CMNA 2012

the 12th workshop on

Computational Models of Natural Argument

27 August, Montpellier, France

a workshop of ECAI 2012

CMNA XII
The 12th Workshop on
Computational Models of Natural Argument

Welcome to the 12th edition of the CMNA workshop!!

The workshop series in Computational Models of Natural Argument has been an important annual showcase of work in natural argumentation since its inception in 2000. This edition is no exception in demonstrating the breath and depth of work in the area.

The longest standing event on Argument and Computation, CMNA has always been characterised by its aiming at the broader interdisciplinary audience, interested in natural, that is real argumentation. Naturalness may involve the use of means which are more immediate than language to illustrate a point, such as graphics or multimedia. Naturalness can also relate to the preference for one particular style of reasoning as opposed to another to structure complex arguments. Or to the use of more sophisticated rhetorical devices, interacting at various layers of abstraction. Or the exploitation of extra-rational characteristics of the audience, taking into account emotions and affective factors.

The current edition is no different in presenting a broad showcase of uses and models of natural argumentation, from the linguistic perspective to the interface issues, from the more formal approaches to the rhetorical and persuasive aspects.

CMNA will not be possible without the unvalued contribution of our programme committee. They have worked very hard in providing top quality, de- tailed reviews: on busy times like these, we are profoundly grateful for their tremendous support.

We hope you will enjoy CMNA XII variegated programme, and we look for- ward, as always, to creative and stimulating discussions.

Floriana Grasso
Nancy Green
Chris Reed

Program Committee

Leila Amgoud	IRIT - CNRS, France
Trevor Bench-Capon	University of Liverpool, UK
Timothy Bickmore	Northeastern University, UK
Tom Gordon	Fraunhofer FOKUS, Germany
Marco Guerini	FBK-Irst, Italy
Helmut Horacek	Saarland University, Germany
Anthony Hunter	University College London, UK
Helena Lindgren	Umeå University, Sweden
David Moore	Leeds Metropolitan University, UK
Fabio Paglieri	ISTC-CNR, Italy
Vincenzo Pallotta	University of Fribourg, Switzerland
Cecile Paris	CSIRO ICT Centre, Australia
Patrick Saint-Dizier	IRIT-CNRS, France
Paul Piwek	The Open University, UK
Henry Prakken	University of Utrecht & University of Groningen, The Netherlands
Sara Rubinelli	University of Lugano, Switzerland
Doug Walton	University of Windsor, Canada
Adam Wyner	University of Liverpool, UK
Tommy Yuan	University of York, UK

Table of Contents

The Language of Learner Proof Discourse: A Corpus Study on the Variety of Linguistic Forms

Magdalena Wolska[1]

Abstract. The paper presents an analysis of linguistic diversity in learner language used in argumentative tutorial dialogues on mathematical proofs conducted in German. The analysis is based on two corpora of dialogues with a tutoring system simulated in a Wizard of Oz setup. The purpose of the analysis is to inform and motivate the choice of computational input processing methodology for an intelligent tutoring system for proofs. After lexical normalisation of mathematical domain-specific vocabulary, learner utterances are classified with respect to, first, linguistic "modality" (natural language vs. symbolic notation) and second, their dialogue function. Proof-contributing utterances are further classified with respect to their function in the proof under construction (proof steps, declarations of proof strategy to be adopted, etc.) and the type of content verbalised in natural language (logical connectives only, domain-specific vocabulary, etc.) Linguistic diversity is quantified in terms of type-token ratios over the normalised linguistic patterns, frequency spectra, and pattern-vocabulary growth curves. The analysis shows that even this seemingly linguistically predictable argumentative domain of mathematical proofs is characterised by a large variety of linguistic patterns of expression along all the above dimensions and by a large number of idiosyncratic verbalisations. Interesting is, moreover, a conversational character of the non-proof-contributing utterances, suggesting learners' informal attitude towards the computer-based dialogues and high expectations on the input interpretation resources. This calls for a combination of shallow and deep semantic processing methods for the discourse in question: shallow pattern-based approaches for contributions which do not add to the proof and deep lexicalised grammars for the proof-relevant content, in order to optimise coverage.

1 MOTIVATION

Mathematical proof can be considered *the* argumentative discourse par excellence: premises must be stated, claims justified, hypotheses discharged, only valid rules of inference followed. Narrative flair is of secondary importance; rigorous argumentation in mathematical proofs is characterised by a highly stylised language which combines formal symbolic expressions and worded natural language structures.

While proofs are central to mathematics, learners often lack skill in constructing proofs or even lack understanding of the need for proof in the first place [12, 1, 28]. Since proofs cannot be learnt other than by practice, the idea of building automated proof tutoring systems is appealing. Indeed, a number of mathematical assistance systems have been adapted for teaching proofs [30, 7, 17]. These systems, however, rely on controlled template-like input of proof structuring language and a formal language for mathematical expressions. EXCHECK [24] was a notable example of a system in which learners could use some natural language, however, its successor, EPGY TPE [25] uses menu-based input and a formula editor. This tendency toward controlled formal input as an interaction mode goes against findings on cognitive difficulties experienced by students while learning to do formal mathematics which show that the formal language and notation are among the major obstacles in proofs [26].

Support for open-ended natural language in a proof tutoring system requires that the language understanding component be capable of translating the learners' input into a symbolic representation required by a deduction system responsible for reasoning. With the view to provisioning such input processing capabilities we collected corpora of learner proofs constructed in a flexible natural language interaction (in German) with an anticipated dialogue-based tutoring system, simulated by a human. In this paper we present an analysis of linguistic diversity of the language the learners used in the course of the interaction. The purpose of the analysis is to inform and motivate the choice of computational input processing methodology for an intelligent tutoring system for proofs.

Outline The paper is organised as follows: In Section 2 the proof corpora are briefly presented. Section 3 describes data preparation: encapsulation of mathematical symbolic content, turn, utterance and word tokenisation, and textual normalisations. Section 4 presents a classification of utterance types. Section 5 presents the analyses: Linguistic diversity has been quantified in terms of (i) type-token ratios over normalised linguistic patterns along different dimensions, (ii) frequency spectra, and (iii) pattern-vocabulary growth curves. The results are discussed in Section 6.

2 PROOF TUTORING CORPORA

Our analysis of proof tutoring is based on two corpora of tutorial dialogues on mathematical theorem proving collected in Wizard of Oz experiments [19]. The domain of mathematics in the first corpus, C-I, was naive set theory and in the second corpus, C-II, binary relations.

In both experiments dialogues were conducted in German using the keyboard and a graphical user interface. The subjects were instructed to enter proof steps, rather than complete proofs at once, to encourage interaction with the system. The set theory corpus contains dialogues conducted in three experimental tutoring conditions: minimal feedback, didactic, or socratic tutoring strategy. Tutor's verbosity of the minimal feedback condition was limited, while in both other conditions as well as in the second experiment, the subjects and the tutors were unconstrained in terms of the linguistic realisation of their turns. The binary relations corpus contains dialogues conducted in two experimental study-material conditions: subjects

[1] Computational Linguistics, Saarland University, Saarbrücken, Germany, email: magda@coli.uni-saarland.de

Table 1. Basic descriptive statistics on the two corpora.

	Set theory (C-I)	Binary relations (C-II)
Proof tasks	3	4
Tutors	1	4
Subjects/Sessions	22	37
Turns	775	1906
Mean No. of turns per session (SD)	35 (12)	51 (19)
Subjects' turns (% of No. turns)	332 (43%)	927 (49%)
Mean No. of subjects' turns per session (SD)	15 (6)	25 (10)
Mode No. of attempted proofs per subject	3	2

received background reading on binary relations presented in either a verbose or a formal variant. In both experiments, the simulated systems followed strict turn-taking rules on the subject's end of the interaction: the interface did not allow the subjects to contribute a new turn until the wizard completed their turn.

The graphical user interface of the simulated system enabled button- and/or keyboard-based insertion of symbolic mathematical expressions. Unlike in the experiments described in [13] no structured editor for the symbolic expressions nor a dedicated area for mathematical formula entry was provided; mathematical expressions could be smoothly interleaved with natural language, as in mathematical narrative discourse in textbooks or scientific publications. The interactions were logged in plain ASCII format. Mathematical symbols were logged as their corresponding unicode numeric tokens (in C-I) or as their LaTeX commands (in C-II).

To illustrate the type of data under analysis here, in Figure 1 we give excerpts from both corpora which are illustrative of the type of language used.[2] C-I comprises 775 turns (332 student and 443 tutor turns, respectively), C-II has 1906 turns (927 student and 979 tutor turns). Table 1 summarises basic descriptive information on the experiments and the collected corpora. More details on the proof tutoring corpora and the experiments can be found in [35, 5].

3 PRE-PROCESSING

3.1 Pre-processing mathematical expressions

In both corpora, mathematical expressions were identified semi-automatically, using a regular-expression grammar. The grammar comprised a vocabulary of letters, mathematical symbols (unicode or LaTeX), brackets, braces, delimiters, etc. The parser's output was manually verified and corrected where necessary.[3] The quantitative analyses were conducted based on turns and utterances in which the identified mathematical expressions have been substituted with a symbolic token MATHEXPR.[4]

[2] Here and in further examples, German utterances have been translated into English preserving sense and grammatical structure as close as possible.

