i.e. the probability of a conclusion given its premises) is inversely proportional to the probability of the disjunction of all the what-if-statements of successful counterexamples against the support; described where we should stop in the construction of counterexamples; illustrated some of the practical limitation of considering the interdependence of what-if-statements when estimating the probability of their disjunction.

If computer models of arguments in natural language are to be successful, they must be able to model all the natural and effective ways of assessing the support of premises. The construction of counterexamples by possible conjunction is a natural and effective way of assessing the sufficiency of premises (i.e., assessing the validity of an argument), and they can be used to estimate the degree of support that is less than conclusive. So, computer models of arguments in natural language should attempt to model the construction, use, and evaluation of these counterexamples. Therefore, programmers face the following challenges:

1. The models must distinguish counterexamples by possible conjunction and counterexamples by analogy.
2. The models must represent the concept of possibility.
3. The models must identify effective what-if statements of counterexamples by possible conjunction against an argument. This identification will depend on the models' ability to determine whether the what-if statements make the proponents of the argument understand how it is possible for all their premises to be true and their conclusion false. So the models must (a) handle the nebulous concept of understanding. They must also (b) be very context sensitive, for the understanding of an audience varies according to its knowledge, experience, and imagination.
4. The models must estimate the probability of the effective what-if statements of each counterexample, and estimate the disjunction of all the probabilities of the effective what-if statements against the same argument.
5. The models must determine when it is appropriate to consider the dependence among effective what-if statements, and how their dependence affects the combined probability of all the effective counterexamples against an argument.
6. The models must determine when it is appropriate to stop constructing counterexamples by possible conjunction.

Since I do not want to discourage any of you from investigating the modeling of these counterexamples, I would like to end by identifying two areas where programmers would probably *not* face any serious challenges:

7. It will be easy for models to identify certain kinds of implicit assumptions. For once an effective what-if statement of a counterex-

ample is identified, it follows that the argument against which the counterexample is advanced assumes the negation of that what-if statement.

8. At a certain stage it will be easy to estimate the degree of support of premises. For when what-if statements are independent, or when their dependence is insignificant, the probability of a conclusion is simply 1 minus the estimated combined probability of the effective what-if statement of each counterexample.

REFERENCES

[1] Cederblom, Jerry, Paulsen, David W. *Critical Reasoning*, 4th ed. Belmont, CA: Wadsworth, 2001.

[2] Feldman, Richard. *Reason & Argument*, Upper Saddle River, NJ: Prentice-Hall, 1999.

[3] Gratton, Claude. Counterexamples by Conjunction and Counterexamples by Analogy: Some Overlooked Logically Significant Differences In Thomas A. Hollihan, editor, *Argument at Century's End, Reflecting on the Past and Envisioning the Future*, pages 109–113, 2000, Annandale VA: National Communication Association.

[4] Rudinow, Joel, Barry, Vincent E. *Invitation to Critical Thinking*, 3rd ed., Forth Worth: Harcourt Brace College Publishers, 1994.

[5] Wilson, David C. *A Guide to Good Reasoning*, Boston: McGraw-Hill, 1999.

Argumentation within Deductive Reasoning

Armin Fiedler and **Helmut Horacek**[1]

Abstract. Deductive reasoning is an area related to argumentation where machine-based techniques, notably theorem proving, can contribute substantially to the formation of arguments. However, making use of the functionality of theorem provers for this issue is associated with a number of difficulties and, as we will demonstrate, requires considerable effort for obtaining reasonable results. Aiming at the exploitation of machine-oriented reasoning for human-adequate argumentation in a broader sense, we present our model for producing proof presentations from machine-oriented inference structures. Capabilities of the model include adaptation to human-adequate degrees of granularitiy and explicitness in the underlying argumentation and interactive exploration of proofs. Enhancing capabilities in all these respects, even just those we have addressed so far, does not only improve the interactive use of theorem provers, but they are essential ingredients to support the functionality of dialog-oriented tutorial systems in formal domains.

1 Introduction

Deductive reasoning is an area related to argumentation where machine-based techniques, notably theorem proving, can contribute substantially to the formation of arguments. However, making use of the functionality of theorem provers for this issue is associated with a number of difficulties and, as we will demonstrate, requires considerable effort for obtaining reasonable results.

Aiming at the exploitation of machine-oriented reasoning for human-adequate argumentation in a broader sense, we present our model for producing proof presentations from machine-oriented inference structures. Capabilities of the model include adaptation to human-adequate degrees of granularitiy and explicitness in the underlying argumentation and interactive exploration of proofs. However, this model has inherent limitations in its argumentative behavior, since arguments giving motivations or justifications on a more strategic or dynamic perspective cannot be obtained from machine-found proofs. Enhancing capabilities in all these respects does not only improve the interactive use of theorem provers, but they are essential ingredients to support the functionality of dialog-oriented tutorial systems in formal domains.

This paper is organized as follows. We first provide some background information about presentation of machine-found proofs in natural language, including empirical motivations that substantiate divergent demands for human-adequate presentations. We describe techniques for building representations meeting these psychological requirements in a formal model, comprising some kinds of proof transformation and adaptations. We illustrate the functionality of our model by discussing a moderately complex example. Finally, we sketch some limitations of our model.

[1] University of the Saarland, P.O.Box 151150, D-66041, Saarbrücken, F.R.G.

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

2 Background

2.1 Proof Presentation in Natural Language

The problem of obtaining a natural language proof from a machine-found proof can be divided into two subproblems: First, the proof is transformed from its original machine-oriented formalism into a human-oriented calculus, which is much better suited for presentation. Second, the transformed proof is verbalized in natural language.

Since the lines of reasoning in machine-oriented calculi are often unnatural and obscure, algorithms (see, e.g., [1, 18]) have been developed to transform machine-found proofs into more natural formalisms, such as the *natural deduction (ND) calculus* [8]. ND inference steps consist of a small set of simple reasoning patterns, such as forall-elimination ($\forall x P(x)$ leads to $P(a)$) and implication elimination, that is, modus ponens. However, the obtained ND proofs often are very large and too involved in comparison to the original proof. Moreover, an inference step merely consists of the syntactic manipulation of a quantifier or a connective. [15] gives an algorithm to abstract an ND proof to an *assertion level* proof, where a proof step may be justified either by an ND inference rule or by the application of an assertion (i.e., a definition, axiom, lemma or theorem).

One of the earliest proof presentation systems was introduced by Chester [2]. Several theorem provers have presentations components that output proofs in pseudo-natural language using canned text (e.g., [3, 4]). Employing several isolated strategies, the presentation component of THINKER [5] was the first system to acknowledge the need for higher levels of abstraction when explaining proofs. PROVERB [16] expresses machine-found proofs abstracted to the assertion level and applies linguistically motivated techniques for text planning, generating referring expressions, and aggregation of propositions with common elements. Drawing on PROVERB, we have developed the interactive proof explanation system *P.rex* [7], which additionally features user adaptivity and dialog facilities. [10] presents another recently developed NLG system that is used as a back end for a theorem prover.

In order to produce reasonable proof presentations, many systems describe some complex inference steps very densely, and they leave certain classes of proof steps implicit in their output, for example, by abstracting from intermediate inference steps that are recoverable from inductive definitions, or by omitting instantiations of axioms. However, leaving out information on the basis of purely *syntactic* criteria, as this has been done so far, easily leads to incoherent and hardly understandable text portions. In order to get control over the inferability and comprehensibility in presenting inference steps, an explicit model is required that incorporates semantic and pragmatic aspects of communication, which is what we try to achieve by our approach.

2.2 Empirical Motivation

Issues in presenting deductive proofs, as a special case of presenting argumentative discourse, have attracted a lot of attention in the fields of psychology, linguistics, and computer science. Central insights relevant to deductive argumentation are the following:

- Logical consequences of certain kinds of information are preferably conveyed implicitly, through relying on capabilities of the audience to exploit the discourse context and default expectations.
- Human performance in comprehending deductive syllogisms varies significantly from one syllogism to another.

The study in [23] demonstrates that humans easily uncover missing pieces of information left implicit in discourse, most notably in sequences of events, provided this information conforms to their expectations in the given context. Similarly to the expectations examined in that study, which occur frequently in everyday conversations, a number of elementary and very common inferences are typically left implicit in mathematical texts, too, including straightforward instantiations, generalizations, and associations justified by domain knowledge.

Another presentation aspect is addressed by studies on human comprehension of deductive syllogisms (see the summary in [17]). These studies have unveiled considerable performance differences among individual syllogisms (in one experiment, subjects made 91% correct conclusions for modus ponens, 64% for modus tollens, 48% for affirmative disjunction, and 30% for negative disjunction). The consequences of this result are demonstrated by the elaborate essay in [24], which presents a number of hypotheses about the impacts that human resource limits in attentional capacity and in inferential capacity have on dialog strategies. These hypotheses are acquired from extensive empirical analysis of naturally occurring dialogs and, to a certain extent, statistically confirmed. One that is of central importance for our investigations says that an increasing number of logically redundant assertions to make an inference explicit are made, in dependency of how hard and important an inference is (modus tollens being an example for a hard inference which requires a more detailed illustration).

However, these crucial issues in presenting deductive reasoning are insufficiently captured by current techniques, which typically suffer from two kinds of deficits:

- A large number of easily inferable inference steps is expressed explicitly.
- Involved inferences, though hard to understand, are presented in single shots.

The first deficit suggests the omission of contextually inferable elements in the proof graph, and the second demands the expansion of compound inference steps into simpler parts. We illustrate the appearance of these deficits and measures to remedy them in the subsequent sections.

3 An Example

Throughout this paper, we will use the proof of a well-known problem, Schubert's Steamroller [22], to demonstrate the functionality of our presentation model:

Axioms:

(1) Wolves, foxes, birds, caterpillars, and snails are animals, and there are some of each of them. Also there are some grains, and grains are plants.

(2) Every animal either likes to eat all plants or all animals much smaller than itself that like to eat some plants.

(3) Caterpillars and snails are much smaller than birds, which are much smaller than foxes, which in turn are much smaller than wolves. Wolves do not like to eat foxes or grains, while birds like to eat caterpillars, but not snails. Caterpillars and snails like to eat some plants.

Theorem:

(4) Therefore there is an animal that likes to eat a grain-eating animal.

Proving that theorem (4) is based on applying given pieces of simplified real world knowledge (1) to (3).

In a nutshell, the proof runs along the following lines: Through applying axiom (2) three times, it is first derived that birds eat plants, then that foxes do not eat grains and, finally, that foxes eat the smaller grain-eating birds, the last being the witness needed to prove theorem (4).

Within the theorem proving community, the Steamroller problem is famous, because solving it requires several variables to be instantiated purposefully without having a guidance how to do this through the formulation of the theorem to be proved — it has only existentially quantified variables in it, but no constants. Until some years ago, automated theorem provers were unable to apply this technique with sufficient degrees of efficiency, so that they were originally unable to solve this problem. For our purposes, this problem is attractive for completely different reasons: its definition is easily comprehensible without mathematical knowledge, and a full-detailed solution path is sufficiently complex so that exploring it interactively seems to be well motivated.

4 Our Model of Argument Building

In order to meet the deficits identified when discussing empirical motivations, we propose the application of an optimization process that enhances an automatically generated proof at the assertion level. Through this process, pragmatically motivated expansions, omissions, and short-cuts are introduced, and the audience is assumed to be able to mentally reconstruct the details omitted with reasonable effort. In a nutshell, the modified proof graph is built through two subprocesses:

- *Building expansions*
  Compound assertion level steps are expanded into elementary applications of deductive syllogisms, while marking the original larger steps as summaries.
- *Introducing omissions and short-cuts*
  Shorter lines of reasoning are introduced by skipping individual reasoning steps, through omitting justifications (marked as inferable) and intermediate reasoning steps (marking the 'indirect' justifications as short-cuts).

4.1 Levels of Abstraction

The purpose underlying the expansion of assertion level steps is to decompose presentations of complex theorem applications or involved applications of standard theorems into easier comprehensible pieces. This operation is motivated by performance difficulties humans typically have in comparable discourse situations. At first, assertion level steps are completely expanded to the natural deduction (ND) level according to the method described in [15]. Thereafter, a partial recomposition of ND steps into inference steps encapsulating the harder comprehensible deductive syllogisms, modus tollens and disjunction elimination steps, is performed, in case the sequence of

ND rules in the entire assertion level step contains more than one of these. To do this, the sequence of ND rules is broken after each but the last occurrence of a modus tollens or disjunction elimination, and the resulting subsequences of ND steps are composed into a sequence of reasoning steps at some sort of *partial assertion* level. This sequence is then inserted in the proof graph as a potential substitute for the original assertion level step, which is marked as a *summary*.

An example for such an expansion and partial recomposition is shown in Figure 1, which exposes a crucial inference in the Steamroller proof in two levels of abstraction. Both variants show subproofs indirectly deriving the categorization of the fox (f) as a meat eater, that is, the fox f does not eat grain g, $\neg EATS(f, g)$.

When the derivation is carried out by a single assertion level step ((1) in Figure 1), this can be paraphrased by 'The wolf either eats grain or, in case the fox eats grain and is smaller than the wolf, the wolf eats the fox. Since the wolf does not eat grain, the wolf does not eat the fox, and the fox is smaller than the wolf, it follows that the fox does not eat grain'. Apparently, this is a very bad argumentation. Though the facts mentioned provide a complete account of the justifications underlying the required reasoning, the way how this works is completely obscure at first sight. However, this is not surprising, since the assertion level step underlying this reasoning is composed of several cognitively complex inference steps, as the expansion to the ND level ((2) in Figure 1) demonstrates. In the general case, this expansion would be followed by a recomposition encompassing cognitively simple deductive syllogisms, yielding a representation on the partial assertion level. Since there are only cognitively difficult inference steps in this instance, the representations on ND and partial assertion levels are identical. Through this expansion, the compound inference step is decomposed into three simpler ones, two disjunction eliminations with a modus tollens in between. The sequence of inference steps can be paraphrased by 'Since wolves do not eat grain, it follows that wolves like to eat all animals smaller than themselves that like to eat plants. Since wolves do not eat foxes, it follows that foxes do not eat grain or that they are not smaller than wolves. Since foxes are smaller than wolves, it follows that foxes do not eat grain.' With more skillful references to instantiations of the central axiom of this problem, this text can be improved to 'Since wolves do not eat grain, their eating habits imply that they are meat eaters. Since they do not eat foxes, it follows that foxes are not plant eaters or not smaller than wolves. Since foxes are smaller than wolves, foxes are not plant eaters, hence they are meat eaters' (see [14] for details on how these referring expressions are built).

4.2 Degrees of Explicitness

Unlike expanding summaries, creating omissions and short-cuts is driven by communicatively motivated *presentation rules*. They express aspects of human reasoning capabilities with regard to contextually motivated inferability of pieces of information on the basis of explicitly mentioned facts and relevant background knowledge [9]. These rules provide an interface to stored assumptions about the intended audience. They describe the following sorts of situations:

Cut-prop: omission of a *proposition* (premise) appearing as a reason
Cut-rule: omission of a *rule* (axiom instance) appearing as a method
Compactification: short-cut by omitting an *intermediate* inference step

These *reduction* rules aim at omitting parts of a justification that the audience is considered to be able to infer from the remaining justification components of the same line of the proof, or even at omitting an entire assertion level step that is considered inferable from the adjacent inference steps. In order for these rules to apply successfully, presentation preferences and conditions about the addressees' knowledge and inferential capabilities are checked.