[3] We do not report precision results on mathematical expression identification and parsing as this is not the focus of this paper. It is assumed that an end-to-end system provides an entry method for mathematical expressions which would enable clear, possibly real-time, identification of mathematical expressions. This could be accomplished by explicitly defining "math mode" delimiters, for instance, as key combinations indicating the start and end of mathematical expression strings or as textual delimiters analogous to the \$-symbols in LaTeX.

[4] As shown in [33] utterances normalised this way can be parsed using a lexicalised grammar if the information on the expression's type – term or formula – is known. With this in mind, we therefore also classify the symbolic expressions into one of the following categories: i) atomic terms: VAR, for set, relation, or individual variables, ii) non-atomic terms: TERM or _TERM_ (object-forming operation symbols appearing in isolation (as in

C-I

S1: Wenn $A \subseteq K(B)$, dann $A \cap B = \emptyset$
(en. *If $A \subseteq K(B)$, then $A \cap B = \emptyset$*)
...
S5: in $K(B)$ sind alle x, die nicht in B sind
(en. *in $K(B)$ are all x which are not in B*)
S6: Da $A \subseteq K(B)$ gilt, alle x, die in A sind sind auch nicht in B
(en. *Since $A \subseteq K(B)$ holds, all x which are in A are also* not in B)
...
S8: Dann gilt auch: Alle x, die in B sind, sind nicht in A
(en. *Then it also holds: All x which are in B are not in A*)

C-II

S1: Ich moechte zunaechst $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$ beweisen
(en. *First I would like to prove $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$*)
S2: Sei $(a, b) \in (R \circ S)^{-1}$
(en. *Let $(a, b) \in (R \circ S)^{-1}$*)
...
S6: Nach der Definition von $\circ$ folgt dann (a, b) ist in $S^{-1} \circ R^{-1}$
(en. *By definition of $\circ$ it follows then that (a, b) is in $S^{-1} \circ R^{-1}$*)
...
S8: Der Beweis geht genauso wie oben , da in Schritt 2 bis 6 nur Aequivalenz umformungen stattfinden
(en. *The proof goes exactly as above since in step 2 to 6 there are only equivalences*)
S9: wie kann ich jetzt weitermachen?
(en. *how can I continue now?*)
...
S11: 1. Fall: Sei $(a, b) \in R$
(en. *1. Case: Let $(a, b) \in R$*)
S12: Ich habe mich vertippt. Korrektur: Sei $(a, z) \in R$
(en. *I made a typo. Correction: Let $(a, z) \in R$*)
...
S17: Ich habe gezeigt: $(a, b) \in (R \cup S) \circ T \Rightarrow (a, b) \in R \circ T \vee (a, b) \in S \circ T$
(en. *I have shown: $(a, b) \in (R \cup S) \circ T \Rightarrow (a, b) \in R \circ T \vee (a, b) \in S \circ T$*)
...
S24: Dann existiert ein z, so dass $(a, z) \in (R \cup S)$ und $(z, b) \in T$
(en. *Then there exists an z such that $(a, z) \in (R \cup S)$ and $(z, b) \in T$*)
S25: Nach Aufgabe A gilt $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$
(en. *By Exercise A $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$ holds*
...
S29: Da die Mengenvereinigung kommutativ ist, koennen wir dieses in student 25 einsetzen und erhalten die Behauptung
(en. *Since set union is commutative, we can use what's in student 25 and obtain the theorem*)
S30: nach Aufgabe W und dem Beispiel-Beweis gilt ...
(en. *By Exercise W and the example proof it holds ...*
...

Figure 1. Examples of learner utterances from both corpora

Examples of utterances from Figure 1 before and after mathematical expression pre-processing are shown below:

(1) Da $A \subseteq K(B)$ gilt, alle x, die in A sind sind auch nicht in B [C-I S6]

Da MATHEXPR$_{\text{FORMULA}}$ gilt, alle MATHEXPR$_{\text{VAR}}$, die in MATHEXPR$_{\text{VAR}}$ sind sind auch nicht in MATHEXPR$_{\text{VAR}}$

(2) Nach der Definition von $\circ$ folgt dann (a, b) ist in $S^{-1} \circ R^{-1}$ [C-II S6]

Nach der Definition von MATHEXPR$_{\text{_TERM_}}$ folgt dann MATHEXPR$_{\text{TERM}}$ ist in MATHEXPR$_{\text{TERM}}$

3.2 Turn and utterance pre-processing

Turns in both corpora were sentence-tokenised based on a standard set of end-of-sentence punctuation marks. The output of the sentence tokeniser was manually verified and corrected where necessary. Word-tokenisation was performed using a standard tokeniser.

Turns were then segmented into utterances. While a sentence is typically defined as a unit of speech containing a subject and a predicate, there is no precise linguistic definition as to what constitutes an utterance. Broadly understood, an utterance is an intentional, meaningful communicative act in an interaction. An utterance may consists of a word, a phrase, or a complex sentence with embedded clauses. It may form a complete turn, but a turn may also consist of more than one utterance. For the purpose of this study the notion of an utterance was operationalised as follows:

- An utterance never spans more than one turn or one sentence;
- Multiple clauses conjoined with conjunctions ("und" (en. *and*), "oder" (en. *or*), "aber" (en. *but*), "weil" (en. *because*), "für" (en. *for*), "also" (en. *so*), "wenn" (en. *if*), "als"/"wann" (en. *when*), etc.) were considered one utterance;
- Multiple clauses conjoined without conjunction words were considered separate utterances;
- "If-then" constructions, also those omitting the words "if" and "then", were considered a single utterance;
- The following non-sentential fragments, not containing a subject, were considered utterances: noun phrases, discourse markers (also inserts, such as "acha", "oh", "naja", "schoen" (en. *nice*)), colloquial subject-drop phrasings in indicative and interrogative mood, single question words and ellipted questions (for instance, "Fertig?" (en. *Done?*)), politeness phrases (such as "sorry", "Danke"), exclamatives (for instance, "Weitere Hilfe!" (en. *Further help!*)), non-sentential answers to questions, including acknowledgments ("ok", "klar" (en. *that's clear*)), yes/no answers.

Examples of tokenised multi-utterance turns from Figure 1 are shown below:[5]

(3) $\langle u \rangle$|Dann|gilt|auch|:|Alle|x|,|die|in|B|sind|,|sind|nicht|in|A|$\langle /u \rangle$ [C-I S8]

(4) $\langle u \rangle$|1.|Fall|:$\langle /u \rangle$
$\langle u \rangle$Sei|MATHEXPR|$\langle /u \rangle$ [C-II S11]

(5) $\langle u \rangle$|Ich|habe|mich|vertippt|.|$\langle /u \rangle$
$\langle u \rangle$|Korrektur|:|$\langle /u \rangle$
$\langle u \rangle$|Sei|MATHEXPR|$\langle /u \rangle$ [C-II S12]

the example utterance (2)), etc. and iii) formulas, FORMULA, for truth-valued statements, _FORMULA_ (statement-forming operators appearing in isolation), etc.

[5] | marks token boundaries. $\langle u \rangle$ and $\langle /u \rangle$ mark utterance boundaries.

3.3 Textual normalisations

Following extensive research into the properties of spoken and written discourse [10, 6], recent studies of computer-mediated communication (CMC) – or electronic discourse more generally – have shown that, much like spoken language differs from written language, the language of type-written computer-mediated communication shares some properties with spoken language, however, it also possesses textual and linguistic characteristics which are not typical for standard written language [23, 11, 18, 3]. Among those non-standard characteristics are the frequent use of abbreviations and acronyms, words and phrases written in all capitals or all lower-case, extensive use of certain punctuation marks and lack or incorrect (random) use of other punctuation (for instance, excessive use of the exclamation mark, lack of or incorrect use of commas, lack of valid end-of-sentence punctuation), and the use of emoticons. Also type-written tutorial dialogue shows qualities which are found both in spoken and written language and those of CMC. It is prone to textual ill-formedness due to the informal setting and the telegraphic nature of the linguistic production.

In order to avoid the effects of CMC-specific qualities of the learners' productions at the utterance-level, prior to the quantitative analysis learners' utterances were normalised with respect to certain writing mechanics phenomena (alternative spelling variants, capitalisation, punctuation) and with respect to the wording of common abbreviations. Moreover, lexical normalisations were performed on lexemes and phrases in order to avoid spurious diversity due to domain-specific terminology and context-specific references. Different lexical realisations of single and multi-word domain terms and conventional speech acts were substituted with symbolic tokens representing their lexical, in case of the former, or communicative, in case of the latter, types. Discourse-specific references were likewise normalised. Details of textual normalisations are summarised below.