The functionality of the reduction rules can be explained by a simple example. If trivial facts, such as $0 < 1$, or axioms assumed to be known to the audience, such as *transitivity*, appear in the set of justifications of some inference step, they are marked as *inferable* ($0 < 1$ through *Cut-prop*, and *transitivity* through *Cut-rule*, provided the use of an axiom is likely to appear evident from the instantiated form). Consequently, the derivation of $0 < a$ can simply be explained by $1 < a$ to an informed audience. Moreover, single facts appearing as the only non-inferable reason are candidates for being omitted through applying *Compactification*. If, for instance, $0 < a$ is the only non-inferable reason of $0 \neq a$, and $0 < a$, in turn, has only one non-inferable reason, $1 < a$, the coherence maintaining similarity between $0 < a$ and $1 < a$ permits omitting $0 < a$ in the argumentative chain. Altogether, $0 \neq a$ can be explained concisely by $1 < a$ to an informed audience.

For problems such as the Steamroller, which make reference to (pseudo-)real world knowledge, similar expectation-based omissions and short-cuts occur. For example, mentioning the size relation between two animals as an argument can be omitted, as in 'It follows that foxes are not plant eaters or not smaller than wolves. Hence, foxes are not plant eaters.' (an instance of a *Cut-prop*).

Let us look into more detail on how the inferential capabilities and assumptions about the background knowledge are expressed. Modeling these mental capabilities is done by distinguishing the following sorts of knowledge and communicative competence:

- knowledge per se, comprising (static) domain knowledge and (dynamic) referential knowledge,
- the attentional state of the addressee, determined by the pieces of knowledge in the current focus of attention,
- inferential skills, which comprise abilities to draw taxonomic, logical, and communicatively adequate inferences. The last kind of inferences concerns the capability to augment logically incomplete pieces of information in a given context.

The first component as well as taxonomic inferences are fairly standard, while logical inferences are a novel part in our model. Its operationalization, however, needs to reflect particularities of the domain. In our application, we use some simple stereotypes to express assumptions about the addressee's domain knowledge (see [6]). Domain knowledge is composed of the addressee's acquaintance with mathematical theories in terms of axioms, definitions, and associated hierarchical relations, while referential knowledge is incrementally built from the assertions made in the course of a proof presentation. For example, if a proof makes reference to a mathematical group, a competent addressee is immediately aware that there are unit and inverse elements in this group because they belong to the definition of groups, and he/she also knows the associated definitions. Moreover, if the proof mentions a subgroup, the addressee is also aware of the fact that the properties of ordinary groups apply to it. Consequently, proof presentation can directly make reference to these propositions without mentioning explicitly the underlying connections that are entailed in the explicit content representation. Thus, taxonomic inferences comprise the following kinds of reasoning:

- Propagating properties of mathematical objects along hierarchical relations.

$$(1) \quad \frac{(EATS(w,g) \lor (EATS(f,g) \land (f < w)) \Rightarrow EATS(w,f)) \quad \neg EATS(w,g) \quad \neg EATS(w,f) f < w}{\neg EATS(f,g)}\text{Assertion}$$

$$(2) \quad \cfrac{\cfrac{\cfrac{(EATS(w,g) \lor (EATS(f,g) \land (f < w)) \Rightarrow EATS(w,f)) \quad \neg EATS(w,g)}{(EATS(f,g) \land (f < w)) \Rightarrow EATS(w,f)} \lor E \qquad \neg EATS(w,f)}{\neg EATS(f,g) \lor \neg (f < w)}\text{Modus Tollens} \qquad (f < w)}{\neg EATS(f,g)} \lor E$$

Figure 1. An involved assertion level inference at two different levels of abstraction.

- Expanding componential properties of mathematical objects.

The remaining components of our model, awareness and logical inferences, are expressed by the predicates AWARE-OF, COHERENT, and ABLE-INFER which are given domain-specific interpretations, elaborated for the domain of mathematics (formal details are given in [12]). For assessing the addressee's awareness (AWARE-OF), we test whether a piece of knowledge required is entailed in a list of theorems, definitions, and hierarchical relations assumed to be known to the addressee, which is expressed in a user model as simple stereotypes (see [6]). The underlying simplifying assumption is that being acquainted with some piece of generic knowledge is sufficient to be aware of it in the course of the entire proof. Inferential capabilities (ABLE-INFER) express whether a user is able to infer the missing pieces of knowledge to justify some conclusion, given only a subset of the premises. This reasoning process is approximated by the requirements that (1) composing the information given is sufficient to fully instantiate the entire inference step, and (2) matching the instantiated form with the relevant generic piece of knowledge is within the complexity limitations the addressee is assumed to be able to handle. The following inferential skills are distinguished, with limitations on the complexity of their applications:

- Generalizations of natural categories and instantiations of basic everyday knowledge; pieces of this sort of knowledge are represented as axioms in mathematical problems.
- And-eliminations to obtain an element on top level of a conjunction.
- Applications of modus ponens without any additional equivalence operations.
- Substitutions in axioms with constants or variables and at most one additional operator (such as a factor, or an exponent) replacing corresponding variables in generic expressions.
- Chaining inference steps with structurally identical conclusions, which differ only by constants or operators (operators must be related, such as '=' and '$<$').

The first three inferential skills are attributed to every user, the remaining ones only to users with some experience in mathematics.

A further issue to consider is the composition of such inference steps, which reflects the concept of coherence. According to psychological experiments, leaving out intermediate steps in a chain of argumentation should still be understood as a "direct" cause, while "indirect" causes negatively affect the reasoning effort [23]. In a previous approach to expert system explanations, this aspect has been modeled by requiring purposes of domain rules involved to be identical [11]. For proofs, we try to capture this coherence requirement by a structural similarity between intermediate and final conclusions: they must be joined by instantiation, generalization, part, or abstraction relations. Precise definitions for a larger set of operators and validation by associated empirical tests are still to be carried out. However, mentally inserting the missing pieces of information into a condensed representation in these sorts of situation is not without limitations. For example, the number of elements in a conjoined expression and its given presentation certainly influence the effort to pick a specific element, and the complexity of the substitution needed to obtain a required instantiation of some axiom or parts of an axiom may make this inference difficult. Hence, understanding the relation between expressions that are transducible into one another by the subsequent application of a substitution and several equivalence operations requires the exposition of some intermediate steps. For an extensive study examining the consequences of human memory limitations on the suitability of discourse contributions, see [24].

Applying the presentation rules to optimize the entire proof graph from an argumentative perspective is carried out in two processing cycles. In each cycle, the proof graph is traversed by starting from its leaf nodes and successively continuing to the root node, without back-tracking (that is, some sort of inverse depth-first search is invoked): In cycle one, *Cut-prop* and *Cut-rule* apply, marking locally inferable justifications. In cycle two, *Compactification* is invoked, adding alternative justifications through short-cuts, on the basis of the inferables. This order takes into account dependencies among the rules. It is also reasonably efficient, since only short-cuts require processing alternative lines of reasoning.

4.3 Interactive Exploration

In order to convey the information specified completely in view of the assumptions made about the audience, summaries are avoided and inferables are omitted. Depending on the target item, giving such an explanation in all details required for full understanding may result in a long text. Therefore, it is better to present a reduced first-shot contribution, which can be further investigated interactively, according to user reactions.

The associated communication is handled by *P.rex* by allowing three types of user interaction: A *command* tells the system to fulfill a certain task, such as explaining a proof. An *interruption* interrupts the system to inform it that an explanation is not satisfactory or that the user wants to insert a different task. In clarification dialogs, finally, the user is prompted to give *answers* to questions that *P.rex* asks when it cannot identify a unique task to fulfill. In this paper, we concentrate on interruptions.

The user can interrupt *P.rex* anytime to enter a new command or to complain about the current explanation. The following speech acts are examples for messages that can be used to interrupt the system:

`(too-detailed :Conclusion` C`)`
: The explanation of the step leading to C is too detailed, that is, the step should be explained at a more abstract level.

`(too-abstract :Conclusion` C`)`
: The explanation of the step leading to C is too abstract, that is, the step should be explained in more detail.

`(too-implicit :Conclusion` C`)`
: The explanation of the step leading to C is too implicit, that is, the step should be explained more explicitly.

`(too-difficult :Conclusion` C`)`
: The explanation of the step leading to C is too difficult.

---

(1) *P.rex*: Hence, $a \in U \cup V$ by the $\cup$-Lemma.

(2) User: This step is too difficult.

(3) *P.rex*: Since $a \in U$ or $a \in V$, $a \in U \cup V$ by the $\cup$-Lemma. Do you understand this step?

(4) User: No.

(5) *P.rex*: Do you understand the premises?

(6) User: Yes.

(7) *P.rex*: In order to prove that $a \in U \cup V$ let us consider the following cases.

Case 1: Let $a \in U$. That leads to $a \in U \cup V$ by the definition of $\cup$.

Case 2: Let $a \in V$. Then $a \in U \cup V$ by the definition of $\cup$.

Do you understand this step?

(8) User: Yes.

Figure 2. A clarification dialog in *P.rex*.

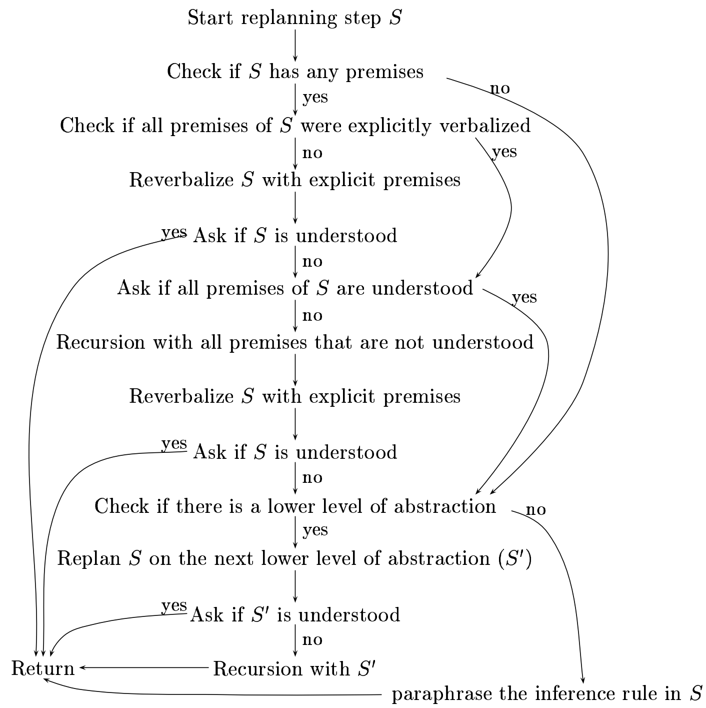


Figure 3. The reaction of the dialog planner if a step S was too difficult.

In *P.rex*, `too-difficult` is considered as an underspecified interruption. Figure 2 gives an example: When the user complains that the derivation of the conclusion in utterance (1) was too difficult, the dialog planner enters a clarification dialog to find out which part of the explanation failed to remedy this failure. During the clarification dialog, the system tries to distill whether the user failed to follow some implicit references (utterance (3)), whether one of the premises is unclear (utterance (5)), or whether the explanation was too abstract (utterance (7)). The control of the behavior of the dialog planner is displayed in Figure 3.

When generating a first-shot description, all possible reductions amount to relaxing the degree of completeness in which the information is presented. Four alternatives are examined, in ascending order of increasing information reduction:

1. Omitting the way how a piece of knowledge (a domain regularity) is applied.
2. Omitting that piece of knowledge.
3. Omitting premises of the inference (eventually, only some of them).
4. Omitting intermediate inference steps.

The choice among these options is based on assumptions about the audience and on the resulting balance of textual descriptions. In [13] we have defined and motivated some strategies for that, examples will be given in the next section.

When one or several intermediate inference steps are omitted (option 4 in the above list of items), some sort of ad-hoc abstraction is carried out. The sequence of enclosing inferences is abstracted into a set of propositions consisting of its conclusion and its premises, while the method how the conclusion is obtained, that is, the underlying sequence of inferences, is omitted. If there is evidence that some of the premises are more important or of more interest to the audience than the remaining ones, larger sets of premises can be reduced to subsets of these. In particular, this measure comprises preferring summaries over detailed exposition of involved inference steps. Moreover, in case these inferences constitute the expansion of a pre-designed proof method [19], which underlies the construction of a partial proof, the functionality of that method can be expressed by a descriptive phrase.

5 Explaining the Steamroller Proof

In this section, we demonstrate two strategies of building one-shot presentations of the solution to the Steamroller problem. In the examples, we paraphrase the expected output focusing on the structure and content of the produced text. Apparently, the proof sketch given when introducing the Steamroller is far from being a complete and fully comprehensible explanation of the proof, since many details that are necessary to understand how the central axiom is applied in each case are not mentioned. On the other hand, a full exploration of the proof is inappropriate for interactive environments because of its length.

A full description of the proof (see Figure 4) is produced by introducing a basic structure according to the main proof steps. These proofs steps, which are easily recognizable in the underlying proof graph, are routed in the application of those domain rules, which are not part of the addressee's background knowledge. In our example, only the rule about the food of animals is considered to be of this

It is first derived that foxes do not eat grain. This ultimately follows from the assumptions that wolves do not eat grain and foxes are smaller than wolves, because animals who do not eat plants eat plant eaters smaller than themselves. Thus, either foxes do not eat grain or they are not smaller than wolves. Hence, only the first alternative is valid. Moreover, it is derived that birds eat grain because animals eat plants if they do not eat plant eaters smaller than themselves. Birds do not eat plant eaters because it is assumed that they do not eat snails, but snails are smaller than birds and they eat plants. Finally, it is derived that foxes eat birds, because animals either eat plants, which foxes don't do, or they eat plant eaters smaller than themselves. Birds are such plant eaters, and they are smaller than foxes. Since foxes eat birds, an animal is known that eats a grain-eating animal, q.e.d.

Figure 4. Fully-detailed presentation of the proof of Schubert's Steamroller.

kind, in contrast to rules about categories ('a fox is an animal') and size relations ('birds are smaller than foxes').

The task of the presentation module is then to suitably mediate between such a concise proof sketch and a fully expanded proof description. One option, reducing the *quality*, leads to the text in Figure 5, achieves a compromise by fully explaining only the derivation of the first key assertion (foxes do not eat grain), while it merely states the other two key assertions derived. Since all three key assertions are derived by the same rule, this information can be stated compactly, preceding the derivation descriptions. The resulting description aims at reducing the set of propositions to be conveyed by explaining only a part of the proof in detail. This is done by selecting the propositions omitted in such a way that they are maximally connected, to minimize the number of potential clarification questions, which might address the derivations of one of the two key assertions, but not any more specific detail.