Spelling The German umlaut diacritics were replaced with their underlying vowels and an "-e". The *eszett* ligatures were replaced with double "s". Spelling mistakes were identified and corrected using the German aspell, a Linux spell-checker, whose general dictionary has been extended with a custom dictionary of relevant domain terms (e.g. "Distributivität"/"Distributivitaet" (en. *Distributivity*));

Punctuation Repeated consecutive occurrences of the same punctuation symbols were replaced with a single occurrence ("!!!" → "!"; "...." → ".", etc.) Punctuation in abbreviations, missing or incorrect, has been normalised (e.g. "b..zw" → "bzw.", "d.h" → "d.h."). In the final analysis inter-sentential and end of sentence/utterance punctuation was ignored;

Abbreviations Upon correcting punctuation different correct and incorrect lexical variants of common abbreviations were substituted with symbolic tokens. These included, BSP for different spelling and capitalisation variants of "z.B." (en. *e.g.*), BZW for "bzw." (en. *respectively*), OBDA for "o.B.d.A." (en. *without loss of generality*), DH for "d.h." (en. *that is*), QED for "q.e.d.", ST for "s.t." (en. *such that*), OK for "ok", "oki", "Okay", etc.

Common speech acts and inserts Conventional expressions of gratitude, such as "Danke", "VIELEN DANK" and apologies, for instance, "Tut mir leid", "Sorry", "Verzeihung", were substituted with tokens THANKYOU and APOLOGY respectively. “Ja”/“Nein” responses were substituted with the token YESNO. Conversational inserts and other discourse markers such as “So”, “Na ja” were substituted with the token DISCOURSEMARKER.

Domain terms and domain-specific references Different lexical variants of nominal and adjectival domain terms which were included in the preparatory material have been mapped to a single form, DOMAINTERM. If single-word domain terms were part of a multi-word term which can be considered a named entity, the multi-word term was normalised. For instance, “DE-MORGAN-1”, “DeMorgan-1”, “DeMorgan-Regel-1”, “de morgan regel 2” all mapped to DOMAINTERM, as did “Distributivitaet von Vereinigung ueber den Durchschnitt” as a multi-word term (a name of a statement/theorem), as well as “symmetrisch” as a single-word term.

Non-deictic references to proof exercises, such as “Aufgabe W” (en. *Exercise W*), theorems provided in the preparatory material, such as “Theorem 9” or “9”, parts of proof structure, such as “Schritt 1” (en. *Step 1*), or turns in the dialogue history, such as “Student 25”[6], were mapped to the token REFERENCE.

Different conventional wordings used to signal the end of a proof, such as “quod erat demonstrandum”, “was zu zeigen war” (en. *which was to be shown*), “woraus der beweis folgt” (en. *from which the proof follows*), “Damit ist der Beweis fertig” (en. *which completes the proof*), etc., were mapped to the token corresponding to the “q.e.d.” abbreviation, QED.

Capitalisation The analyses presented in Section 5 were performed on corpus utterances normalised as above with case-insensitive matching.

Examples of utterances from Figure 1 pre-processed as outlined in this section are shown below:

(6) dann existiert ein MATHEXPR so dass MATHEXPR und MATHEXPR [C-II S24]

(7) nach REFERENCE gilt MATHEXPR [C-II S25]

(8) da DOMAINTERM DOMAINTERM ist koennen wir dieses in REFERENCE einsetzen und erhalten die Behauptung [C-II S29]

(9) nach REFERENCE und REFERENCE gilt MATHEXPR [C-II S30]

4 CLASSIFICATION OF UTTERANCE TYPES

Learner contributions in a tutoring interaction may fulfill several functions. As illustrated in the dialogue excerpts in Figure 1, learners contribute not only proof steps – complete or incomplete (C-I S5: a justification of the statement is not given), explicit or implicit (as in C-II S8: a high-level description of a set of steps is given rather than explicit proof steps) – but also other content which adds to the solution indirectly (as in C-II S1: a solution strategy to be adopted is described or C-II S11: a proof structure to follow – case distinction – is signalled) or which does not add to the solution at all (C-II S9: help is requested).

[6] References of this form are artifacts of our dialogue display interface. In the dialogue history, student turns were numbered and labelled “Student 1”, “Student 2”, etc. while tutor turns were labelled “Tutor 1”, etc.

In order to investigate linguistic diversity of learner proof discourse at a level corresponding to the different functional contribution types, we designed a typology of learner utterances based on the corpus data at hand. The present classification builds on previously proposed dialogue move taxonomies for tutorial dialogue [22, 32, 9, 4] and has been adapted specifically for the proof tutoring domain based on the analysis of our data. The classification, shown in Table 2, has a shallow hierarchical structure focusing on *Solution-contributing* content. All utterances which do not contribute solution proposals are grouped into one category, *Other*, with an extra class. *Uninterpretable* for utterances whose semantics or pragmatic intent could not be interpreted; for instance, because they were cut off mid-utterance.

The distinction between the *Solution-contributing* class and *Other* is that with *solutions* a learner is adding information to the solution he is constructing, be it by contributing an explicit or implicit solution step or steps, changing the meta-level status of the solution (for instance, stating that a new attempt at a solution will be made) or by signalling a revision or an evaluation of an already contributed solution part. The *Other* class may comprise utterances which express learner’s knowledge, but only those explicitly elicited by the tutor and which do not add to the solution being constructed. Since in the scope of this paper we are mainly interested in the analysis of argumentative language of mathematical proofs and so focus on contributions with solution-relevant content, the classification of utterances which do not contribute solution steps is coarse-grained.[7]

Note that the present classification can be mapped to previously proposed classifications of dialogue actions in tutoring. For instance, the category *Proof contribution* corresponds to *Contribute domain content* in the classification proposed in [32], to *Information Exchange : Assert* in [4] and *Assertions* in [22], and comprises the categories *Solution-step* and *Solution-strategy* from [8]. Following the general scheme proposed in [9] our class of *Proof contributions* further coded in the *Novelty* dimension for steps which contribute new content (C-II S17 is a counter-example) and in the *Motivation* dimension as *Internal* or *External*, depending on whether they have been elicited by the tutor. Utterances in the *Motivation : External* category would be found, among others, in our *Answer* category.

5 QUANTITATIVE ANALYSIS OF THE LINGUISTIC FORMS

We begin the quantitative analysis with a high-level overview of the amount of natural language verbalisation in the learner language by looking at the distribution of turns and utterances formulated using mathematical symbols alone, using natural language alone, and using natural language interleaved with mathematical symbols. Following this overview, we focus on the latter two categories; that is on utterances formulated using *some* natural language. We first look at the distribution of utterance types, as defined in Section 4, in the two corpora. Then we take a closer look at the *Proof contribution* utterances, in particular at the *Proof step* category in terms of the type of content that is verbalised. We summarise the most frequently encountered linguistic forms – linguistic *verbalisation patterns* – by category, and analyse the growth of the diversity of forms with the increasing corpus size. In all analyses we consider the two corpora in isolation (C-I and C-II) and also a larger corpus consisting of the two corpora combined into one data set (C-I $\cup$ C-II).

[7] We provide the full utterance classification, including the non-solution-related categories, for the sake of completeness.

Table 2. Classification of learner utterances

Category	Description	Examples
Solution-contributing		
Proof contribution		
Proof step	Contributes a proof step or part of a proof step	“From $A \cap B = \emptyset$ follows: $A \subseteq K(B)$” “Justification: $A \subseteq (U \setminus B)$”
Proof strategy	States a solution strategy to be adopted	“I’m using the Axiom of Extensionality” “Proof by $\subseteq$ and $\supseteq$”
Proof structure	Signals solution structure	“I’m making a case distinction:” “Forward direction:”
Proof status	Signals the status of the (partial) solution	“And so one subset relation is shown.” “q.e.d.”
Meta-level		
Self-evaluation	States an evaluation of own step	“I’ve made a typo.” “Correction:”
Restart	Signals a new attempt at a proof being started	“new start” “Once again from the beginning.”
Give up	Signals abandoning the solving task	“I would like to know the solution” “I’m giving up”
Other		
Request help	Requests assistance	“I need a hint” “How is $R \circ S$ defined?” “am I on the right track?”
Answer	Provides a non-Yes/No answer to a question posed	*T: What are the possible properties of binary relations?* “symmetry” *T: What does the variable x mean?* $\langle u \rangle$”x has two meanings”$\langle / u \rangle$ $\langle u \rangle$”it occurs in two different sets”$\langle / u \rangle$
Address	Provides a non-elicited reaction to a previous contribution	“This answers my question only halfway!” “The bracket could just as well be in a different place if you ask me!”
Agree	Expresses agreement with a statement	“indeed you’re right”
Cognitive state	Expresses the state of knowledge or understanding	“i don’t know what i can do with this hint!” “I know that.”
P/E/A	Politeness/Emotion/Attitude	“Sorry!” “I will exchange you at the shop!”
Session	Expresses a meta-level session-related statement	“Actually Exercise E (as you call it) is called Exercise A here!” “how about postponing Exercise W and starting with A?”
Self talk	Expresses an unelicited comment	“The difference between $=$ und $\cap$ is questionable” “Must have something to do with the difference.”
DM	Discourse Marker	“Right...” “Good then.”
OK	Simple acknowledgment	
Yes/No	“yes” or “no” answer	

Table 3. Descriptive information on learner proof discourse in terms of content modality: symbolic (ME), natural language (NL), and natural language interleaved with symbolic expressions (ME & NL)

	C-I Unique / Total	C-II Unique / Total	C-I ∪ C-II Unique / Total
Turns	147 / 332	497 / 927	628 / 1259
ME	2 / 153	2 / 274	2 / 427
NL	34 / 51	134 / 162	163 / 213
ME & NL	111 / 128	361 / 491	463 / 619
Utterances[1]	200 / 443	531 / 1118	702 / 1561
ME	2 / 189	1 / 300	2 / 489
NL	64 / 92	185 / 278	240 / 370
ME & NL	134 / 162	345 / 540	460 / 702

[1] Non-empty utterances after removing punctuation (see pre-processing in Section 3; A single occurrence of an utterance consisting of a question mark alone (in C-II) is included in the NL category.