The other possibility is reducing the *convenience*, which leads to the text in Figure 6. It achieves a compromise by providing details about all key assertion derivations. The reduction here is obtained by merely stating the key assertions derived in connection with the underlying facts without elaborating how the responsible rule is applied. As in the previous case, that rule is only mentioned once, preceding the exposition of further details. A potential justification for

The proof runs through applying three times the rule that animals either eat plants or all plant eaters smaller than themselves. It is first derived that foxes do not eat grain. This ultimately follows from the assumptions that wolves do not eat grain and foxes are smaller than wolves, because animals who do not eat plants eat plant eaters smaller than themselves. Thus, either foxes do not eat grain or they are not smaller than wolves. Hence, only the first alternative is valid. Similarly, it is derived that birds eat grain, and finally, that foxes eat birds. Since foxes eat birds, an animal is known that eats a grain-eating animal, q. e. d.

Figure 5. Quality-reduced presentation of the proof of Schubert's Steamroller.

The proof runs through applying three times the rule that animals either eat plants or all plant eaters smaller than themselves. It is first derived that foxes do not eat grain. This ultimately follows from the assumptions that wolves do not eat grain and foxes are smaller than wolves. Thus, either foxes do not eat grain or they are not smaller than wolves. Hence, only the first alternative is valid. Moreover, it is derived that birds eat grain. Birds do not eat plant eaters because it is assumed that they do not eat snails, but snails are smaller than birds and they eat plants. Finally, it is derived that foxes eat birds, because they are plant eaters, and smaller than foxes. Since foxes eat birds, an animal is known that eats a grain-eating animal, q. e. d.

Figure 6. Convenience-reduced presentation of the proof of Schubert's Steamroller.

this presentation lies in augmenting the assumptions about the addressee's inferential capabilities – he/she is assumed to mentally apply a previously unknown recently mentioned rule to a number of facts.

The production of longer, but information-reduced, utterances can naturally serve the purpose of a summary meeting certain length parameters and content preferences. Moreover, these texts are well-suited as first-shot explanations in comparable discourse situations, based on known requirements or on tentatively made assumptions about the addressee. Further details may be exposed, guided by vague hints or by specific demands of the other conversant, who has at least the following options at his/her disposal:

- Assessments concerning choices made in building the condensed descriptions, such as 'be more concise' or 'be less concise', and 'emphasize why some intermediate conclusion holds', that is, elaborate on the underlying justifications, or 'emphasize how it is derived', that is, illustrate the underlying method.
- A general statement expressing that the description produced is considered insufficient in view of the explanatory demands.
- A request statement about extending specific portions of the description produced, such as 'Why does P(x) hold?' or 'How do Q(x1) to Q(xn) cause P(x)?'

The assessments are interpreted in terms of changes in the parameter setting that guide choices in producing descriptions. Moreover, they may be treated as indirect requests asking for modified or extended follow-up presentations, in case a presentation built according to these assessments provides additional information. Similarly, an unspecific request for clarification triggers the presentation of extensions to the previously given one, eventually by following an alternative strategy, in the spirit of reactive explanations, as demonstrated by Moore and Swartout [21]. Finally, when being confronted with a specific request, either a request for expanding the underlying method, or one for expanding the underlying justification, a dedicated description with the required information is produced. In all cases, communicating the additionally required material follows the same lines as the preceding first-shot dialog contribution. For complicated and argumentatively rich elaborations, similar constraints concerning the length of the utterance to be produced may prove to be relevant even for the chosen subset of information.

Being confronted with a shortened explanation, the addressee can request more details selectively, according to his/her interests (see Figure 7). For instance, details about the derivation of one or more

User: Why do birds eat grain?

System: Because animals eat all plants or all plant eaters smaller than themselves. Snails eat plants and are smaller than birds, and birds do not eat snails.

User: How is that derived?

System: For animals, it holds that (1) they eat all plants or (2) they eat all plant eaters smaller than themselves. (2) cannot hold for birds, because snails eat plants and they are smaller than birds, but it is known that birds do not eat snails. Hence (1) holds for birds.

User: Why do foxes and birds have their eating habits?

System: Foxes do not eat grain because wolves do not eat grain, so they eat plant eaters smaller than themselves. Since foxes are smaller than wolves, and it is assumed that wolves do not eat foxes, foxes cannot be plant eaters, because wolves eat plant eaters smaller than themselves. Birds eat grain because animals eat plants if they do not eat plant eaters smaller than themselves. Finally, foxes eat birds because animals either eat plants, which foxes don't do, since they do not eat grain, or they eat plant eaters smaller than themselves.

Figure 7. Follow-up questions to content-reduced presentation of the proof of Schubert's Steamroller.

of the unexplained facts in the explanation following with reduced quality can be focused on by asking dedicated how-questions. Similarly, why-questions can be asked to learn about the rules underlying specific derivations in the explanation with reduced convenience.

6 A Potential Extension

In this section, we demonstrate that the material for deductive argumentation, when provided by machine-generated proofs, restricts the associated argumentation in its scope. We illustrate the kind of limitations and describe additional sources for argumentation, exemplified by a new perspective on the Steamroller proof.

6.1 An Inherent Limitation

Arguments about a proof as considered so far merely consist of two components:

- *What* is derived, that is, the claims, which are intermediate steps in a proof, and may serve as arguments for other derivations.
- *Why* some results has been derived, that is, the proper arguments, which are the justifications of a proof step.

In essence, the entire proof is made up of a sequence of arguments of this kind. It may be varied so that it is more detailed or more condensed, more implicit or more explicit, but it merely specifies the facts that make up a proof. Such a presentation is inherently limited in its communicative function - it supports a "passive" understanding, which is restricted to a *control* or *verification* perspective on a proof. As opposed to that, an essential task in deduction is not merely *understanding*, but actually *finding* a proof. This puts a *search* or *performance* perspective on a proof, an "active" understanding for which there are no clues in the proper proof presentation.

6.2 The Performance Perspective

In order to provide an argumentative basis for showing how the search for a proof is carried out, high-level strategic conceptualizations are essential driving forces. These conceptualizations must consist in a rather limited repertoire of fundamental and adaptive techniques, which are relevant for different kind of proofs, but with varying details in concrete uses. Hence, assuming the principled acquaintance with such a conceptualization, recognizing its applicability in a concrete case, and a skillful performance in actually applying it must be addressed in an argumentative conversation. This characterization is typical for human-oriented problem-solving, with a mixture of limited, but highly diverse pieces of knowledge and operational skills to combine them. It is in sharp contrast to the large-scale uniform knowledge representation and schematic reasoning, which is the typical process organization for machine-oriented purposes. Therefore, even high-level characterizations of a machine-found proof, such as the level of proof plans [20] constitutes an inappropriate level of description for human-oriented purposes - the plans are too many and each of them contains too many details to be meaningful to humans as memorizable conceptualizations.

For elementary mathematics and logics, which are the most realistic areas for being subject to tutorial purposes, there are only a few fundamental proof techniques. Among them are the partitioning into simpler subproblems and the transformation to a different representation/calculus which allows for operations for which the original representation is inappropriate. The latter concept, for example, may be applicable in various contexts, including a transformation of assertions about residue classes into integer equations, and a transformation of operations on sets into propositional logic expressions. For humans, it is essential to recognize the commonality between the measures in each of these contexts. For addressing the domain of limit theorems, a method called "complex-estimate" has been developed as part of a proof planning for this domain [19]. This method is a specific form of the fundamental concept "partitioning into simpler subproblems", with a specific interpretation suitable for polynomial expressions. Since the method does not separate the (general) underlying concept from the (domain-specific) interpretation, which would render its application in automated proof planning considerably more difficult, it does not provide an adequate basis for argumentation about human problem-solving.

6.3 An Example – the Steamroller Proof

In our running example, the general concept underlying the problem-solving process is the *reduction of alternatives*. The relevance of this concept becomes apparent from the relation between the theorem to be proved and the formulation of the major piece of knowledge introduced in the problem definition. While the former states an eating relation between two animals, the latter specifies alternative possibilities for the eating habits of animals. Once the stategic value of the crucial problem-solving concept is recognized, the question arises how it can be applied in the given case. Since the alternatives are directly encapsulated in a rule, it is advisable to simply instantiate this rule so that it becomes evident which of the alternatives is true and which is false for a concrete instantiation. However, actually performing the instantiation may impose difficulties on a person unexperienced in problem-solving, since there are five animals in the context and each of them is a candidate for instantiating the two slots in the domain rule in question. In order to avoid exhaustive searching whenever possible, another general problem-solving principle can be

…

System: How do you propose to solve that problem?

User: I do not know.

System: In order to obtain more information, you should try to resolve undecided alternatives. Can you identify where they are?

User: The eating habits of the animals.

System: Exactly. Can you find out about some?

User: I should look at that longish rule, but I do not know where to start.

System: This refers to animals, so what animals?

User: Any pair may be relevant.

System: Look more closely at the rule about their eating habits.

User: One of them must be much smaller than the other.

System: Perfect. Can you name such a pair?

User: The fox and the bird.

…

Figure 8. Fragment of a tutorial dialog about finding a proof.

taken into account, namely ”look for most plausible instantiations first”. The key for establishing a preference among all candidates lies in the relation ”much smaller than” which must hold between the two animals referred to in the central axiom. Through this relation, more than half of the possible combinations are excluded. It is even more than that since the relation ”much smaller than” is interpreted in a non-transitive way in the Steamroller problem, otherwise there would be multiple solutions. Altogether, less than a handful of combinations remain, which can be tested in turn. Carrying out these tests amounts to applying ND inferences, which depends on the structure of instantiated subclauses. For rules of comparable complexity as the central axiom in the Steamroller problem, an unskilled person may require argumentative support as well. In particular, testing this rule with the instantiations ’fox’ and ’bird’ may turn out to be difficult – it is not known a priori whether or not the fox eats grain, hence, both eating options for the fox must be maintained at first. It is only the indirect clue from examining the eating habits of the wolf which clarifies the fact that the fox does not eat grain. Appropriate arguments about guiding the search in such a setting will certainly be profitable for a student.

Finally, we illustrate the functionality of this broader argumentation by a fragment of an interactive construction of the proof to the Steamroller problem, in a tutorial environment, exemplified by the hypothetical dialog in Figure 8.

The system statements in this conversation are, in fact, not proper argumentations, but some kind of hints. However, in order to produce these hints, relying on an argumentative basis as outlined above is absolutely necessary. Hence, putting these extensions to live is very likely to improve dialog capabilities in tutorial environments in an essential way.

7 Conclusion

The analysis of human proof explanations shows that certain logical inferences are only conveyed implicitly drawing on the discourse context and default expectations. Moreover, different syllogisms call for different presentation strategies to account for human performance. In this paper, we proposed the partial assertion level as an appropriate representation of proofs to plan the content of an explanation and different degrees of explicitness and condensation. Then, driven by the unfolding dialog, a reactive planner allows for an interactive, user-adaptive navigation through the proofs.

So far we have implemented *P.rex* and some tools for mediating between levels of abstraction. We are currently investigating manipulations of the proof structure to realize different degrees of explicitness. We will soon incorporate this work into a newly starting project on dialog-oriented tutoring systems. Moreover, we believe that our approach also proves useful for argumentative dialog systems in general.

REFERENCES

[1] Peter B. Andrews. Transforming matings into natural deduction proofs. In *Proceedings of the 5th International Conference on Automated Deduction*, pages 281–292. Springer Verlag, 1980.

[2] Daniel Chester. The translation of formal proofs into English. *AI*, 7:178–216, 1976.

[3] Yann Coscoy, Gilles Kahn, and Laurent Théry. Extracting text from proofs. In Mariangiola Dezani-Ciancaglini and Gordon Plotkin, editors, *Typed Lambda Calculi and Applications*, number 902 in Lecture Notes in Computer Science, pages 109–123. Springer Verlag, 1995.

[4] Bernd Ingo Dahn, J. Gehne, Thomas Honigmann, and A. Wolf. Integration of automated and interactive theorem proving in ILF. In William McCune, editor, *Proceedings of the 14th Conference on Automated Deduction*, number 1249 in LNAI, pages 57–60, Townsville, Australia, 1997. Springer Verlag.

[5] Andrew Edgar and Francis Jeffry Pelletier. Natural language explanation of natural deduction proofs. In *Proceedings of the 1st Conference of the Pacific Association for Computational Linguistics*, Vancouver, Canada, 1993. Centre for Systems Science, Simon Fraser University.

[6] Detlef Fehrer and Helmut Horacek. Exploiting the Addressee's inferential capabilities in presenting mathematical proofs. In Martha E. Pollack, editor, *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 959–964, Nagoya, Japan, 1997. Morgan Kaufmann.

[7] Armin Fiedler. Dialog-driven adaptation of explanations of proofs. In Bernhard Nebel, editor, *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1296–1300, Seattle, WA, 2001. Morgan Kaufmann.

[8] Gerhard Gentzen. Untersuchungen über das logische Schließen I & II. *Mathematische Zeitschrift*, 39:176–210, 572–595, 1935.

[9] H. Grice. Logic and conversation. *Syntax and Semantics*, 3:43–58, 1975.

[10] Amanda M. Holland-Minkley, Regina Barzilay, and Robert L. Constable. Verbalization of high-level formal proofs. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99) and Eleventh Innovative Application of Artificial Intelligence Conference (IAAI-99)*, pages 277–284. AAAI Press, 1999.

[11] Helmut Horacek. A model for adapting explanations to the user's likely inferences. *User Modeling and User Adapted Interaction*, 7:1–55, 1997.

[12] Helmut Horacek. Presenting proofs in a human-oriented way. In Harald Ganzinger, editor, *Proceedings of the 16th Conference on Automated Deduction*, number 1632 in LNAI, pages 142–156, Trento, Italy, 1999. Springer Verlag.

[13] Helmut Horacek. Tailoring inference-rich descriptions through making compromises between conflicting cooperation principles. *Int. J. Human-Computer Studies*,53:1117–1146, 2000.

[14] Helmut Horacek. Expressing references to rules in proof presentations. In R. Goré, A. Leitsch, T. Nipkov, *International Joint Conference on Automated Reasoning*, Short papers, Technical Report DII 11/01, University of Siena, pages 76–85, 2001.

[15] Xiaorong Huang. Reconstructing proofs at the assertion level. In Alan Bundy, editor, *Proceedings of the 12th Conference on Automated Deduction*, number 814 in LNAI, pages 738–752, Nancy, France, 1994. Springer Verlag.

[16] Xiaorong Huang and Armin Fiedler. Proof verbalization as an application of NLG. In Martha E. Pollack, editor, *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 965–970, Nagoya, Japan, 1997. Morgan Kaufmann.

[17] Philip Johnson-Laird and Ruth Byrne. *Deduction*. Ablex Publishing, 1990.

[18] Christoph Lingenfelder. *Transformation and Structuring of Computer Generated Proofs*. PhD thesis, Universität Kaiserslautern, Kaiserslautern, Germany, 1990.

[19] Erica Melis. AI-techniques in proof planning. In Henri Prade, editor, *Proceedings of of the 13th European Conference on Artifical Intelligence*, pages 494–498, Brighton, UK, 1998. John Wiley & Sons, Chichester, UK.

[20] Erica Melis and Jörg Siekmann. Knowledge-Based Proof Planning. *Artificial Intelligence*, 115(1):65–105, 1999.

[21] Johanna Moore and Bill Swartout. A reactive approach to explanation. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1504–1510, Detroit, MI, USA, 1989. Morgan Kaufmann.