Two frequency counts are given in the descriptive statistics tables throughout the rest of this paper: "Total" denotes the number of turn/utterance instances (tokens or "vocabulary size"; where by "vocabulary" here we mean linguistic patterns). "Unique" denotes the number of *distinct* types (unique pattern types). The proportion of these two measures is known as "type-token ratio". The two raw frequencies rather than the summarised measure are provided because the number of tokens is different for each cell in the tables, so the raw counts are more informative.

Aside from the frequency distributions, we plot graphs of frequency spectra. Spectrum visualisations are typically used with word frequencies. They show a frequency distribution in terms of number of types by frequency class, where a frequency class is a set of (sets of) instances with the same number of occurrences in the data. In other words, it shows how many *distinct types* (y-axis) occur once, twice, and so on (x-axis), thus revealing the degree of skewedness of the types distribution; the earlier the tail with y around 1 starts, the more idiosyncratic types are likely to exist in the data. We use verbalisation patterns – pre-processed utterances – as units of analysis.[8]

5.1 Mathematical symbols vs. natural language

The most prominent surface characteristic of mathematical discourse is that it is the familiar mixture of symbols and natural language, the mother tongue of the author or, in case of most of scientific publications, English, which has become the de facto language of science. While, in principle, proofs can be presented using the symbolic language of mathematics alone – as in formal logic, for instance – this presentation style is not common in communicating mathematics. In fact, it has been argued that symbolic notation does not have to dominate in a proof for it to make a "better" proof [16]. There is, however, no "prescribed" presentation style other than guidelines, and even on those authors differ (see [29, 15, 20, 21], to mention just a few).

In the context of learning mathematics, mathematical notation, its mastery, has been shown to be one of the major obstacles in learning to do proofs [26]. Interestingly, the presentation style of the study-material – mainly formal vs. mainly natural language, verbose proof presentation – has an influence on the learners' use of natural language in computer-based tutoring [34]; that is, learners mimic the linguistic style they are presented with. As the first approximation of linguistic variety in learner proof discourse, we therefore analyse the learners' contributions in terms of the two types of content modalities: natural language and symbolic expressions.

Table 4. Distribution of utterance types

	C-I Total	C-II Total	C-I ∪ C-II Total
Solution-contributing	187	548	735
Proof contribution	180	539	719
Proof step	171	469	640
Proof strategy	4	30	34
Proof status	5	24	29
Proof structure	-	16	16
Meta-level	7	9	16
Self-evaluation	2	5	7
Restart	2	3	5
Give up	3	1	4
Other	64	267	331
Request help	16	154	170
Yes/No	18	24	42
Cognitive state	15	16	31
Politeness/Emotion/Attitude	3	21	24
Discourse marker	1	21	22
Answer	5	15	20
OK	1	6	7
Address	1	5	6
Session	-	4	4
Agree	2	1	3
Self talk	2	-	2
Uninterpretable	3	4	7

Table 3 shows the distribution of turns and utterances in both corpora with respect to natural language and symbolic content. ME denotes turns and utterances consisting of symbolic expressions alone, NL those consisting of natural language alone (as in C-II S8), and ME & NL those consisting of natural language interleaved with mathematical expressions (C-I S1 or C-II S29).

In both corpora the majority of turns and utterances contain some natural language (turns: 54% NL/ME & NL vs. 46% ME in C-I and 70% vs. 30%, respectively, in C-II; utterances: 57% NL/ME & NL vs. 43% ME in C-I and 73% and 27%, respectively, in C-II). Only 14 NL/ME & NL turn-level patterns and only 28 utterance-level patterns occur both in C-I and C-II (turn-level: 640 NL/ME & NL patterns in C-I and C-II considered in isolation vs. 626 in C-I ∪ C-II; utterances: 728 in C-I and C-II in isolation vs. 700 in C-I ∪ C-II). There is proportionally more natural language in C-II even though, as shown in [34], the participants in the formal material condition were less verbose than those in the verbose material condition.

Overall, 69% of the utterances in C-I ∪ C-II contain some linguistic material, among which there are 700 distinct utterances (verbalisation patterns). From this point on we focus on a subset of the data: we look at utterances only and only those which do contain natural language. We start by looking at the distribution of utterance types.

5.2 Distribution of utterance types

Table 4 shows the distribution of utterance types, as defined in Section 4, in both corpora.[9] The majority of utterances in both corpora are solution-contributing, 74% of all utterances in C-I and 67% in C-II, and most of them proof steps. This is not surprising of course. The second experiment involved more complex proofs requiring, for

[8] R [27] was used to create for the plots and the zipfR package [14] for the frequency spectra. Only the first 15 frequency classes are shown since in all cases the frequency of the larger classes oscillated between 0 and 5.

[9] Only the utterance types with more than five occurrences will be discussed here. Utterance types with lower frequency of occurrence appear too sparse for any conclusions about their wording.

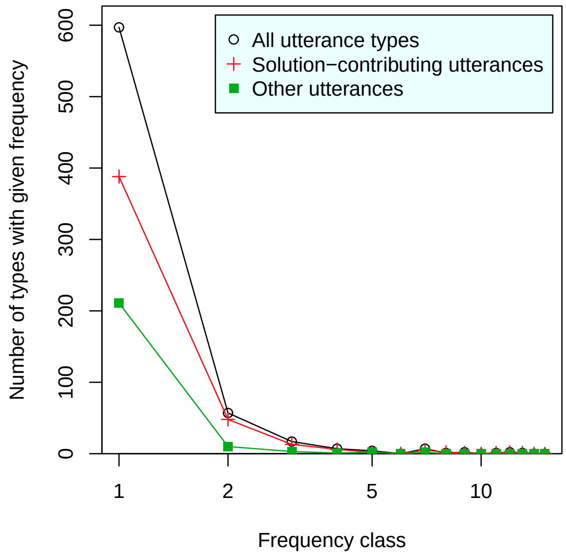


Figure 2. Frequency spectra: Utterance types (x-axis log-scaled)

instance, considering cases and proving both directions of a bi-conditional, which resulted in explicit verbalisations of the proving strategy, the proof structure, and in learners signalling that a complex proof (or its part; e.g. one direction of a bi-conditional) is completed.

Among the non-solution-contributing utterances, the largest class, 51%, are help requests of different specificity; from general requests (such as “Hilfe!” (en. *Help!*) or “Einfaches Beispiel wuerde mir weiter helfen” (en. *A simple example would help me*)) to specific requests of a definition (such as “Wie lautet die Definition der Operation $^{-1}$?” (en. *What’s the definition of* $^{-1}$*?*) or “Erklaere die Definition $R \circ S$ in Worten!” (en. *Explain the definition of* $R \circ S$ *in words!*)), or enquiries whether propositions hold (such as “Ist (a, z) in R?” (en. *Is* (a, z) *in* R*?*) or “Elemente von $(R \circ S) \circ T$ sind Tripel der Form (x, y, z), oder?” (en. *Elements of* $(R \circ S) \circ T$ *are triples of the form* (x, y, z)*, right?*)) The second largest category are closed-class types, Yes/No and OK, which together make up 15% of all the non-solution-contributing utterances.

The second largest category of open-ended verbalisations are meta-cognitive statements on the state of knowledge (or, for the most part, of the *lack* of knowledge), 31 occurrences. Statements such as “Keine Ahnung mehr wie der Nachweis korrekt erbracht werden kann” (en. *No idea how the proof can be correctly produced*) or “Verstehe die definition nicht” (en. *Don’t understand the definition*), can be interpreted as indirect requests of help. Interestingly, only one wording appeared more than once, “Dann weiss ich nicht weiter” (en. *So I’m lost*).

Aside from the two common variants of expressions of gratitude (“Danke”/“Vielen Dank” (en. *Thank you/Thank you very much*)) and the four common German variants of apologies (“Tut mir leid”/“Entschuldigung”/“Verzeihung”/“Sorry”), the remaining expressions of emotions and attitude (Politeness/Emotion/Attitude class) were idiosyncratic and unpredictable, and spanned both positive polarity emotions, for instance, “Das macht Spass mit Dir” (en. *It’s fun with you!*) and negative polarity (“Wollen Sie mir nun Mathematik beibringen oder wollen Sie mich pruefen???” (en. *Do you want to teach me math now or do you are you giving me a test???*), “NERV!!” (en. *[anger]*)). Not surprisingly, idiosyncratic were also the occurrences of the remaining open-ended classes, answers and addresses, whose content is entirely determined be the preceding context, i.e. the tutor’s contribution which triggered them.