[22] Mark E. Stickel. Schubert's steamroller problem: Formulations and solutions. *Journal of Automated Reasoning*, 2:89–101, 1986.

[23] Manfred Thüring and Kurt Wender. Über kausale Inferenzen beim Lesen. *Sprache und Kognition*, 2:76–85, 1985.

[24] Marylin Walker. The effect of resource limitations and task complexity on collaborative planning in dialogue. *Artificial Intelligence*, 85:181–243, 1996.

Argumentative Deliberation for Autonomous Agents

Antonis Kakas and Pavlos Moraïtis[1]

Abstract. This paper presents an argumentation based framework, developed as an extension of an existing framework for non-monotonic reasoning, in order to support an agent's self deliberation process. The framework allows the agent to draw conclusions taking into account in a natural way a given preference policy. After developing the argumentation framework we examine two general cases of such argumentative deliberation: (a) under a preference policy that takes into account the roles agents can have within a context pertaining to an environment of interaction and (b) under a preference policy for the current needs of the agent emerging from his profile. In the first case we apply the argumentative deliberation model within a simple agent interaction scenario where each agent's self-deliberation determines, according to his own policy, his position at each step of the interaction process. In the second case we apply the framework to model motivational factors that apparently drive human behaviors and therefore can define agent personalities. Agents will thus similarly, as it is claimed in psychological literature for human beings, chose at any moment to pursue, those goals that are most compatible with their current motivations.

The proposed argumentation framework allows us to define policy preferences at different levels of deliberation resulting in modular representations of the agent's knowledge or personality profile. This high degree of modularity gives a simple computational model in which the agent's deliberation can be naturally implemented.

1 Introduction

Argumentation has had a renewed interest in Artificial Intelligence with several recent works studying its links to various problems such as the formalization of law, non-monotonic and common-sense reasoning, agent deliberation and dialogue and others. Abstract frameworks of argumentation are very powerful as they can encode many different problems but they face the challenge of doing so in a direct and natural way that at the same time is amenable to a simple computational model.

In this paper, we study an argumentation framework developed over the last decade as a result of a series of studies [12, 8, 7, 11, 10, 6] on the links of argumentation to non-monotonic reasoning. This framework, called Logic Programming without Negation as Failure ($LPwNF$), was proposed originally in [10] and can be seen as a realization of the more abstract frameworks of [7, 4]. The abstract attacking relation, i.e. its notion of argument and counter-argument, is realized through monotonic proofs of contrary conclusions and a priority relation on the sentences of the theory that make up these proofs. We extend the framework, following the more recent approach of other works [23, 5] to allow this priority relation and thus the attacking relation to be dynamic, making the framework more suitable for applications.

We claim that this extended argumentation framework is a natural argumentation framework. But how should we define the naturality of an argumentation framework? To do so we can set the following desiderata for naturality:

- the framework must be *simple* employing a small number of basic notions e.g. a uniform single notion of attack between arguments
- the encoding of a problem within the framework must be *directly* related to the high-level specification of the problem
- the representations of problems must be *modular*, with changes in the problem accommodated locally within the argumentation theory
- the argumentative reasoning and its computation must be *modular and local* to the problem task or query at hand

These properties are motivated from the perspective of a viable computational model of argumentation. This list of desiterata is not meant to be a complete list but rather that these are good properties that one would expect from a natural argumentation framework. Ultimately, the best criterion of the naturality of a framework is the test whether it can be applied, exhibiting the above properties, to capture different forms of natural human argumentative reasoning thus formalizing natural behaviour.

For this reason after developing our argumentation framework we test this by studying in detail how it can be used to capture agent deliberation in a dynamic external environment. In particular, we examine two problems: (a) argumentative deliberation of an agent according to a given decision policy on a domain of interest that takes into account the roles filled by the agents and the context of the external environment, and (b) argumentative deliberation of an agent about his needs according to a meta theory of "personality" related preferences.

In this work, we adopt the idea that an agent is composed of a set of modules each of them being responsible for a particular functionality, and all together implementing the agent's overall behavior (e.g. problem solving, cooperation, communication, etc.). Therefore we consider that the proposed argumentative deliberation model can be used in order to implement the various decision making processes needed by different modules of an agent. For example, the decision for the choice and achievement of a goal (within the problem solving module) or the decision for the choice of the appropriate partners according to a specific cooperation protocol (within the cooperation module), etc.

Over the last few years argumentation is becoming increasingly important in agent theory. Several works have proposed argumentation models in the multi-agent field [28, 27, 21, 16, 3, 1, 2]. Our work can be seen as bringing together work from [27, 2] who have suggested that roles can affect an agent's argumentation, especially within the context of a dialogue, and work from [23, 5] who have shown the need for dynamic priorities within an argumentation framework when we want to apply this to formalize law and other related problems. In this paper, we put together these ideas proposing a new argumentation framework for agent deliberation obtained by extending the argumentation framework of ($LPwNF$) [10, 6] to include dynamic priorities. We also employ a simple form of abduction to deal with the incompleteness and evolving nature of the

[1] Dept. of Computer Science, University of Cyprus, P.O.Box 20537, CY-1678 Nicosia, Cyprus, antonis@ucy.ac.cy, moraitis@ucy.ac.cy

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

agent's knowledge of the external environment.

We show how our framework can encompass the influence that the different relative roles of interacting agents and the context of the particular interaction can have on the deliberation process of the agents. Roles and context define in a natural way dynamic priorities on the argumentative decision rules of the agent at two different levels in the deliberation process. These priorities are represented within the overall argumentation theory of the agent in two corresponding modular parts. The use of this argumentative deliberation framework is demonstrated within an interaction protocol where the agent's deliberation helps him to decide his position.

Our use of the same argumentation framework to model agent motivations and through that, agent personalities, is inspired by the classical work of Maslow [17] in which he sets up a theory of hierarchy of human needs (physiological, safety, affiliation, achievement, learning) corresponding to motivational factors that drive human behavior. According to this theory human beings consider their unsatisfied needs in an order and decide to satisfy first those that are lower (and hence more important) in the hierarchy before considering higher needs. In the agent literature, Maslow's theory is already used by [18, 19] for guiding the behavior of deliberative and reactive agents in various unpredictable environments. To our knowledge our work is the first time where argumentation is used to model Maslow's hierarchy and other similar agent personalities where the mechanism for choosing which need to address next is carried out via a process of argumentative deliberation.

Section 2 presents the extension of the basic argumentation framework of $LPwNF$ with dynamic priorities. It also gives the basic concepts of roles and context and how these are captured through dynamic priorities in argumentation. Section 3 studies a simple interaction protocol based on argumentative deliberation. Section 4 presents how we model within our argumentation framework a hierarchy of needs of an agent and how these are chosen via argumentative deliberation. Section 5 discusses related and future work.

2 Argumentative Deliberation

An agent has his own theory expressing the knowledge under which he will take decisions. This decision process needs to compare alternatives and arrive at a conclusion that reflects a certain policy of the agent. In this paper we formalize this type of agent reasoning via argumentation where the deliberation of an agent is captured through an argumentative evaluation of arguments and counter-arguments.

There are several frameworks of argumentation proposed recently (e.g. [22, 4]) that could be adopted for formalizing an agent's deliberation. We will use the framework presented in [10, 6], called *Logic Programming without Negation as Failure* $(LPwNF)$ (The historical reasons for this name are not directly relevant to this paper). We briefly review this framework and then study its extension needed to accommodate roles and context in argumentative deliberation.

In $LPwNF$ a non-monotonic argumentation theory is viewed as a pool of sentences (or rules) from which we must select a suitable subset, i.e. an argument, to reason with, e.g. to support a conclusion. Sentences in a $LPwNF$ theory are written in the usual extended logic programming language with an explicit negation, but without the Negation as Failure (NAF) operator. We will often refer to the sentences of a theory as argument rules. In addition, these rules may be assigned locally a "relative strength" through a partial ordering relation. For example, we may have

$fly(X) \leftarrow bird(X) \qquad \neg fly(X) \leftarrow penguin(X)$
$bird(X) \leftarrow penguin(X) \qquad bird(tweety)$

with an ordering relation between the rules that assigns the second rule higher than the first. This theory captures the usual example of "flying birds" with its exceptions, without the use of explicit qualifications of the default rules with abnormality conditions. We can conclude that $tweety$ flies since we can derive this from the first rule and there is no way to derive $\neg fly(tweety)$. We have an argument (i.e. a proof) for $fly(tweety)$ but no argument for $\neg fly(tweety)$. If we add to the theory $penguin(tweety)$ then we can derive both $fly(tweety)$ and $\neg fly(tweety)$ - we have an argument for either conclusion. But in the non-monotonic argumentation semantics of the theory we can only conclude $\neg fly(tweety)$. This overrides $fly(tweety)$ since the argument that derives $\neg fly(tweety)$ contains the second rule which is designated higher than the first rule which belongs to the argument that derives $fly(tweety)$. We say that the argument for $\neg fly(tweety)$ **attacks** the argument for $fly(tweety)$ but not vice-versa. In general, the argumentation-based framework of $LPwNF$ is defined as follows.

Definition 1 *Formulae in the* **background logic**[2] $(\mathcal{L}, \vdash)$ *of the framework are defined as* $L \leftarrow L_1, \ldots, L_n$*, where* $L, L_1, \ldots, L_n$ *are positive or explicit negative literals. The derivability relation,* $\vdash$*, of the logic is given by the single inference rule of modus ponens.*

Together with the set of sentences of a theory $\mathcal{T}$, we are given an ordering relation $<$ on these sentences (where $\phi < \psi$ or $<(\phi, \psi)$ means that ϕ has lower priority than ψ). The role of the priority relation is to encode locally the relative strength of argument rules in the theory. The relation $<$ is required to be irreflexive.

Definition 2 *An* **argumentation theory** $(\mathcal{T}, <)$ *is a set of sentences* $\mathcal{T}$ *in* $\mathcal{L}$ *together with a priority relation* $<$ *on the sentences of* $\mathcal{T}$*. An* **argument** *for a literal* L *in a theory* $(\mathcal{T}, <)$ *is any subset of* $\mathcal{T}$ *that derives* L*,* $T \vdash L$*, under the background logic.*

In general, we can separate out a part of the theory $\mathcal{T}_0 \subset \mathcal{T}$ (e.g. the last two rules of the example above) and consider this as a non-defeasible part from which any argument rule can draw information that it might need. The notion of attack between arguments in a theory $\mathcal{T}$ is based on the possible conflicts between a literal L and its explicit negation $\neg L$ and on the priority relation $<$ on $\mathcal{T}$.

Definition 3 *Let* $(\mathcal{T}, <)$ *be a theory and* $T, T' \subseteq \mathcal{T}$*. Then* T' **attacks** T *(or* T' *is a* **counter argument** *of* T*) iff there exists* L*,* $T_1 \subseteq T'$ *and* $T_2 \subseteq T$ *s.t.:*

(i) $T_1 \vdash_{min} L$ *and* $T_2 \vdash_{min} \neg L$
(ii) $(\exists r' \in T_1, r \in T_2 \text{ s.t. } r' < r) \Rightarrow (\exists r' \in T_1, r \in T_2 \text{ s.t. } r < r')$.

Here $T \vdash_{min} L$ means that $T \vdash L$ under the background logic and that L can not be derived from any proper subset of T. The second condition in this definition states that an argument T' for L attacks an argument T for the contrary conclusion only if the set of rules that it uses to prove L are at least of the same strength (under the priority relation $<$) as the set of rules in T used to prove the contrary. Note that the attacking relation is not symmetric.

Using this notion of attack we then define the central notions of an *admissible argument* of a given theory and the non-monotonic argumentation consequence relation of a given theory as follows.

Definition 4 *Let* $(\mathcal{T}, <)$ *be a theory and* T *a subset of* $\mathcal{T}$*. Then* T *is* **admissible** *iff* T *is consistent and for any* $T' \subseteq \mathcal{T}$ *if* T' *attacks* T *then* T *attacks* T'*.*

Definition 5 *Let* $T = (\mathcal{T}, <)$ *be a theory and* L *a ground literal. Then* L *is a* **credulous (resp. skeptical) consequence** *of* T *iff* L *holds in a (resp. every) maximal (wrt set inclusion) admissible subset of* $\mathcal{T}$*.*

[2] The background logic of this argumentation framework can be replaced with any monotonic first order logic.

2.1 Roles and Context

Agents are always integrated within a (social) environment of interaction. We call this the *context* of interaction. This determines relationships between the possible roles the different agents can have within the environment. We consider, in line with much of the agent literature, (e.g. [20, 30]), a *role* as a set of behaviour obligations, rights and privileges determining its interaction with other roles.

Generally, the substance of roles is associated to a *default context* that defines shared social relations of different forms (e.g. authority, friendship, relationship, etc.) and specifies the behaviour of roles between each others. Consequently, it implicitly installs a partial order between roles that can expresses preferences of behaviour. For instance in the army context an officer gives orders that are obeyed by a soldier, or in a everyday context we respond in favour more easily to a friend than to a stranger. However, a default context that determines the basic roles filled by the agents is not the only environment where they could interact. For example, two friends can also be colleagues or an officer and a soldier can be family friends in civil life. Therefore we consider a second level of context, called *specific context*, which can overturn the pre-imposed, by the default context, ordering between roles and establish a different social relation between them. For instance, the authority relationship between an officer and a soldier would change under the specific context of a social meeting at home or the specific context of treason by the officer.

2.2 Argumentation with Roles and Context

In order to accommodate in an agent's argumentative reasoning the roles and context as described above we can extend the framework of $LPwNF$ so that the priority relation of a theory is not simply a static relation but a dynamic relation that captures the non-static preferences associated to roles and context. There is a natural way to do this. Following the same philosophy of approach as in [23], the priority relation can be defined as part of the agent's theory $\mathcal{T}$ and then be given the same argumentation semantics along with the rest of the theory.

We distinguish the part of the theory that defines the priority relation by $\mathcal{P}$. Rules in $\mathcal{P}$ have the same form as any other rule, namely $L \leftarrow L_1, \ldots, L_n$ where the head L refers to the higher-priority relation, i.e. L has the general form $L = h\_p(rule1, rule2)$. Also for any ground atom $h\_p(rule1, rule2)$ its negation is denoted by $h\_p(rule2, rule1)$ and vice-versa. For simplicity of presentation we will assume that the conditions of any rule in the theory do not refer to the predicate h_p thus avoiding self-reference problems. We now need to extend the semantic definitions of attack and admissibility.

Definition 6 *Let* $(\mathcal{T}, \mathcal{P})$ *be a theory,* $T, T' \subseteq \mathcal{T}$ *and* $P, P' \subseteq \mathcal{P}$. *Then* (T', P') **attacks** (T, P) *iff there exists a literal* L, $T_1 \subseteq T'$, $T_2 \subseteq T$, $P_1 \subseteq P'$ *and* $P_2 \subseteq P$ *s.t.:*

(i) $T_1 \cup P_1 \vdash_{min} L$ *and* $T_2 \cup P_2 \vdash_{min} \neg L$
(ii) $(\exists r' \in T_1 \cup P_1, r \in T_2 \cup P_2 \text{ s.t. } T \cup P \vdash h\_p(r, r')) \Rightarrow (\exists r' \in T_1 \cup P_1, r \in T_2 \cup P_2 \text{ s.t. } T' \cup P' \vdash h\_p(r', r))$.