What is interesting is that there were 22 occurrences of discourse markers, the kinds typical of spoken language: “na ja” (en. *oh well*), “oh”, “hm”. The variety of discourse markers suggests that computer-mediated dialogue was treated by the subjects much like natural spoken interaction, even though it was type-written.

Figure 2 shows the frequency spectra of all the utterance types and of the two major utterance classes. It is clear from the plot that the distribution of distinct verbalisations is heavily skewed. For all sets of utterance types, already the number of patterns with at least between three to five occurrences is less than 10. The tail of patterns with frequency 1 starts between 5-10 or more occurrences.

Frequency spectra also show that the data is sparse and even though some utterance types have a high frequency of occurrence (Table 4) they consist of mainly idiosyncratic linguistic patterns. Of course, most interesting from the point of view of formalisation are the core argumentative utterances which build up a proof. Thus, we now take a closer look at the verbalisations of proof contributions.

5.3 Proof contributions

Since we are interested in the diversity of wording, we first consider the type of content that proof contributions verbalise. Considering that the *ultimate* goal of this work is to computationally translate the natural language verbalisations into a formal language of a deduction system, aside from the three classes of proof-level descriptions – proof strategy, proof structure, and proof status (see Table 2) – three classes of proof steps are distinguished in the analysis that follows. The sub-categorisation of proof steps takes into account, on the one hand, the type of content the natural language expresses and, on the other hand, the type of linguistic knowledge which needs to be encoded in order for formalisation to be possible.

The simplest case for translation are steps in which natural language is used only for logical operators (connectives and binders/quantifiers), to signal proof step components, and where no discourse context nor domain-specific linguistic information is needed. By proof step components we mean elements of a deduction system’s proof language such as the declarative proof script language presented in [2]. In order to formalise proof steps of this kind, the only linguistic knowledge needed is the natural language vocabulary and syntax of logical connectives and of the proof structural components (proof discourse connectives); that is, only a basic interpretation lexicon. Examples of this class of proof steps include:[10]

(10) Wenn $A \subseteq K(B)$, dann $A \cap B = \emptyset$ [C-I S1]

(11) Sei $(a, b) \in (R \circ S)^{-1}$ [C-II S2]

We will refer to this class of steps as *NL logic & proof step components* which stands for “natural language logical connectives and proof step components”.

The second and third class of proof steps are those which require context and linguistic domain knowledge for interpretation and formalisation: if beyond the type of content described above, only domain concepts from the domain(s) to which the proof refers (here: set theory and binary relations) and discourse-specific references have to be translated, then the proof step belongs to the second category,

[10] The example sentences are worded here as they occur in the corpus. For the analysis, they have been pre-processed as described in Section 3.

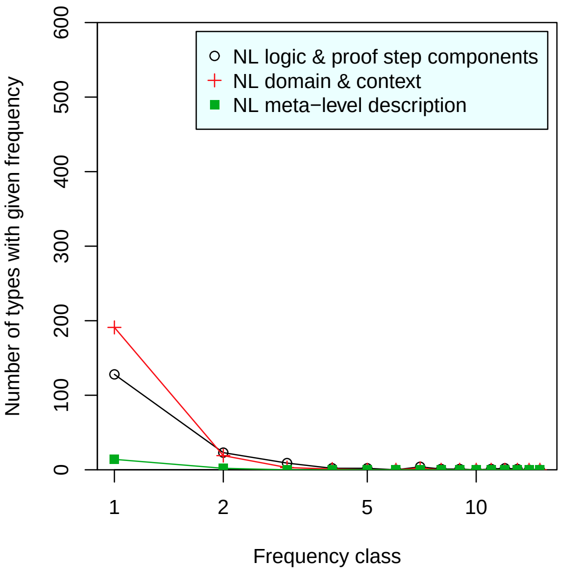


Figure 3. Frequency spectra: Proof step types (x-axis log-scaled; y-axis range extended to match Figure 2 for comparison)

to which we will refer to as *NL domain & context*. The verbalised domain concepts may be single and multi-word domain terms[11] but also informal verbalisations of domain relations, such as the locative prepositional phrase with "in" for set membership. Examples of the second class of proof steps include:

(12) in $K(B)$ sind alle x, die nicht in B sind [C-I S5]

(13) Nach der Definition von $\circ$ folgt dann (a, b) ist in $S^{-1} \circ R^{-1}$ [C-I S6]

(14) Nach Aufgabe A gilt $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$ [C-II S25]

In C-II S25 the reference "Aufgabe A" needs to be resolved. Note, however, that the utterance "Es gilt nach Definition ausserdem $S^{-1} \circ R^{-1} = (x, y)|\exists z(z \in M \wedge (x, z) \in S^{-1} \wedge (z, y) \in R^{-1})$" (en. *By the definition it moreover holds that* ...) belongs to the first class, *NL logic & proof step components*: no domain-specific vocabulary is used; the word "definition" is in the basic lexicon of mathematics.[12]

Finally, the third class comprise those steps which are not specified explicitly, but rather indirectly as high-level meta-descriptions of a (possibly complex) transformation which needs to be performed in order to reconstruct the intended step. An example of such as complex proof step is C-II S8. Other examples include:

[11] See the paragraph on normalisation of domain terms and domain-specific references in Section 3.3

[12] The verbalisation-oriented proof step classification proposed in [31], while similar to ours and designed with a similar motivation, is imprecise. First, it is not clear whether the class *simple connections* would accommodate utterances with adverbs or adverbial phrases, such as "Moreover, as previously shown, it follows that ..." Second, and more importantly, the distinction between *weakly verbalised* and *strongly verbalised* formulas is unclear based on the definitions given. *Weakly verbalised* formulas are defined as those "where some relations or quantifiers are partly verbalised", while *strongly verbalised* formulas as those "where all relations and quantifiers are fully verbalised". Based on these definitions it is not clear why the example "a is the limit of $(a_n)_{n \in N}$", given in the paper, should be classified as *weakly verbalised*, whereas "For all ϵ holds: there exists a $n_0(\epsilon) \in N$ with ..." as *strongly verbalised*; clearly, the set membership relation in $n_0(\epsilon) \in N$ is not verbalised.

Table 5. Descriptive information on proof contributions

	C-I Unique / Total	C-II Unique / Total	C-I ∪ C-II Unique / Total
Proof step	138/ 171	287 / 469	407 / 640
NL logic & proof step components	54 / 80	136 / 286	175 / 366
NL domain & context	78 / 85	140 / 171	216 / 256
NL meta-level description	6 / 6	11 / 12	16 / 18
Proof strategy	4 / 4	25 / 30	29 / 34
Proof structure	- / -	7 / 16	7 / 16
Proof status	1 / 5	7 / 24	7 / 29

(15) Analog geht der Fall, wenn $(a, z) \in S$.
(en. *The case for* $(a, z) \in S$ *is analogous*)

(16) de morgan regel 2 auf beide komplemente angewendet
(en. *de morgan rule 2 applied to both complements*)

(17) $(S \circ T)$ ist genauso definiert
(en. $(S \circ T)$ *is defined the same way*).

Complex proof steps of this kind will be referred to as *NL meta-level description*.

Table 5 shows descriptive statistics on proof contributions with proof steps sub-classified as described above. Not surprisingly, the wording of two types of proof contributions which refer to the proof-level concepts – proof strategy and proof structure – is diverse. Wording of proof status information is repetitive; indeed, most often only the end of the proof is signalled explicitly and most often using the abbreviation "q.e.d." Now, also not surprisingly, within the class of proof steps, the more complex the content, the more varied the wording. Meta-level descriptions of proofs are almost entirely idiosyncratic. Only two utterance patterns occurred more than once: "MATHEXPR ist analog definiert" (en. MATHEXPR *is defined analogously*) and "das gleiche gilt fuer MATHEXPR" (en. *The same holds for* MATHEXPR). The wording of proof steps in the *NL domain & context* category is also diverse: 92% of instances are distinct in C-I, 82% in C-II, and 84% overall. Most repetitive patterns are found in the *NL logic & proof step components* class: 67% of all utterance instances in this category are distinct in C-I, only 47% in C-II, and 48% in both corpora combined. Overall, 63% of proof steps (from the three categories) are distinct.

Figure 3 shows the frequency spectra of the three proof step categories in C-I ∪ C-II. Again, the distribution of verbalisation patterns is heavily skewed. In the largest category, *NL domain & context*, 210 out of the 216 unique patterns occur only once or twice; that is 97%. In the *NL logic & proof step components* category, around 150 out of the 175 unique patterns, 73%, occur once or twice. However, within this class there are 8 patterns with at least five instances of occurrence. Table 6 shows the top-10 most frequent linguistic patterns in the three classes of proof steps from the combined corpus, C-I ∪ C-II, with their frequency of occurrence.

5.4 Growth of the diversity of forms

Finally, we are interested in how the diversity of forms evolves with the number of conducted dialogues. Specifically, we would like to know how many dialogues are needed to have observed most of the verbalisation patterns. Figure 4 shows a plot of a variant of the type-token (vocabulary growth) curve [36]. On the x-axis is the number of dialogues seen. Rather than the raw type count, the y-axis shows the proportion of observed pattern types out of all pattern types in

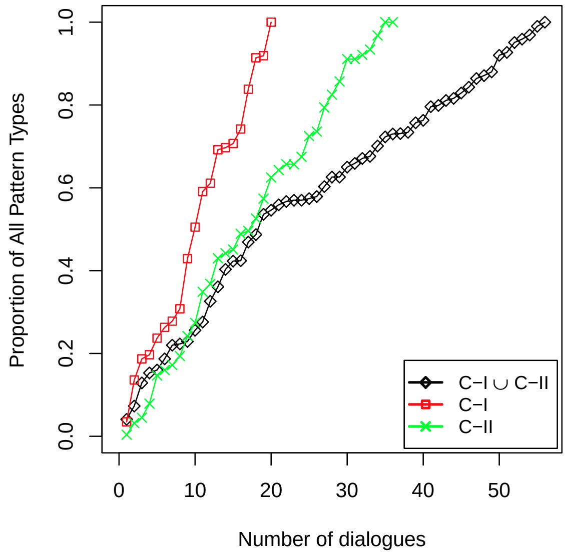


Figure 4. Growth of the utterance patterns over random dialogue sequences

Table 6. Top-10 most frequent utterance patterns expressing proof steps

Linguistic pattern	Frequency
Proof step	
NL logic & proof step components	
sei MATHEXPR	54
es gilt MATHEXPR	13
wenn MATHEXPR dann MATHEXPR	12
also MATHEXPR	12
dann ist MATHEXPR	11
also ist MATHEXPR	9
MATHEXPR und MATHEXPR	8
MATHEXPR ist dann MATHEXPR	7
daraus folgt MATHEXPR	7
daraus folgt dass MATHEXPR	7
NL domain & context	
nach REFERENCE MATHEXPR	7
DOMAINTERM	7
nach REFERENCE ist MATHEXPR	4
MATHEXPR nach REFERENCE	3
DOMAINTERM von MATHEXPR ist DOMAINTERM MATHEXPR	3
aus REFERENCE folgt MATHEXPR	3
wegen der formel fuer DOMAINTERM folgt MATHEXPR	2
oder MATHEXPR wegen DOMAINTERM von MATHEXPR	2
nach REFERENCE gilt MATHEXPR	2
nach DOMAINTERM gibt es ein MATHEXPR mit MATHEXPR	2
NL meta-level description	
MATHEXPR ist analog definiert	2
das gleiche gilt fuer MATHEXPR	2
gleiches gilt mit MATHEXPR	1
DOMAINTERM auf beide DOMAINTERM angewendet	1
der fall MATHEXPR verlaeuft analog	1
der beweis von MATHEXPR ist analog zum beweis von MATHEXPR	1
beweis geht genauso wie oben da in REFERENCE bis REFERENCE nur DOMAINTERM umformungen stattfinden	1
analog geht der fall wenn MATHEXPR	1
andersrum	1
die zweite DOMAINTERM ergibt sich aus der umkehrung aller bisherigen beweisschritte	1

the given corpus.[13] The order of dialogues in C-I and C-II has been randomised. For the C-I ∪ C-II plot, the corpora were combined and a random sequence drawn from the combined set.

What can be seen from the graphs is that the pattern vocabulary grows linearly (given the random sample drawn). The tendency is similar in both corpora: half of the patterns have been seen at about 40% of the data sets and 80% of the patterns at about 77% into the data set in C-I (ca. 17 dialogues) and 70% in C-II (ca. 26 dialogues). In the combined corpus, however, half of the patterns have been seen already about 32% into the data set. 80% of the patterns have been seen about 70% into the data set (ca. 41 dialogues).

6 DISCUSSION

First, it is clear from the results that the language of learner discourse in proofs is not as repetitive as one might expect. Learners use complex natural language utterances not only during meta-communication with the tutor, but also when contributing proof steps. 57% of all utterances in C-I and 73% in C-II contained some natural language. The fact that natural language was more often used in the C-II corpus may be explained by the fact that the binary relations proofs were more complex than the set theory proofs. However, set theory is very naturally expressed in natural language, so the reason why this was the case needs further investigation.

Second, the wording of proof steps is surprisingly diverse and the language used in the two corpora is different. The fact that there were only 28 utterance verbalisations which occurred in both data sets is surprising.[14] This low number of common patterns is reflected in the type-token plot (Figure 4) which exhibits a steady increase with only one area of slower growth in the combined corpus, about 20-25% into the randomly-ordered data set.

The difference in the linguistic diversity of the proof language (the proof contributions class) in the two corpora can be also seen in the different distributions of distinct linguistic patterns (Table 5). Among the *logic & proof step components* class, 67% of the verbalisations were distinct in C-I and 47% in C-II. In the *domain & context* class, 92% of all the verbalisations were distinct in C-I and 82% in C-II. That is, the language in C-II appears more repetitive. In both corpora, however, the language in the latter class of proof steps is more heterogeneous than in the former. The frequency spectra and the pattern growth curves show further the degree to which the language is indeed diverse. In the *logic & proof step components* class, 81% of the distinct types were single-occurrence utterances (81% in C-I and 72% in C-II). In the *domain & context* class, 90% of the types were single-occurrence (96% in C-I and 85% in C-II).

Not surprisingly, the majority of the meta-level communication are the learners' requests for assistance: requests for hints, definitions, explanations, etc. As these are not the core argumentative utterance types, we did not present a detailed analysis here, however, to roughly illustrate the diversity of wording it is enough to mention that out of the 170 help requests, 149 (88%) were distinct verbalisations. 136 of these were single-occurrence patterns. A further subclassification of help requests might reveal more homogeneity in the wording within the subcategories.

The relatively large number of discourse markers, typical of spoken interaction, suggests that participants had an informal approach

[13] 198, NL + ME & NL, utterance patterns in C-I, 530 in C-II, and 700 in C-I ∪ C-II; see Table 3.

[14] 8 were from the non-solution contributing class and 20 were proof step verbalisations, the majority from the *logic & proof step components* class.

to dialogue style and treated it much like a chat, adapting spoken language they would have otherwise used in a natural setting to the experiments' type-written modality; this is a known phenomenon [18]. The diversity of verbalisations may be partly due to this.

7 CONCLUSIONS AND FURTHER WORK

We have shown that even this seemingly linguistically predictable argumentative domain of mathematical proofs is characterised by a large variety of linguistic patterns of expression and by a large number of idiosyncratic verbalisations and that the meta-communicative part of discourse which does not directly contribute to the solution has an conversational character, suggesting learners' informal attitude towards the computer-based dialogues and high expectations on the input interpretation resources. This calls for a combination of shallow and deep semantic processing methods for the discourse in question: shallow pattern-based approaches for contributions which do not add to the proof and deep lexicalised grammars for the proof-relevant content, in order to optimise coverage. At the time of writing a parsing grammar for German we have been developing is capable of analysing all the linguistic proof-contribution structures which occurred more than twice in the data. Future work will proceed in two directions: (i) we will continue to improve the grammar coverage and (ii) we have started pre-processing proofs from a corpus of open-access scientific publications (in English) in order to perform an analogous analysis of language variety in expert proof discourse.

REFERENCES

[1] D. Almeida, 'A survey of mathematical undergraduates' interaction with proof: Some implications for mathematics education', *International Journal of Mathematical Education in Science and Technology*, **3**(6), 869–890, (2000).

[2] A. Autexier, D. Dietrich, and M. Schiller, 'Towards an Intelligent Tutor for Mathematical Proofs', in *Proceedings of the 1st Workshop on CTP Components for Educational Software*, eds., Pedro Quaresma and Ralph-Johan Back, volume 79 of *EPTCS*, pp. 1–28, (2012).

[3] N. Baron, 'Language of the Internet', in *The Stanford Handbook for Language Engineers*, ed., Ali Farghali, 59–127, Stanford, CA, (2003).

[4] L. Becker, W. Ward, S. Van Vuuren, and M. Palmer, 'DISCUSS: A dialogue move taxonomy layered over semantic representations', in *Proceedings of the 9th International Conference on Computational Semantics (IWCS-11)*, pp. 310–314, (2011).

[5] C. Benzmüller, H. Horacek, H. Lesourd, I. Kruijff-Korbayová, M. Schiller, and M. Wolska, 'A corpus of tutorial dialogs on theorem proving; the influence of the presentation of the study-material', in *Proceedings of the 5th LREC*, pp. 1766–1769.

[6] D. Biber, *Variation across speech and writing*, CUP, 1988.

[7] W. Billingsley and P. Robinson, 'Student Proof Exercises Using MathsTiles and Isabelle/HOL in an Intelligent Book', *Journal of Automated Reasoning*, **39**(2), 181–218, (2007).

[8] M. Buckley and M. Wolska, 'A classification of dialogue actions in tutorial dialogue', in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling-08)*, pp. 73–80, (2008).

[9] G.C. Campbell, N.B. Steinhauser, M.O. Dzikovska, J.D. Moore, C.B. Callaway, and E. Farrow, 'DeMAND coding scheme: A "common language" for representing and analyzing student discourse', in *Proceedings of 14th AIED Conference (AIED-09)*, pp. 665–667, (2009).