Here, when L does not refer to h_p, $T \cup P \vdash_{min} L$ means that $T \vdash_{min} L$. This extended definition means that a composite argument (T', P') is a counter-argument to another such argument when they derive a contrary conclusion, L, and $(T' \cup P')$ makes the rules of its counter proof at least "as strong" as the rules for the proof by the argument that is under attack. Note that now the attack can occur on a contrary conclusion L that refers to the priority between rules.

Definition 7 *Let* $(\mathcal{T}, \mathcal{P})$ *be a theory,* $T \subseteq \mathcal{T}$ *and* $P \subseteq \mathcal{P}$. *Then* (T, P) *is* **admissible** *iff* $(T \cup P)$ *is consistent and for any* (T', P') *if* (T', P') *attacks* (T, P) *then* (T, P) *attacks* (T', P').

Hence when we have dynamic priorities, for an object-level argument (from $\mathcal{T}$) to be admissible it needs to take along with it priority arguments (from $\mathcal{P}$) to make itself at least "as strong" as the opposing counter-arguments. This need for priority rules can repeat itself when the initially chosen ones can themselves be attacked by opposing priority rules and again we would need to make now the priority rules themselves at least as strong as their opposing ones.

Let us illustrate this extended form of argumentative reasoning with an example adapted from [23]. In this example, we are trying to formalise a piece of legislation that refers to whether or not we should modify an old building. In the first part,$\mathcal{T}$, of the theory we have the object-level law that refers directly to this particular topic:
$r_1(X) : \neg modify(X) \leftarrow protected(X)$
$r_2(X) : modify(X) \leftarrow needs\_repair(X)$

In addition, we have a theory $\mathcal{P}$ that represents the priorities between these (and other) laws as captured by another (more general) part of the law that deals with the relative strength of different types of regulations:
$rr_1(L_a, L_b) : h\_p(L_a(X), L_b(X)) \leftarrow art\_protect\_law(L_a(X)), planning\_law(L_b(X))$
$rr_2(L_a, L_b) : h\_p(L_a(X), L_b(X)) \leftarrow art\_protect\_law(L_b(X)), preservation\_law(L_a(X))$
$rr_3(rr_2, rr_1) : h\_p(rr_2(L_a(X), L_b(X)), rr_1(L_a(X), L_b(X))) \leftarrow dangerous(X)$.

The first of these states that a law for artistic protection is generally stronger than a planning law whereas the second says that a law for the preservation of an old building is generally stronger than an artistic protection law. The third statement stipulates that in the particular case of a building that is dangerous to the public then the law that gives higher priority to preservation laws over artistic protection laws is stronger than the law that gives higher strength to artistic protection laws over planning laws.

We also have in the non-defeasible part $\mathcal{T}_0$ of the theory some general information on the type of these laws together with information on a particular case for a $villa_0$:
$preservation\_law(r_2(X)) \leftarrow serious\_damage(X)$
$art\_protect\_law(r_1(X)) \qquad planning\_law(r_2(X))$
$protected(villa_0) \qquad needs\_repair(villa_0)$
$serious\_damage(villa_0) \qquad dangerous(villa_0)$.

Should we modify $villa_0$ or not and how do we argue the case for our conclusion? We have two conflicting object-level arguments relating to the modification of $villa_0$. These are $\Delta_1 = (\{r_1(villa_0)\}\{\})$ for $\neg modify(villa_0)$ and $\Delta_2 = (\{r_2(villa_0)\}\{\})$ for $modify(villa_0)$. We can strengthen these arguments by adding priority rules in them. If we extend Δ_1 to $\Delta_1^{'} = (\Delta_1, \{rr_1(r_1(villa_0), r_2(villa_0))\})$ then for Δ_2 to attack back $\Delta_1^{'}$ it needs to extend itself to $\Delta_2^{'} = (\Delta_2, \{rr_2(r_2(villa_0), r_1(villa_0))\})$. Now these extended arguments have another conflict on the priority between the object level rules r_1, r_2, i.e. on $h\_p(r_1(villa_0), r_2(villa_0))$. $\Delta_1^{'}$ and $\Delta_2^{'}$ attack each other on this. But $\Delta_2^{'}$ can strengthen its argument for $h\_p(r_2(villa_0), r_1(villa_0))$ by adding in its priority rules the rule $\{rr_3(rr_2, rr_1)\}$. In fact, if we consider the attack on $\Delta_1^{'}$ given by $(\{\}, \{rr_2(r_2(villa_0), r_1(villa_0)), rr_3(rr_2, rr_1)\})$ there is no way to extend Δ_1' so that it attacks this back. Hence Δ_1' (and Δ_1) is not admissible. We only have admissible sets that derive $modify(villa_0)$ and hence this is a skeptical conclusion.

This example illustrates in particular how we can take into account the relative strength that different types of law have on the reasoning. The types of law act as roles with relative importance which depends on the particular context under which we are examining the case.

We can now define an agent's argumentation theory for describing his policy in an environment with roles and context as follows.

Definition 8 *An agent's* **argumentative policy theory or theory,** T**,** *is a triple* $T = (\mathcal{T}, \mathcal{P}_R, \mathcal{P}_C)$ *where the rules in* $\mathcal{T}$ *do not refer to* h_p*, all the rules in* $\mathcal{P}_R$ *are priority rules with head* $h\_p(r_1, r_2)$ *s.t.* $r_1, r_2 \in \mathcal{T}$ *and all rules in* $\mathcal{P}_C$ *are priority rules with head* $h\_p(R_1, R_2)$ *s.t.* $R_1, R_2 \in \mathcal{P}_R \cup \mathcal{P}_C$*.*

We therefore have three levels in an agent's theory. In the first level we have the rules $\mathcal{T}$ that refer directly to the subject domain of the agent. We call these the **Object-level Decision Rules** of the agent. In the other two levels we have rules that relate to the policy under which the agent uses his object-level decision rules according to roles and context. We call the rules in $\mathcal{P}_R$ and $\mathcal{P}_C$, **Role (or Default Context) Priorities** and **(Specific) Context Priorities** respectively.

As an example, consider the following theory $\mathcal{T}$ representing (part of) the object-level decision rules of an employee in a company.
$r_1 : give(A, Obj, A_1) \leftarrow requests(A_1, Obj, A)$
$r_2 : \neg give(A, Obj, A_1) \leftarrow needs(A, Obj)$
$r_3 : \neg give(A, Obj, A_2) \leftarrow give(A, Obj, A_1), A_2 \neq A_1.$
In addition, we have a theory $\mathcal{P}_R$ representing the general default behaviour of the code of contact in the company relating to the roles of its employees: a request from a superior is in general stronger than an employee's own need; a request from another employee from a competitor department is in general weaker than its own need. (Here and below we will use capitals to name the priority rules but these are not to be read as variables).
$R_1 : h\_p(r_1(A, Obj, A_1), r_2(A, Obj, A_1)) \leftarrow higher\_rank(A_1, A)$
$R_2 : h\_p(r_2(A, Obj, A_1), r_1(A, Obj, A_1)) \leftarrow competitor(A, A_1)$
$R_3 : h\_p(r_1(A, Obj, A_1), r_1(A, Obj, A_2)) \leftarrow higher\_rank(A_1, A_2)$
Between the two alternatives to satisfy a request from a superior from a competing department or not, the first is stronger when these two departments are in the specific context of working together on a common project. On the other hand, if we are in a case where the employee who has an object and needs it, needs this urgently then s/he would prefer to keep it. Such policy is represented at the third level in $\mathcal{P}_C$:
$C_1 : h\_p(R_1(A, Obj, A_1), R_2(A, Obj, A_1)) \leftarrow common(A, Obj, A_1)$
$C_2 : h\_p(R_2(A, Obj, A_1), R_1(A, Obj, A_1)) \leftarrow urgent(A, Obj).$
Note the *modularity* of this representation. For example, if the company decides to change its policy "that employees should generally satisfy the requests of their superiors" to apply only to the direct manager of an employee we would simply replace R_1 by the new rule R_1' without altering any other part of the theory:
$R_1' : h\_p(r_1(A, Obj, A_1), r_2(A, Obj, A_1)) \leftarrow manager(A_1, A).$

Consider now a scenario where we have two agents ag_1 and ag_2 working in competing departments and that ag_2 requests an object from ag_1. This is represented by extra statements in the non-defeasible part, $\mathcal{T}_0$, of the theory, e.g. $competitor(ag_2, ag_1)$, $requests(ag_2, obj, ag_1)$. Should ag_1 give the object to ag_2 or not?

If ag_1 does not need the object then, there are only admissible arguments for giving the object, e.g. $\Delta_1 = (\{r_1(ag_1, obj, ag_2)\}, \{\})$ and supersets of this. This is because this does not have any counter-argument as there are no arguments for not giving the object since $needs(ag_1, obj)$ does not hold. Suppose now that $needs(ag_1, obj)$ does hold. In this case we do have an argument for not giving the object, namely $\Delta_2 = (\{r_2(ag_1, obj, ag_2)\}, \{\})$. This is of the same strength as Δ_1 but the argument Δ_2', formed by replacing in Δ_2 its empty set of rules of priority with $\{R_2(r_2(ag_1, obj, ag_2), r_1(ag_1, obj, ag_2))\}$, attacks Δ_1 and any of its supersets but not vice-versa: R_2 gives higher priority to the rules of Δ_2 and there is no set of priority rules with which we can extend Δ_1 to give its object-level rules equal priority as those of Δ_2. Hence we conclude skeptically that ag_1 will not give the object. This skeptical conclusion was based on the fact that the theory of ag_1 cannot prove that ag_2 is of higher rank than himself. If the agent learns that $higher\_rank(ag_2, ag_1)$ does hold then Δ_2' and Δ_1', obtained by adding to the priority rules of Δ_1 the set $\{R_1(r_1(ag_1, obj, ag_2), r_2(ag_1, obj, ag_2))\}$, attack each other. Each one of these is an admissible argument for not giving or giving the object respectively and so we can draw both conclusions credulously.

Suppose that we also know that the requested object is for a common project of ag_1 and ag_2. The argument Δ_2' is now not admissible since now it has another attack obtained by adding to the priority rule of Δ_1' the extra priority rule $C_1(R_1(ag_1, obj, ag_2), R_2(ag_1, obj, ag_2))$ thus strengthening its derivation of $h\_p(r_1, r_2)$. The attack now is on the contrary conclusion $h\_p(r_1, r_2)$. In other words, the argumentative deliberation of the agent has moved one level up to examine what priority would the different roles have, within the specific context of a common project. Δ_2' cannot attack back this attack and no extension of it exists that would strengthen its rules to do so. Hence there are no admissible arguments for not giving and ag_1 draws the skeptical conclusion to give the object.

We have seen in the above example that in several cases the admissibility of an argument depends on whether we have or not some background information about the specific case in which we are reasoning. For example, ag_1 may not have information on whether their two departments are in competition or not. This means that ag_1 cannot build an admissible argument for not giving the object as he cannot use the priority rule R_2 that it might like to do. But this information maybe just unknown and if ag_1 wants to find a way to refuse the request he can reason further to find *assumptions* related to the unknown information under which he can build an admissible argument. Hence in this example he would build an argument for not giving the object to ag_2 that is *conditional* on the fact that they belong to competing departments. Furthermore, this type of information may itself be dynamic and change while the rest of the theory of the agent remains fixed, e.g. ag_1 may have in his theory that ag_2 belongs to a competing department but he has not yet learned that ag_2 has changed department or that his department is no longer a competing one.

We can formalize this conditional form of argumentative reasoning by defining the notion of *supporting information* and extending argumentation with *abduction* on this missing information.

Definition 9 *Let* $T = (\mathcal{T}, \mathcal{P})$ *be a theory, and* $\mathcal{A}$ *a distinguished set of predicates in the language of the theory, called* **abducible** *predicates. Given a goal* G*, a set* S *of abducible literals consistent with the non-defeasible part* $\mathcal{T}_0$ *of* T*, is called a* **strong (resp. weak) supporting evidence** *for* G *iff* G *is a skeptical (resp. credulous) consequence of* $(\mathcal{T} \cup S, \mathcal{P})$*.*

The structure of an argument can also be generalized as follows.

Definition 10 *Let $T = (\mathcal{T}, \mathcal{P})$ be a theory and $\mathcal{A}$ its abducible predicates. A* **supported argument** *in T is a tuple (Δ, S), where S is a set of abducible literals consistent with $\mathcal{T}_0$ and Δ is a set of argument rules in T, which is not admissible in T, but is admissible in the theory $(\mathcal{T} \cup S, \mathcal{P})$. We say that S supports the argument Δ.*

The supporting information expressed through the abducibles predicates refers to the incomplete and evolving information of the external environment of interaction. Typically, this information pertains to the context of the environment, the roles between agents or any other aspect of the environment that is dynamic. We will see in section 3 how agents can acquire and/or validate such information through an interaction protocol where they exchange missing information.

Given the above framework the **argumentative deliberation** of an agent can be formalized via the following basic reasoning functions.

Definition 11 *Let Ag be an agent, T his argumentation theory, G a goal and S a set of supporting information consistent with $\mathcal{T}_0$. Then we say that Ag* **deliberates** *on G to produce s^{ag}, denoted by $deliberate(Ag, G, S; s^{ag})$, iff $s^{ag} \neq \{\}$ is a strong supporting evidence for G in the theory $T \cup S$. If $s^{ag} = \{\}$ then we say that Ag* accepts *G under $T \cup S$ and is denoted by* **accept(Ag,G,S)**. *Furthermore, given an opposing goal $\overline{G}$ (e.g $\neg G$) to G and s' produced by deliberation on $\overline{G}$, i.e. that $deliberate(Ag, \overline{G}, S; s')$ holds, we say that s' is supporting evidence for agent Ag to* **refuse** *G in $T \cup S$.*

2.3 Modularity and Computation

As mentioned above, the proposed framework allows modular representations of problems where a change in the policy of an agent can be effected locally in his theory. The following results formalize some of the properties of modularity of the framework.

Proposition 12 *Let Δ be a set of arguments that is admissible separately with respect to the theory $T_1 = (\mathcal{T}, \mathcal{P}_{R1}, \{\})$ and the theory $T_2 = (\mathcal{T}, \mathcal{P}_{R2}, \{\})$. Then Δ is admissible with respect to the theory $T = (\mathcal{T}, \mathcal{P}_{R1} \cup \mathcal{P}_{R2}, \{\})$. Similarly, we can decompose $\mathcal{P}_C$ into $\mathcal{P}_{C1}$ and $\mathcal{P}_{C2}$.*

Proposition 13 *Let Δ be a set of arguments that is admissible with respect to the theory $T_1 = (\mathcal{T}, \mathcal{P}_R, \{\})$. Suppose also that Δ is admissible with respect to $T_2 = (\mathcal{T} \cup \mathcal{P}_R, \{\}, \mathcal{P}_C)$. Then Δ is admissible with respect to $T = (\mathcal{T}, \mathcal{P}_R, \mathcal{P}_C)$.*

The later proposition shows that we can build an admissible argument $\Delta = (O, R)$ by joining together an object-level argument O together with a set of priority rules R that makes O admissible and is itself admissible with respect to the higher level of context priorities. These results provide the basis for a modular computational model in terms of interleaving levels of admissibility processes one for each level of arguments in the theory.