[10] W. Chafe and Deborah T., 'The Relation between Written and Spoken Language', *Annual Review of Anthropology*, **16**, (1987).

[11] D. Crystal, *Language and the Internet*, CUP, 2001.

[12] T Dreyfus, 'Why johnny can't prove', *ESM*, **38**, 85–109, (1999).

[13] M. Dzikovska, D. Reitter, J. Moore, and C. Zinn, 'Data-driven Modeling of Human Tutoring in Calculus', in *Proceedings of the Language-enhanced Educational Technology Workshop*, pp. 22–28, (2006).

[14] S. Evert and M. Baroni, '*zipfR*: Word frequency distributions in R', in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 29–32, (2007). (R v. 0.6-6 of 2012-04-03).

[15] L. Gillman, *Writing Mathematics Well: A Manual for Authors*, Mathematical Association of America, 1987.

[16] Paul R. Halmos, 'How to write mathematics', *L'Énseignement Mathématique*, **16**, (1970).

[17] M. Hendriks, C. Kaliszyk, F. van Raamsdonk, and F. Wiedijk, 'Teaching logic using a state-of-art proof assistant', *Acta Didactica Napocensia*, **3**(2), 35–48, (2010).

[18] Y. Hård af Segerstad, *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*, Ph.D. dissertation, Göteborg University, 2002.

[19] J.F. Kelley, 'An iterative design methodology for user-friendly natural language office information applications', *ACM Transactions on Office Information Systems*, **2**(1), 26–41, (1984).

[20] D.E. Knuth, T. Larrabee, and R.M. Roberts, *Mathematical Writing*, MAA Notes, Mathematical Association of America, 1989.

[21] Steven G. Krantz, *A Primer of Mathematical Writing*, AMS, 1997.

[22] J. Marineau, P. Wiemer-Hastings, D. Harter, B. Olde, P. Chipman, A. Karnavat, V. Pomeroy, S. Rajan, and A. Graesser, 'Classification of speech acts in tutorial dialogue', in *Proceedings of the ITS-00 Workshop on Modeling Human Teaching Tactics and Strategies*, pp. 65–71, (2000).

[23] N. Maynor, 'The Language of Electronic Mail: Written Speech?', in *Centennial Usage Studies*, 48–54, (1994).

[24] J. McDonald, 'The EXCHECK CAI system', in *University-level computer-assisted instruction at Stanford: 1968–1980*, ed., Patrick Suppes, 765–790, Stanford University, Institute for Mathematical Studies in the Social Sciences, Stanford, CA, (1981).

[25] D. McMath, M. Rozenfeld, and R. Sommer, 'A computer environment for writing ordinary mathematical proofs', in *Proceedings of the 8th Conference on Logic for Programming*, pp. 507–516, (2001).

[26] R.C. Moore, 'Making the transition to the formal proof', *ESM*, **27**, 249–266, (1994).

[27] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[28] A. Selden and J. Selden, 'Errors and misconceptions in college level theorem proving', Technical Report 2003–3, Tennessee Technological University, Cookeville, TN, (2003).

[29] N.E. Steenrod, P.R. Halmos, M.M. Schiffer, and J.A. Dieudonné, *How to write mathematics*, Americal Mathematical Society, 1981.

[30] Andrzej Trybulec and Piotr Rudnicki, 'Using Mizar in Computer Aided Instruction of Mathematics', in *Norwegian-French Conference of CAI in Mathematics*, (1993). Retrieved on May 17 from `http://www.mizar.org/project/oslo.pdf`.

[31] M. Wagner and H. Lesourd, 'Using TeX macs in Math Education: An exploratory Study', in *Proceedings of MathUI-08*, (2008). Retrieved on May 27, 2012 from `http://www.activemath.org/workshops/MathUI/08/proceedings/iMathEU.html` (Online proceedings).

[32] M. Wolska and M. Buckley, *Text Resources and Lexical Knowledge, Selected Papers from the 9th Conference on Natural Language*, volume 8 of *Text, Translation, Computational Processing*, chapter A Taxonomy of Task-related Dialogue Actions: The Cases of Tutorial and Collaborative Planning Dialogue, 105–118, Mouton De Gruyter, 2008.

[33] M. Wolska and I. Kruijff-Korbayová, 'Analysis of Mixed Natural and Symbolic Language Input in Mathematical Dialogs.', in *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL-04)*, pp. 25–32, (2004).

[34] M. Wolska and I. Kruijff-Korbayová, 'Factors influencing input styles in tutoring systems: the case of the study-material presentation format', in *Proceedings of the ECAI-06 Workshop on Language-enabled Educational Technology*, pp. 86–91, (2006).

[35] M. Wolska, I. Kruijff-Korbayová, and H. Horacek, 'Lexical-semantic interpretation of language input in mathematical dialogs', in *Proceedings of the ACL 2nd Workshop on Text Meaning and Interpretation*, pp. 81–88, (2004).

[36] G. Youmans, 'Measuring Lexical Style and Competence: The Type-Token Vocabulary Curve', *Style*, **24**, 584–0599, (1990).

Developing Software for Training Argumentation Skills

Mare Koit[1]

Abstract. The paper introduces a dialogue model and software we have been developing for training the user´s argumentation skills. We consider dialogues in natural language where one participant *A* is influencing his partner *B* to make a decision about performing an action *D*. In communication, *A* presents various arguments for *D* in order to direct *B*´s reasoning process; he stresses the positive and down-grades the negative aspects of *D*. When playing *B*´s role, the user can develop her skills - how to oppose, how to avert the partner´s arguments, and how to find suitable counter-arguments.

1 INTRODUCTION

When one person initiates communication with another (s)he proceeds from the fact that the partner is a human being who feels, reasons and has wishes and plans like every human being. In order to be able to foresee what processes will be triggered in the partner after a move and what will the outcome of these be, the agent must know the inner workings of the partner´s psychological mechanisms [10]. When aiming at a certain goal in communication, the agent must know how to direct the functioning of these mechanisms in order to bring about the intended result in the partner. This knowledge forms a necessary part of human communicative competence. Without it the intentional, goal-directed communication is impossible.

We are considering dialogues where the goal of one of the partners, *A*, is to get another partner, *B*, to carry out a certain action *D*. This type of dialogue constitutes one kind of so-called agreement negotiation dialogues [12]. Such a dialogue can be considered, on a more general level, as rational behaviour of conversational agents which is based on beliefs, desires and intentions of agents, at the same time being restricted by their resources [7].

Because of this, we have modeled the reasoning processes that people supposedly go through when working out a decision whether to perform an action or not. In a model of conversational agent its cognitive states as well as cognitive processes will be represented. One of the most well-known models of this type is the BDI (belief-desire-intention) model [1, 2].

The paper has the following structure. Section 2 introduces our dialogue model that includes a reasoning model. The model is implemented as a conversational agent, which interacts with a user in Estonian and can be used for training argumentation skills of the user. Section 3 represents interaction as update of information states of the conversational agent. Section 4 discusses some implementation details and section 5 draws conclusions.

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

2 MODELLING ARGUMENTATION DIALOGUE

Let us consider dialogue between two participants (*A* and *B*) in a natural language. In the goal base of the initiator (let it be *A*) a certain goal G^A related to *B*´s activities gets activated. In constructing his first turn (request the partner to perform some action *D*), *A* must plan the dialogue acts and determine their verbal form as a turn tr_1^A. The partner *B* interprets *A*´s turn and generates her own response tr_1^B. *B*´s response triggers in *A* the same kind of process in the course of which he has to evaluate how the realization of his goal G^A has proceeded, and depending on this he may activate a new sub-goal of G^A, and the cycle is repeated: *A* builds a new turn tr_2^A. Dialogue comes to an end, if *A* has reached his goal or abandoned it.

After *A* has requested *B* to perform *D*, *B* can respond with agreement or rejection, depending on the result of her reasoning about the action. *B*´s rejection can be supported with an argument. These arguments can be used by *A* as giving information about the reasoning process that brought *B* to the (negative) decision.

2.1 Reasoning Model

Our reasoning model is based on the studies in the common-sense conception of how the human mind works in such situations, cf. [4] since in natural communication people depart from this conception, not from any scientific one.

In our model we try to reflect the main types of determinants that motivate humans to act. Thus the strategy used depends on which determinant is chosen as the target of influence.

In general lines our reasoning model follows the ideas realized in the BDI model. But it has a certain particular feature - we want to model a ´´naïve´´ theory of reasoning, a ´´theory´´ that people use when interacting with other people and trying to predict and influence their decisions [5].

The reasoning model consists of two parts [9]: (1) a model of human motivational sphere; (2) reasoning procedures. In the motivational sphere three basic factors that regulate reasoning of a subject concerning *D* are differentiated. First, subject may *wish* to do *D*, if pleasant aspects of *D* for him/her overweight unpleasant ones; second, subject may find reasonable to do *D*, if *D* is *needed* to reach some higher goal, and useful aspects of *D* overweight harmful ones; and third, subject can be in a situation where (s)he *must* (is obliged) to do *D* - if not doing *D* will lead to some kind of punishment. We call these factors WISH-, NEEDED- and MUST-determinants, respectively.