In general, the basic $LPwNF$ has a simple and well understood computational model [6] that can be seen as a realization of a more abstract computational model for argumentation [14]. It has been successfully used [13] to provide a computational basis for reasoning about actions and change. The simple argumentation semantics of $LPwNF$, where the attacking relation between arguments depends only on the priority of the rules of a theory, gives us a natural "dialectical" proof theory for the framework. In this we have two types of interleaving derivations one for considering the attacks and one for counter attacking these attacks. The proof theory then builds an admissible argument for a given goal by incrementally considering all its attacks and, whenever an attack is not counter-attacked by the argument that we have build so far, we extend this with other arguments (rules) so that it does so. This in turn may introduce new attacks against it and the process is repeated.

The priorities amongst the rules help us move from one type of derivation to the other type e.g. we need only consider attacks that come from rules with strictly higher priority than the rules in the argument that we are building (as otherwise the argument that we have so far will suffice to counter attack these attacks.) For the more general framework with dynamic priorities we apply the same proof theory extended so that a derivation can be split into levels. Now a potential attack can be avoided by ensuring that its rules are not of higher priority than the argument rules we are building and hence we move the computation one level up to attacks and counter attacks on the priorities of rules. This move one level can then be repeated to bring us to a third level of computation.

This extended proof theory has been implemented and used to build agents that deliberate in the face of complete (relevant) information of their environments. We are currently investigating how to extend this implementation further with (simple forms of ground) abduction, required for the computation of supporting evidence in the face of incomplete information about the environment, using standard methods from abductive logic programming.

3 Argumentation based Agent Interaction

In this section, we study the use of the argumentative deliberation of an agent, defined above, within a simple interaction protocol where two agents are trying to agree on some goal, as an example of how this argumentation framework can be used within the different decision making processes of an agent. In our study of this we will be mainly interested how agents can use their argumentative deliberation in order to decide their position at each step of the interaction process. We will not be concerned with the conversation protocol supporting the agent interaction.

Each agent builds his reaction according to his internal argumentative policy theory, his current goal and other supporting information about the external environment that he has accumulated from the other agent. This extra supporting information is build gradually during the interaction and it allows an incremental deliberation of the agents as they acquire more information.

In the specific interaction protocol that we will consider, each agent insists in proposing his own goal as long as his deliberation with his theory and the accumulated supporting information (agreed by the two agents so far) produces new supporting evidence for this goal, suitable to convince the other agent. The first of the two interacting agents, who is unable to produce a new such supporting evidence, abandons his own goal and searches for supporting information, if any, under which he can accept the goal of the other agent (e.g. a seller agent unable to find another way to support his high price considers selling at a cheap price, provided that the buyer has a regular account and pays cash). In such a case, if the receiver agent can endorse the proposed supporting information the interaction ends with an agreement on this goal and the supporting information accumulated so far. Otherwise, if the receiver refuses some of the proposed supporting information the sender takes this into account and tries again to find another way to support the goal of the other agent. If this is not possible then the interaction ends in failure to agree on a common goal.

The following algorithm describes the steps of the interaction process presented above. Let us denote by X and Y the two agents, by G^X, G^Y their respective goals, by S the knowledge accumulated during the interaction exchanges and by s_i^X, s_i^Y the various supports that the agents generate in their deliberation. Note that when G^X, G^Y are opposing goals any supporting evidence for one of these goals also forms a reason for refusing the other goal.

Besides the argumentative functions $deliberate$ and $accept$ given in the previous section, we need three more auxiliary functions, which are external to the argumentative reasoning of an agent and relate to other functions of the agent in the present interaction protocol. The function $propose(Goal, e_j, s_i)$ is used by a sender agent to determine what information to send to the other agent: $Goal$ is a goal proposed, e_j is the evaluation by the sender of the supporting information s_j sent to him in the previous step by the other agent, and s_i is a new supporting evidence produced by the deliberation function of the sender. The function $evaluate(Ag, s_i)$ produces e_i where each (abducible) literal in the supporting information s_i may remain as it is or negated according to some external process of evaluation of this by an agent Ag. The function $update(S, e)$ updates, through an external mechanism, the accumulated supporting information S with the new information e consisting of the agent's evaluation of the supporting evidence sent by the other agent and the evaluation of his own supporting information by the other agent.

As described above, the interaction protocol has two phases. Phase 1 where each agent insists on its own goal and Phase 2 where they are trying to agree on the goal of one of the two agents. In the definition below Phase 2 supposes that agent X initiates this phase.

Phase 1
Step 1 (Agent X starts the Interaction)
Agent X *propose*(G^X, ∅, s_n^X) to Y
Step 2 (Agent Y)
e_n^X←*evaluate*(Y, s_n^X); S←*update*(S, $e_{n-1}^Y \cup e_n^X$)
If Y *accept*(Y, G^X, S) **then** End(agreement, G^X)
Else n←n+1; agent Y *deliberate* (Y, G^Y, S; s_n^Y)
If s_n^Y exists **then** *propose*(G^Y, e_{n-1}^X, s_n^Y) to X
Else Start Phase 2
Step 3 (Agent Y)
e_n^Y←*evaluate*(X, s_n^Y); S←*update*(S, $e_{n-1}^X \cup e_n^Y$)
If X *accept*(X, G^Y, S) **then** End(agreement, G^Y)
Else n←n+1; agent X *deliberate* (X, G^X, S; s_n^X)
If s_n^X exists then *propose*(G^X, e_{n-1}^Y, s_n^X) to Y
Goto step2
Else Start Phase 2

Phase 2
Step 1 (Agent X)
S←*update*(S, e_n^X); n←n+1
agent X *deliberate*(X, G^Y, S; s_n^X)
If s_n^X exists **then** *propose*(G^Y, ∅, s_n^X) to Y
Else End(Failure)
Step 2 (Agent Y)
e_n^X←*evaluate*(Y, s_n^X)
If e_n^X=s_n^X **then** End(agreement, G^Y)
Else *propose* (G^Y, e^X, ∅); Goto step 1

We illustrate this algorithm with a buying-selling scenario between two agents, a seller called X who has the goal, G^X, to sell a product at a high price to another agent, the buyer, called Y, who has the (opposing) goal, G^Y, to buy this product at a low price. They are trying to find an agreement on the price by agreeing either on G^X or on G^Y. We assume that the seller has the following argumentation policy for selling products. We present only a part of this theory.

The object-level theory $\mathcal{T}^X$ of the seller contains the rules:
$r_1 : sell(Prd, A, high\_price) \leftarrow pay\_normal(A, Prd)$
$r_2 : sell(Prd, A, high\_price) \leftarrow pay\_install(A, Prd)$
$r_3 : sell(Prd, A, low\_price) \leftarrow pay\_cash(A, Prd)$
$r_4 : \neg sell(Prd, A, P_2) \leftarrow sell(Prd, A, P_1), P_2 \neq P_1.$
His role and context priority theories, $\mathcal{P}_R^X$ and $\mathcal{P}_C^X$, are given below. They contain the policy of the seller on how to treat the various types of customers. For example, to prefer to sell with normal paying conditions over payment by installments when the buyer is a normal customer (see R_1). Also that there is always a preference to sell at high price (see R_2, R_3) but for regular customers there are conditions under which the seller would sell at low price (see R_4, R_5). This low price offer to a regular customer applies only when we are not in high season (see C_1, C_2).

$R_1 : h\_p(r_1(Prd, A), r_2(Prd, A)) \leftarrow normal(A)$
$R_2 : h\_p(r_1(Prd, A), r_3(Prd, A))$
$R_3 : h\_p(r_2(Prd, A), r_3(Prd, A))$
$R_4 : h\_p(r_3(Prd, A), r_1(Prd, A)) \leftarrow regular(A), buy\_2(A, Prd)$
$R_5 : h\_p(r_3(Prd, A), r_1(Prd, A)) \leftarrow regular(A), late\_del(A, Prd)$
$C_1 : h\_p(R_2(Prd, A), R_4(Prd, A)) \leftarrow high\_season$
$C_2 : h\_p(R_2(Prd, A), R_5(Prd, A)) \leftarrow high\_season$
$C_3 : h\_p(R_4(Prd, A), R_5(Prd, A)).$

Lets us consider the particular interaction scenario given below and study how the seller uses his own argumentative deliberation in this scenario.

Phase1
Seller X (step 1): propose(G^X, ∅, s_0={pay normal})
Buyer Y (step 2): NO; e_0=s_0; S=(e_0); deliberate (Y, G^Y, S; s_1={expensive price}); propose (G^Y, e_0, s_1)
Seller X (step 3): NO; e_1=s_1, S=($e_0 \cup e_1$); deliberate (X, G^X, S; s_2={pay install}); propose (G^X, e_1, s_2)
Buyer Y (step 2): NO; e_2=¬s_2; S=($e_0 \cup e_1 \cup e_2$); deliberate(Y, G^Y, S; s_3={pay cash}); propose(G^Y, e_2, s_3)
Seller X (step 3): NO; e_3=s_3; S=($e_0 \cup e_1 \cup e_2 \cup e_3$), deliberate (X, G^X, S, s); fails

Phase 2
Seller X (step 1): S=($e_0 \cup e_1 \cup e_2 \cup e_3$); deliberate (X, G^Y, S; s_4={regular cust, buy 2}); propose (G^Y, ∅ s_4)
Buyer Y (step 2): NO; e_4={regular cust, ¬buy 2}; propose(G^Y, e_4, ∅)
Seller X (step 1): S=($e_0 \cup e_1 \cup e_2 \cup e_3 \cup e_4$); deliberate (X, G^Y, S; s_5={later delivery}); propose (G^Y, ∅, s_5)
Buyer Y (step 2): e_5=s_5;YES; End(agreement, G^Y)

At the third step of Phase1 the seller needs to see if he can find an argument to support his goal (of selling high) given the fact that the buyer considers the price expensive. Deliberating on his goal, he now finds another argument for selling high, using the object-level rule r_2 since he can no longer consider the buyer a normal customer and R_1 does not apply (the seller derives this from some general background knowledge that he has in $\mathcal{T}_0$ e.g. from a rule $\neg normal(A) \leftarrow expensive(A, high\_price)$). This new argument needs the support $pay\_install(buyer, prd)$ and the seller offers this information to the buyer.

At the last step of Phase1 the seller deliberates again on his own goal (to sell high) but cannot find a new solution anymore. He therefore initiates phase2 where he considers the goal of the buyer, i.e. to sell at low_price and finds that it is possible to do so if the customer is a regular one and he accepts some other conditions. He finds an admissible argument for low price using the object-level rule r_3 and the role priority rule R_4. This is conditional on the information that the buyer is indeed a regular customer, will pay cash and that he will buy two of the products. Note that for this argument to be admissible the context rule C_1 must not apply, i.e. the seller knows that currently they are not in a $high\_season$. The buyer confirms the first two conditions but refuses the third. The seller then has another solution to sell low to a regular customer conditional on late delivery.

It is easy to show the following result of termination and correctness of the above interaction protocol.

Proposition 14 *Let X, Y be two agents with T_X, T_Y their respective argumentation policy theories such that for each goal, G, there exists only a finite number of different supporting evidence for G in T_X or T_Y. Then any interaction process between X and Y will terminate. Furthermore, if an interaction process terminates with agreement on a goal G and S is the final set of supporting information accumulated during the interaction then G is a skeptical conclusion of both $T_X \cup S$ and $T_Y \cup S$.*

We also remark that the evaluation function, $evaluate(Ag, s_i)$, used by an agent within the interaction process in order to decide if he can accept a proposed supporting information s_i, can vary in complexity from a simple check in the agent's database on the one hand to a new (subsidiary) argumentative deliberation on s_i according to a related argumentative policy theory that the agent may have.

4 Agent Deliberation on Needs and Motivations

In this section, we will study how the argumentation framework proposed in this paper can help us model the needs and motivations of an agent. In particular, we will examine the argumentative deliberation that an agent has to carry out in order to decide which needs to address at any current situation that he finds himself. This will then allows us to use the argumentation framework to specify different personalities of agents in a modular way independently from the other architectural elements of an agent.

We will apply the same approach as when we model a preference policy of an agent in a certain knowledge or problem domain, described in the previous sections. We now simply consider the domain of an agent's needs and motivations where, according to the type or personality of an agent, the agent has a default (partial) preference amongst the different types of needs. Hence now the type of need, or the motivation that this need addresses, plays an analogous role to that of Roles in the previous section. The motivations will then determine the basic behaviour of the agent in choosing amongst his different needs and whenever we have some specific context this may overturn the default decision of the agent for a particular need.

We will follow the work of Maslow [17] from Cognitive Psychology (see also [18, 19]) where needs are categorized in five broad classes according to the motivation that they address. These are **Physiological, Safety, Affiliation or Social, Achievement or Ego and Self-actualization or Learning**. As the world changes a person is faced with a set of potential goals from which it selects to pursue those that are "most compatible with her/his (current) motivations". We choose to eat if we are hungry, we protect ourselves if we are in danger, we work hard to achieve a promotion etc. The theory states that in general there is an ordering amongst these five motivations that we follow in selecting the corresponding goals. But this ordering is only followed in general under the assumption of "other things being equal" and when special circumstances arise it does not apply.

Our task here is then to model and encode such motivating factors and their ordering in a natural way thus giving a computational model for agent behaviour and personality.

Let us assume that an agent has a theory $\mathcal{T}$ describing the knowledge of the agent. Through this, together with his perception inputs, he generates a set of needs that he could possibly address at any particular situation that he finds himself. We will consider that these needs are associated to goals, G, e.g. to fill with petrol, to rest, to help someone, to promote himself, to help the community etc. For simplicity of presentation and without loss of generality we will assume that the agent can only carry out one goal at a time and thus any two goals activated by $\mathcal{T}$ oppose each other and a decision is needed to choose one. Again for simplicity we will assume that any one goal G is linked only to one of the five motivations above, m_j, and we will thus write G_j, $j = 1, ..., 5$ to indicate this, with $m_1 = Physiological$, $m_2 = Safety$, $m_3 = Affiliation$, $m_4 = Achievement$, $m_5 = Self - actualization$.

Given this theory, $\mathcal{T}$, that generates potential goals an agent has a second level theory, $\mathcal{P_M}$, of priority rules on these goals according to their associated motivation. This theory helps the agent to choose amongst the potential goals that it has and forms part of his decision policy for this. It can be defined as follows.

Definition 15 *Let Ag be an agent with knowledge theory $\mathcal{T}$. For each motivation, m_j, we denote by S_j the set of conditions, evaluated in $\mathcal{T}$, under which the agent considers that his needs pertaining to motivation m_j are* **satisfied**. *Let us also denote by N_j the set of conditions, evaluated in $\mathcal{T}$, under which the agent considers that his needs pertaining to motivation m_j are* **critical**. *We assume that S_j and N_j are disjoint and hence N_j corresponds to a subset of situations where $\neg S_j$ holds. Then the* **default motivation preference theory of** *Ag, denoted by $\mathcal{P_M}$, is a set of rules of the following form:*

- $R^1_{ij} : h\_p(G_i, G_j) \leftarrow N_i$
- $R^2_{ij} : h\_p(G_i, G_j) \leftarrow \neg S_i, \neg N_j$

where G_i and G_j are any two potential goals, $(i \neq j)$, of the agent associated to motivations m_i and m_j respectively.

The first rule refers to situations where we have a critical need to satisfy a goal G_i whereas the second rule refers to situations where the need G_j is not critical and so G_i can be preferred.

Hence when the conditions S_i hold an agent would not pursue goals of needs pertaining to this motivation m_i. In fact, we can assume that whenever a goal G_i is activated and is under consideration that $\neg S_i$ holds. On the other side of the spectrum when N_i holds the agent has an urgency to satisfy his needs under m_i and his behaviour may change in order to do so. Situations where $\neg S_j$ and $\neg N_j$ both hold are in between cases where the decision of an agent to pursue a goal G_j will depend more strongly on the other simultaneous needs that he may have. These conditions S_i and N_i vary from agent to agent and their truth is evaluated by the agent using his knowledge theory.

For example, when a robotic agent has low_energy, that would make it non-functional, the condition N_1 is satisfied and a goal like $G_1 = fill\_up$ has, through the rules R^1_{1j} for $j \neq 1$, higher priority than any other goal. Similarly, when the energy level of the robotic agent is at some middle value, i.e. $\neg S_1$ and $\neg N_1$ hold, then the robot will again consider, through the rules R^2_{1j} for $j \neq 1$, the goal G_1 to fill up higher than other goals provided also that in such a situation there is no other goal whose need is critical. Hence if in addition the robotic agent is in great danger and hence N_2 holds then rule R^2_{12} does not apply and the robot will choose goal $G_2 = self\_protect$ which gets a higher priority through R^1_{21}.

In situations as in this example, the agent has a clear choice of which goal to select. Indeed, we can show that under some suitable conditions the agent can decide deterministically in any situation.

Proposition 16 *Let $\mathcal{P_M}$ be a preference theory for an agent and suppose that $N_i \cap N_j = \emptyset$ $(i \neq j)$ and that $\neg S_j = N_j$ for each j. Then given any two goals G_i, G_j only one of these goals belongs to an admissible extension of the agents theory and thus the agent at any situation has a deterministic choice of which need to address.*

Similarly, if we have $N_i \cap N_j = \emptyset$ and $\neg S_i \cap \neg S_j = \emptyset$ $(i \neq j)$ then the agent can always make a deterministic choice of which goal to choose to address in any current situation. But these conditions are too strong. There could arise situations where, according to the knowledge of the agent, two needs are not satisfied and/or where they are both urgent/critical. How will the agent decide which one to perform? The agent is in a *dilemma* as its theory will give an admissible argument for each need. For example, a robotic agent may at the same time be low in energy and in danger. Similarly, the robotic agent may be in danger but also need to carry out an urgent task of helping someone.

According to Maslow's theory decisions are then taken following a basic hierarchy amongst needs. For humans this basic hierarchy puts the Physiological needs above all other needs, Safety as the second most important with Affiliation, Achievement and Self-Actualization following in this order. Under this hierarchy a robotic agent would choose to fill its battery despite the danger or avoid a danger rather than give help. One way to model in $\mathcal{P_M}$ such a hierarchy of needs that helps resolve the dilemmas is as follows. For each pair $k, l\ s.t.\ k \neq l$ the theory $\mathcal{P_M}$ contains only one of the rules R^1_{kl} or R^1_{lk}. Deciding in this way which priority rules, R^1, to include in the theory gives a basic profile to the agent.

But this would only give us a partial solution to the problem not resolving dilemmas that are not related to urgent needs and a similar decision needs to be taken with respect to the second category of rules, R^2, in $\mathcal{P_M}$. More importantly this approach is too rigid in the sense that the chosen hierarchy in this way can never be overturned under any circumstance. Often we may want a higher degree of flexibility in modeling an agent and indeed Maslow's hierarchy of needs applies under the assumption of "other things being equal". In other words, there maybe special circumstances where the basic hierarchy in the profile of an agent should not be followed. For example, an agent may decide, despite his basic preference to avoid danger rather than help someone, to help when this is a close friend or a child.

We can solve these problems by extending the agent theory with a third level analogous to the specific context level presented in the previous sections.

Definition 17 *An agent theory expressing his decision policy on needs is a theory $T = (\mathcal{T}, \mathcal{P_M}, \mathcal{P_C})$ where $\mathcal{T}$ and $\mathcal{P_M}$ are defined as above and $\mathcal{P_C}$ contains the following types of rules. For each pair of rules R^k_{ij}, R^k_{ji} in $\mathcal{P_M}$ we have the following rules in $\mathcal{P_C}$:*

- $H^k_{ij} : h\_p(R^k_{ij}, R^k_{ji}) \leftarrow true$
- $E^k_{ji} : h\_p(R^k_{ji}, R^k_{ij}) \leftarrow sc^k_{ji}$
- $C^k_{ji} : h\_p(E^k_{ji}, H^k_{ij}) \leftarrow true$

where sc^k_{ji} are (special) conditions whose truth can be evaluated in $\mathcal{T}$. The rules H^k_{ij} are called the **basic hierarchy** *of the theory T and the rules E^k_{ji} the* **exception policy** *of the theory T. The theory $\mathcal{P_C}$ contains exactly one of the basic hierarchy rules H^k_{ij} and H^k_{ji} for each $k = 1, 2$ and $i \neq j$.*

Choosing which one of the basic hierarchy rules H^k_{ij} or H^k_{ji} to have determines the default preference of needs G_i over G_j or G_j over G_i respectively (for $k = 1$ in critical situations and for $k = 2$ in non-critical situations). The special conditions sc_{ij} define the specific contexts under which this preference is overturned. They are evaluated by the agent in his knowledge theory $\mathcal{T}$. They could have different cases of definition that depend on the particular nature of the goals and needs that we are considering in the dilemma.

Each choice of the rules H^k_{ij} to include in the agent theory, determining a basic hierarchy of needs, in effect gives a different agent with a different basic profile of behaviour. For example, if we have H^k_{34} in $\mathcal{P_C}$ (remember that $m_3 = Affiliation$ and $m_4 = Achievement$) we could say that this is an *altruistic* type of agent, since under normal circumstances (i.e. not exceptional cases defined by sc^k_{43}) he would give priority to the affiliation needs over the self-achievement needs. Whereas if we have H^k_{43} we could consider this as a *selfish* type of agent.

To illustrate this let us consider the specific theory $\mathcal{P_C}$ corresponding to Maslow's profile for humans. This will contain the following rules to capture the basic hierarchy of Physiological (m_1) over Safety (m_2) and Safety over Affiliation (m_3):

- $H^k_{12} : h\_p(R^k_{12}, R^k_{21}) \leftarrow true,\ for\ k = 1, 2$
- $H^k_{13} : h\_p(R^k_{13}, R^k_{31}) \leftarrow true,\ for\ k = 1, 2$
- $H^k_{23} : h\_p(R^k_{23}, R^k_{32}) \leftarrow true,\ for\ k = 1, 2$
- $E^2_{21} : h\_p(R^2_{21}, R^2_{12}) \leftarrow sc^2_{21}$
- $C^2_{21} : h\_p(E^2_{21}, H^2_{12}) \leftarrow true$
- $E^2_{31} : h\_p(R^2_{31}, R^2_{13}) \leftarrow sc^2_{31}$
- $C^2_{31} : h\_p(E^2_{31}, H^2_{13}) \leftarrow true$
- $E^2_{32} : h\_p(R^2_{32}, R^2_{23}) \leftarrow sc^2_{32}$
- $C^2_{32} : h\_p(E^2_{32}, H^2_{23}) \leftarrow true.$

The conditions sc^2_{21} are exceptional circumstances under which we prefer a safety need over a physiological need, e.g. sc^2_{21} could be true if an alternative supply of energy exists. Similarly for sc^2_{31} and sc^2_{32}. Note that if we are in a situation of critical physiological need (i.e N_1 holds and hence R^1_{12} applies) then this theory has no exceptional circumstances (there is no E^1_{21} rule) where we would not prefer to satisfy this physiological need over a critical safety need. Similarly, this profile theory does not allow any affiliation need to be preferred over a critical safety need; it does not allow a "heroic" behaviour of helping. If we want to be more flexible on this we would add the following rules in the profile:

- $E^1_{32} : h\_p(R^1_{32}, R^1_{23}) \leftarrow sc^1_{32}$
- $C^1_{32} : h\_p(E^1_{32}, H^1_{23}) \leftarrow true$

where the conditions sc^1_{32} determine the circumstances under which the agent prefers to help despite the risk of becoming non-functional, e.g. when the help is for a child or a close friend in great danger.

Given any such profile theory $\mathcal{P_C}$ we can show that an agent can always decide which goal to pursue once he can evaluate the sc^k_{ij} special conditions independently in $\mathcal{T}$ alone.

Proposition 18 *Let $T = (\mathcal{T}, \mathcal{P_M}, \mathcal{P_C})$ be an agent theory according to definition 17 and G_i, G_j $(i \neq j)$ be any two potential goals addressing different needs. Then given any situation there exists an admissible argument for only one of the two goals.*

In practice, the agent when in a dilemma will need to deliberate on each of the two goals and produce supporting information for each goal. This information is the incomplete information from $N_i, \neg S_i$ and sc^k_{ij} that the agent may be missing at the current situation. He would then be able to test (or evaluate) in the real world which one of these supporting information holds and thus enable him to make the decision which need to pursue.

Our argumentation based approach allows a high degree of flexibility in profiling deliberative agents. An agent's profile, defined via his $\mathcal{P}_{\mathcal{M}}$ and $\mathcal{P}_{\mathcal{C}}$ theories, is parametric on the particular rules we choose to adopt in both of these theories. In this paper we have adopted one possibility but this is certainly not the only one. For example, we could adopt a different underlying theory $\mathcal{P}_{\mathcal{M}}$ containing the basic priority rules amongst needs, rather than the fixed theory we have used in this paper, and use this as a new basis for profiling the agents. This issue needs to be studied further to examine the spectrum of different agents that can be build in this way.

5 Related Work and Conclusions

In this paper we have proposed an argumentative deliberation framework for autonomous agents and presented how this could be applied in different ways. We have argued that this framework has various desired properties of simplicity and modularity and in particular we have shown how it can capture some natural aspects of the behaviour of an autonomous agent. The framework can embody in a direct and modular way any preference policy of the agent and hence can be used to support the various decision making processes of an agent. It can be incorporated within different models of agent architecture. For example, it could be used within the BDI model to implement (with the necessary adaptations) the filter function [29] which represents the agent's deliberation process, for determining the agent's new intentions based on its current beliefs, desires and intentions. The proposed argumentation framework also has a simple and modular computational model that facilitates the implementation of deliberative agents.

The main characteristic of our argumentation framework is its modularity of representation and associated computation. Our work rests on the premise that for a computational framework of argumentation to be able to encapsulate *natural* forms of argumentation it is necessary for this framework to have a high degree of modularity. The argumentation theory of the agent should be able to capture locally and in a direct way the decision policy and accompanied knowledge of the agent. This modularity is needed for the agent to be able to carry out his argumentative deliberation efficiently, where at each particular instance of deliberation the computational argumentative process for this can be localized to the relevant (for this instance) part of the agent's argumentation theory. In a complex problem domain where an agent needs to address different types of problems and take into account different factors this ability to ”home in” on the relevant part of the theory is very important. Furthermore, the dynamic environment of an agent where new information is acquired and changes to his existing theory (or policy) can be made, requires that the representation framework is able to encode the agent's theory in a highly modular way so that these changes can be easily localized and accommodated effectively.

The argumentation framework developed and used in this paper is based on the more general and abstract notions that have emerged from a series of previous studies on argumentation [12, 8, 11, 7, 10]. The basic notion that is used is that of admissibility [7] which is itself a special case of acceptability [10]. It also follows the more recent approach of [23, 5] who have shown the need for dynamic priorities within argumentation when we want to apply this to formalize law and other related problems. Our framework is close to that of [23] in that it uses a similar background language of logic programming. They also both have a computational model that follows a dialectical pattern in terms of interleaving processes one for each level of arguments in the theory. In comparison our framework is simpler using only a single notion of attack and avoids the separate use of negation as failure that is subsumed by the use of rule priorities. In [5] dynamic priorities are related to the argumentation protocols, also called rules of order, describing which speech acts are legal in a particular state of the argumentation. Although the interests for application of our framework are different its formal relation to these frameworks is an interesting problem for further study.

In the development of agent deliberation we have introduced, in the same spirit as [27, 2], roles and context as a means to define non-static priorities between arguments of an agent. This helps to capture the social dimension of agents, as it incorporates in a natural way the influence of the environment of interaction (which includes other agents) on the agents ”way of thinking and acting”. We have shown how we can encompass, within this framework, the relative roles of agents and how these can vary dynamically depending on the external environment. The representation of this role and context information is expressed directly in terms of priority rules which themselves form arguments and are reasoned about in the same way as the object level arguments. This gives a high-level encapsulation of these notions where changes are easily accommodated in a modular way.

The use of roles and dynamic context is a basic difference with most of other works [28, 27, 21, 16, 3, 1] on agent argumentation. Our work complements and extends the approaches of [27, 2] with emphasis on enriching the self argumentative deliberation of an agent. It complements these works by linking directly the preferences between different contexts, which these works propose, to a first level of roles that agents can have in a social context, called default context, showing how roles can be used to define in a natural way priorities between arguments of the agents filling these roles. It extends this previous work by incorporating reasoning on these preferences within the process of argumentative deliberation of an agent. This is done by introducing another dimension of context, called specific context, corresponding to a second level of deliberation for the agent. This allows a higher degree of flexibility in the adaptation of the agents argumentative reasoning to a dynamically changing environment. In [2] the context preferences can also be dynamic but the account of this change is envisaged to occur outside the argumentative deliberation of the agent. An agent decides a-priori to change the context in which he is going to deliberate. In our case the change is integrated within the deliberation process of the agent.

This extra level of deliberation allows us to capture the fact that recognized roles in a context have their impact and substance only within this default context where they are defined, although these roles always ”follow” agents filling them, as a second identity in any other context they find themselves. Therefore agents who have some relationships imposed by their respective roles can be found in a specific context where the predefined (according to their relationships) order of importance between them has changed.

In comparison with other works on agent argumentation our work also integrates abduction with argumentation to handle situations where the information about the environment, currently available to the agent, is incomplete. This use of abduction is only of a simple form and more work is needed to study more advanced uses of abduction drawing from recent work on abduction in agents [26]. Another direction of future work concerns dialogue modeling. Our aim is to use our argumentative deliberation model for determining dialogue acts and protocols thus extending the framework of [15].

We have also studied, following the work of Maslow's hierarchy of needs [17], the use of our argumentative deliberation framework to model an agent's needs corresponding to motivational factors. This allows the expression of different personality profiles of an agent in a modular and flexible way. In the agent literature [18, 19] have already used Maslow's theory for guiding the behaviour of deliberative and reactive agents in various unpredictable environments. However, to our knowledge, this is first time that an argumentative deliberation framework is used to model these motivation factors, in a way that, we believe, allows a more natural expression of several behaviours. Also in comparison with the various behavior-based approaches for agent personalities (e.g. [25, 24]), our work gives an alternative model for specifying different personalities in a modular way independently from the other architectural elements of the agent. In addition, our approach uses a uniform representation framework for encoding an agent's personality and other policies or protocols associated with some of his different functionalities, e.g. with his problem solving capability.

More work is needed in this direction. On the technical side we need to extend the framework to allow an agent to decide amongst goals which address more than one need simultaneously. Also a deeper study is needed to explore the flexibility of the framework in modeling different agent personalities with respect to the way that they address their needs. Here we can draw further from work in cognitive science (see e.g. [9]) on the characteristics of human personalities. It is also important to study how these different personalities play a role in the interaction among agents especially in relation to the problem of forming heterogeneous communities of different types of agents, where the deliberation process of an agent may need to take into account the personality profile of the other agents.

In our work so far we have considered as separate the different processes of (i) generating an agent's needs and associated goals and (ii) deciding which one of these is prevalent under the current circumstances. The potential goals that an agent generates at any situation can be influenced by the personality of the agent and his previous decisions of which goal and need to address. According to Maslow when a more important need is satisfied then new goals for other less important needs are generated. We are currently studying how to integrate together these processes into a unified model for the overall deliberation of an argumentative agent, where these two processes are interleaved into each other, taking also into account the deliberative decision making of the agent on how to satisfy his chosen goals.

REFERENCES

[1] L. Amgoud, N. Maudet, and S. Parsons, 'Modelling dialogues using argumentation', in *ICMAS-00, pp. 31-38*, (2000).

[2] L. Amgoud and S. Parsons, 'Agent dialogues with conflicting preferences', in *ATAL01*, (2001).

[3] L. Amgoud, S. Parsons, and N. Maudet, 'Arguments, dialogue and negotiation', in *ECAI-00, pp. 338-342*, (2000).

[4] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni, 'An abstract, argumentation-theoretic framework for default reasoni ng', *Artificial Inelligence*, **93(1-2)**, 63–101, (1997).

[5] G. Brewka, 'Dynamic argument systems: a formal model of argumentation process based on situation calculus', in *Journal of Logic and Computation, 11(2), pp. 257-282*, (2001).

[6] Y. Dimopoulos and A. C. Kakas, 'Logic programming without negation as failure', in *Proc. ILPS'95, pp. 369-384*, (1995).

[7] P.M. Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', in *Artificial Intelligence, 77, pp. 321-357 (also in IJCAI'93)*, (1995).

[8] P.M. Dung, A. C. Kakas, and P. Mancarella, 'Negation as failure revisited', in *University of Pisa Technical Report*, (1992).

[9] Great Ideas in Personality, 'Five-factor model', in *www.personalityresearch.org/bigfive.html*, (2002).

[10] A. C. Kakas, P. Mancarella, and P.M. Dung, 'The acceptability semantics for logic programs', in *Proc. ICLP'94, pp. 504-519*, (1994).

[11] A.C. Kakas, 'Default reasoning via negation as failure', in *LNAI, Vol. 810, pp. 160-179*, (1994).

[12] A.C. Kakas, R.A. Kowalski, and F. Toni, 'Abductive logic programming.', in *Journal of Logic and Computation, 2(6), pp. 719-770*, (1992).

[13] A.C. Kakas, R.S. Miller, and F. Toni, 'E-res: Reasoning about actions, events and observations', in *LPNMR'01, pp. 254-266*, (2001).

[14] A.C. Kakas and F. Toni, 'Computing argumentation in logic programming', in *JLC 9(4), 515–562, O.U.P*, (1999).

[15] N. Karacapilidis and P. Moraitis, 'Engineering issues in inter-agent dialogues', in *Proc. of 15th European Conference on Artificial Intelligence (ECAI02), Lyon, France,*, (2002).

[16] S. Kraus, K. Sycara, and A. Evenchik, 'Reaching agreements through argumentation: a logical model and impl ementation', in *Artificial Intelliegence, 104 pp. 1-69*, (1998).

[17] A. Maslow, 'Motivation and personality', in *Harper and Row, New York*, (1954).

[18] P. Morignot and B. Hayes-Roth, 'Adaptable motivational profiles for autonomous agents', in *Knowledge Systems Laboratory, Report No. KSL 95-01, Dept of Computer Science, Stanford University*, (1995).

[19] P. Morignot and B. Hayes-Roth, 'Motivated agents', in *Knowledge Systems Laboratory, Report No. KSL 96-22, Dept of Computer Science, Stanford University*, (1996).

[20] P. Panzarasa, N.R. Jennings, and T. Norman, 'Formalising collaborative decision-making and practical reasoning i n multi-agent systems', in *Journal of Logic and Computation 12 (1), to appear*, (2002).

[21] S. Parsons, C. Sierra, and N.R. Jennings, 'Agents that reason and negotiate by arguying', in *Logic and Computation 8 (3), 261-292*, (1998).

[22] J.L. Pollock, 'Justification and defeat', in *Artficial Intelligence Vol 67, pp. 377-407*, (1994).

[23] H. Prakken and G. Sartor, 'A dialectical model of assessing conflicting arguments in legal rea soning', in *Artficial Intelligence and Law Vol 4, pp. 331-368*, (1996).

[24] P. Rizzo, M. Veloso, M. Miceli, and A. Cesta, 'Goal-based personalities and social behaviors in believable agents', *Applied Artificial Inelligence*, **13**, 239–272, (1999).

[25] D. Rousseau and B. Hayes-Roth, 'Improvisational synthetic actors with flexible personalities', in *Technical Report, KSL 97-10, Stanford University*, (1997).

[26] F. Sadri, F. Toni, and P. Torroni, 'Dialogues for negotiation: agent varieties and dialogue sequences', in *ATAL01*, (2001).

[27] C. Sierra, N.R. Jennings, P. Noriega, and S. Parsons, 'A framework for argumentation-based negotiation', in *ATAL-97, pp. 167-182*, (1997).

[28] K. Sycara, 'Argumentation: Planning other agents' plans', in *IJCAI-89, pp. 517-523*, (1989).

[29] M. Wooldridge, *Introduction to Multi-Agent Systems*, John Wiley and Sons, 2002.

[30] M. Wooldridge, N.R. Jennings, and D. Kinny, 'The gaia methodology for agent-oriented analysis and design', in *JAAMAS 3 (3), pp. 285-312*, (2000).

Session IV - Discussion

Natural is Uncertain, Emotional, Deceptive and Still Other. But: How to Get it?

Position Statement and Questions

Fiorella de Rosis[1]

Linguists seem to agree since long in claiming that a 'good' argumentation system should be keen in selecting the arguments that are 'strong' in given circumstances, by envisaging counter-arguments and producing counter-counter arguments (in advance if needed, or on request), in order to eventually produce a well formatted, coherent and 'convincing' message. Argumentation theories go back to the origins of our culture and AI researchers should apparently only find out appropriate techniques to produce such natural results. However, artificial argumentation systems are still far from being natural: indeed, obstacles still found in the production of a satisfying solution are due, in my view, to the unclearness of some concepts in these theories. I'll try to list some of these problems, in the hope that they may contribute to the Workshop discussion.

1 Strength of arguments and theories to treat them

Is an argument strong in itself or does its strength vary according to the Hearer to whom it is addressed and to the context in which the interaction occurs? How should strength be measured? Is there only one measure of strength ('probative weight' or 'plausibility', or 'impact') or should several variables be combined to produce an overall measure of argument strength? If so, which numerical parameters should be associated with the various elements that constitute 'an argumentation scheme' and with the data, to enable calculating its strength when applied to these data?

I am inclined for avoiding to 'invent an *ad hoc* theory' to measure and combine the argument strength: probability and utility theories provide a comfortable environment in which to place such a problem. Belief networks and inference diagrams enable us to represent chaining of arguments and propagation of uncertainty along this chain, from possibly uncertain evidence. They allow, as well, to define how to measure different concepts that contribute to establishing an argument 'strength'; for instance: 'warrant's qualifier' , 'uncertainty in the belief about data', 'impact of data on the 'claim', 'plausibility of data and claim to the Hearer', 'complexity of an argument' (to the Speaker and to the Hearer), 'cost of failing in convincing the Hearer', and so on. Finally, they provide a vivid representation of the strength of those arguments in which 'information sources' are cited (such as in Walton's "Argument from position to know" or "Appeal to Expert opinion" [7]), by enabling a definition of 'positive and negative competence' (the equivalents of 'sensitivity and specificity', in epidemiology), 'positive and negative sincerity', 'informativeness' of the source and how these measures affect the plausibility of the communicated data.

[1] Department of Informatics, University of Bari derosis@di.uniba.it http://aos2.uniba.it:8080/IntInt.html

Workshop on *Computational Models of Natural Argument*
Edited by Giuseppe Carenini, Floriana Grasso and Chris Reed
ECAI 2002. *15th European Conference on Artificial Intelligence*

2 Building argumentation chains from argumentation schemes

Belief networks (BNs) [3] are not a novel formalism, in argumentation. They have been applied, for instance, by Zukerman [8] to build a prototype system that produces, at the same time, arguments and answers to rebuttals. But do arguments produced by these systems show the characteristics that would enable us to label them as 'natural'? Not yet, I believe. Although 'insincere' argumentation may be simulated (and this is, in my view, a clear sign of naturalness [1]), the natural language texts produced by these systems are not much 'natural'. In addition, counter and counter-counter argumentation (rebuttals or responses to them) is still weak. An explanation of this limit is that BNs (as they have been employed so far) do not represent the rich linguistic, psychological and rhetorical knowledge that is embedded in argumentation schemes. They are often not much more than chains of logical rules to which uncertainty, measured in probability terms, is associated. To make BNs more knowledgeable, some semantics should be associated with their nodes and arcs. A rule that results from applying Walton's "Appeal to Expert Opinion" schema might be formulated, for instance, as follows:

(Say X f) and (Expert X f) $\Rightarrow^?$ (T f), with
(Competent X f) and (Sincere X f) $\Rightarrow^?$ (Expert X f)
(NegCompetent X f) and (F f) $\Rightarrow^?$ (Bel X f)
(PosCompetent X f) and (T f) $\Rightarrow^?$ (Bel X f)
(NegSincere X f) and (Say X (f) $\Rightarrow^?$ (Bel X f)
(PosSincere X f) and (Say X f) $\Rightarrow^?$ (Bel X f)

where X is an Agent, f is a fact and the symbol $\Rightarrow^?$ should be read as a 'probabilistic implication' and represented in terms of conditional probability tables. This would enable representing, in the BN, the knowledge that is needed to answer, after the argument: "*The fact f may plausibly be taken to be true because X asserts that f is true*", critical questions such as those mentioned by Walton: "*But how competent and sincere is X as a source? Is X's assertion based on evidence?*" (questions that are aimed, in this case, at checking the truth value of major or minor premises) or "*X is not an expert in the subject domain to which f belongs!*" in which the truth value of a premise that was not mentioned explicitly in the argumentation text is evoked. Another advantage of this formalism is that it enables relaxing the difference between 'observable' and 'not observable' data: as evidence about any node in the network may be propagated, argumentation may be chained in any direction: back from data to other data (the typical means-end reasoning) or forward from claims to other claims (a 'hypothetical reasoning about the implications' of a claim). For instance: when I come to know that (Say X f) and (T f), I may update my belief on X's expertise about f.

However, in translating argumentation schemes into BNs, several problems arise. First of all: how may rebuttals should be represented in these networks? This sends us back to a more basic question: Are Toulmin's rebuttals the same as Walton's critical questions, or are they something different? In the previous example, several objections might be made to the argument: "*The fact f may plausibly be taken to be true because X asserts that f is true*". Some of these objections are the critical questions mentioned by Walton, in which an objection is made about some (direct or indirect) premise of the scheme. But objections of a different kind might be raised, by evoking other argumentation schemes (in the previous example, "Appeal to popular Opinion"); or by applying the same argumentation scheme to different data, that produce contrasting results: for instance, "*But Y asserts that f is false, and he is an expert too*".

Is this the kind of objection that we call a 'rebuttal'?.

If the answer is 'yes', no problem, apparently: we just add to our BN some more arcs towards the same claim-node, and that's all! Old fashioned 'Expert Systems' would have enabled us to do this by combining uncertainty in the two schemes according to 'parallel' and 'sequential' propagation rules. But no one would propose such an *ad hoc* theory anymore: and, with belief networks, uncertainty due to application of different schemes cannot be calculated incrementally, as if the two knowledge sources were independent of each other. So, to be able to reply to rebuttals, all of the possible rebuttals have to be represented in the BN (which increases considerably the network's complexity!).

3 Intertwining 'pathos' with 'logos'

'Rational' argumentation apparently dominates the domain of psycholinguistics, as many place the kind of argumentation in which emotional factors are evoked, among the examples of 'deceptive' or 'unfair' argumentation. As a matter of fact, though, appeal to emotion and to a scale of values ('pathos' or 'ethos') are frequently applied, in human-human communication, to persuade somebody to perform some action. In Sillince's list of warrants [6], for instance, those based on ethical or social rules, or on appeal to goals, are the majority. So, emotions cannot be neglected when reproducing human-like argumentation systems is a goal.

Again, however, emotions are triggered and abandoned according to a mechanism in which uncertainty, weight given to goals, and time decay play a crucial role. Again, then, the emotional impact of a specific argument, for a given Hearer, and in a given context should be modeled through a formalism in which such factors are considered and treated appropriately (again, for instance, belief networks: [2]).

But how should emotional arguments be combined with logical ones? Should they, like someone assumes, be a 'last resort' to which to recur only in case of failure of other strategies? Or isn't it more 'natural' to wisely intermingle rational with emotional steps, in an argumentation message? For instance, emotional arguments might be evoked in a shallow and a bit elusive way, while more 'rational' ones might be spelled out more clearly and in detail. This argumentation style might be achieved by applying different methods to translate knowledge in the BNs into natural language messages, in the two cases. Is there any evidence of how this occurs, in human argumentation? Any corpus of data of public domain?

Silly questions? Too general ones? I hope not.

REFERENCES

[1] V Carofiglio, C Castelfranchi, F de Rosis and R Grassano, 'An interactive system for generating arguments in deceptive communication'. Springer LNAI 2175, 2001.

[2] V Carofiglio and F de Rosis: 'Mixed emotion modeling'. *AISB'02*, London, April 2002.

[3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA, Morgan Kaufmann (1988).

[4] C. Reed: 'Is it a monologue, a dialogue or a turn in a dialogue?'

[5] J A Sillince and R H Minors: 'What makes a strong argument? Emotions, highly-placed values and role-playing'. *Communication and Cognition*, 1991

[6] J A Sillince: 'Warrant selection for an intelligent argumentation tool', *Applied Artificial Intelligence*, **12**, 1998

[7] D N Walton: 'Argumentation and theory of evidence'. In *New Trends in Criminal Investigation and Evidence*, 2000.

[8] I Zukerman, R Mc Conachy, K Korb and D Pickett: 'Exploratory interaction with a bayesian argumentation system'. *Proceedings of IJCAI* 1999.

Contents