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

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

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

Resources of the subject concerning *D* constitute any kinds of internal and external circumstances which create the possibility to perform *D* and which are under the control of the reasoning subject.

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

Of course, in reality people do not operate with numbers. Anyway, existence of certain scales also in human everyday reasoning is apparent. For instance, for the characterization of pleasant and unpleasant aspects of some action there are specific words: *enticing, delightful, acceptable, unattractive, displeasing, repulsive* etc. We may suppose that each of these adjectives can be expressed quantitatively.

The second part of the reasoning model consists of reasoning procedures that supposedly regulate human action-oriented reasoning. A reasoning procedure represents steps that the subject goes through in his/her reasoning process; these consist in computing and comparing the weights of different aspects of *D*; and the result is the decision to do *D* or not.

The reasoning depends on the determinant which triggers it (WISH, NEEDED or MUST). In addition, a reasoning model, as a naïve theory of mind, includes some principles which represent the interactions between determinants and the causal connection between determinants and the decision taken. For instance, the principles fix such concrete preferences as:

- people want pleasant states and do not want the unpleasant ones
- people prefer more pleasant states to less pleasant ones.

We do not go into details concerning these principles here. Instead, we refer to [9].

As an example, let us present a reasoning procedure which is triggered by WISH-determinant, that is, if the subject believes that it would be pleasant to do *D* (JSP diagram in Fig. 1). WISH-determinant gets activated when a reasoning subject finds that the action *D* itself or some of its consequences would be pleasurable to him/her, i.e. *w(pleasant) > w(unpleasant)*.

In the case of other input determinants (NEEDED, MUST) the general structure of the algorithm is analogous, but there are differences in certain steps.

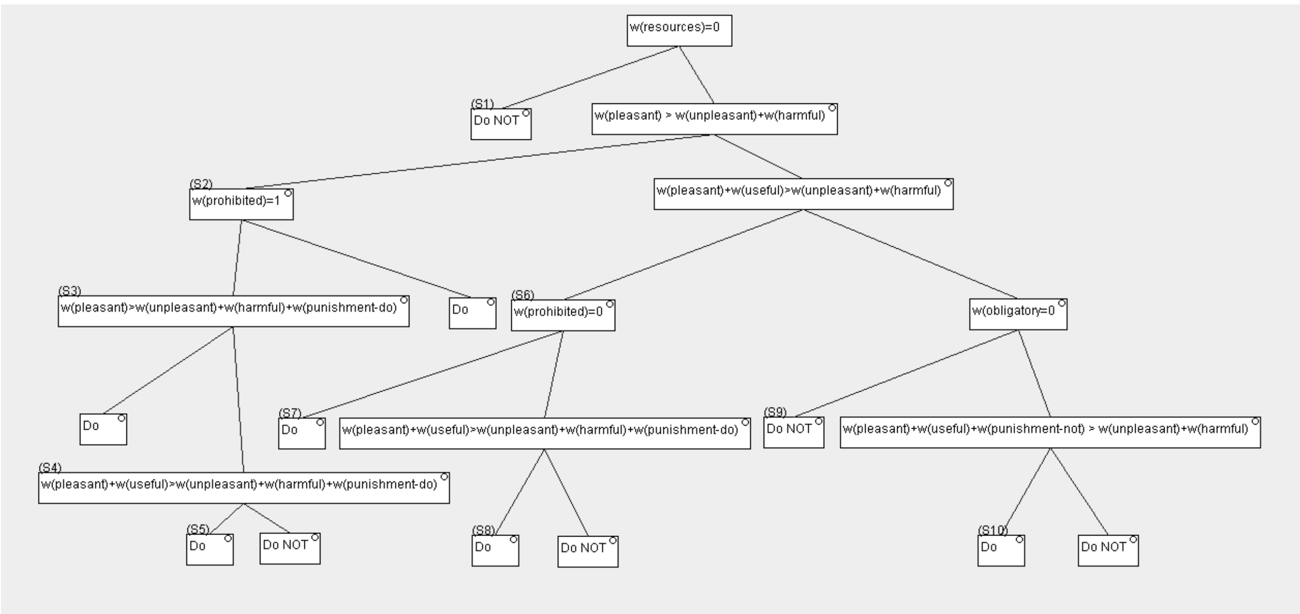


Figure 1. Reasoning procedure WISH.

When comparing our model with BDI model, then beliefs are represented by knowledge of the reasoning subject with reliability less than 1; desires are generated by the vector of weights *w*; and intentions correspond to the goals in the goal base. In addition to desires, from the vector of weights we also can derive some parameters of the motivational sphere that are not explicitly conveyed by the basic BDI model: needs, obligations and prohibitions.

The vector(s) w^{AB} (*A*´s beliefs concerning *B*´s evaluations, where *B* denotes agent(s) *A* may communicate with) are used as partner model(s).

2.2 Communicative Strategies and Tactics

A communicative strategy is an algorithm used by a participant for achieving his/her goal in interaction. The initiator of communication (the participant *A*) can realize his communicative strategy in different ways: stress pleasant aspects of *D* (i.e. *entice* the partner *B*), stress usefulness of *D* for *B* (i.e. *persuade B*), stress punishment for not doing *D* if it is obligatory (*threaten B*), etc. We call communicative tactics these concrete ways of realization of a communicative strategy. *A*, trying to direct *B*´s reasoning to the positive decision (to perform *D*), proposes arguments for doing *D* while *B*, when opposing, proposes counter-arguments.

The simplest tactic, which *A* can use in order to achieve his communicative goal is so-called *defense* - here he does not stress any positive aspects of performing *D* for *B* but only averts (downgrades) counter-arguments presented by *B*. For example, in the following dialogue excerpt (1), *B* repeatedly points to missing resources while *A* tries to indicate how the resources can be obtained [8].

(1)
A: Please prepare a potato salad.
B: I do not have enough time.
A: I will help you.
B: My mother is waiting for me.
A: Call home.

3 CONVERSATION AS UPDATE OF INFORMATION STATES

Several dialogue management architectures have been implemented in dialogue systems [6]. The most powerful are information-state dialogue managers [11]. Information state represents cumulative additions from previous actions in the dialogue, motivating future actions of the conversational agent. The functions of the dialogue manager can be formalized in terms of information state update. In our software, we use information state architecture.

3.1 Representation of Information States

The key of an information state is the partner model, which is changing during the interaction.

There are two parts of an information state of a conversational agent *A* - private (information accessible only for the agent) and shared (accessible for both participants). The private part consists of the following information slots [8]:

- current partner model (vector w^{AB} of weights - *A*´s picture about *B*)
- a tactic t_i^A which *A* has chosen for influencing *B*
- a reasoning procedure r_j which *A* is trying to trigger in *B* and bring it to a positive decision (it is determined by the chosen tactic, e.g. when enticing, *A* triggers the reasoning procedure *wish* in *B*)
- stack of (sub-)goals under consideration. At the beginning, *A* puts its initial goal into the stack (*B* decides to do *D*). In every information state, the stack contains an aspect of *D* under consideration (e.g. while *A* is enticing *B* then pleasantness is on the top)
- set of dialogue acts $DA=\{d_1^A, d_2^A,..., d_n^A\}$. There are the following DA-s for *A*: proposal, assessments for increasing or decreasing weights of different aspects of *D* for *B*, etc.
- (finite) set of utterances as verbal forms of DA-s, incl. utterances for increasing or decreasing the weights (arguments for/against) $U=\{u_{i1}^A, u_{i2}^A,..., u_{iki}^A\}$. Every utterance has its own weight - numerical value: $V=\{v_{i1}^A, v_{i2}^A,..., v_{iki}^A\}$ where v_{i1}^A, etc. is the value of u_{i1}^A, etc., respectively. Every argument can be chosen by *A* only once.

The shared part of an information state contains

- set of reasoning models $R=\{r_1,..., r_k\}$
- set of tactics $T=\{t_1, t_2,..., t_p\}$
- dialogue history - the utterances together with participants´ signs and dialogue acts $p_1{:}u_1[d_1]$, $p_2{:}u_2[d_2]$, ..., $p_i{:}u_i[d_i]$ where $p_1=A$, p_2, etc. is *A* or *B*.

3.2 Update Rules

There are different categories of update rules which will be used for moving from the current information state into the next one:

I. rules used by *A* in order to interpret *B*´s turns

II. rules used by *A* in order to generate its own turns:

1) for the case if the ´´title´´ aspect $a^*(t_i)$ of the current tactic t_i is located on top of the goal stack (e.g. if the tactic is enticing then the ´´title´´ aspect is pleasantness)
2) for the case if another aspect a_j is located on the ´´title´´ aspect of the current tactic t_i (e.g. if *A* is trying to increase the pleasantness of *D* for *B* but *B* argues for usefulness, then the usefulness lies over the pleasantness)
3) for the case if there are no more utterances for continuing the current tactic (and a new tactic should be chosen if possible)
4) for the case if *A* has to abandon its goal
5) for the case if *B* has made the positive decision and therefore, *A* has reached the goal.

Specific rules of the category II exist for updating the initial information state.

For example, the rules of category I have the following general form:

`IF the current tactic is` t_i `AND B´s last utterance was about D´s aspect` a_j `THEN` **`put`**`, firstly, the ´´title´´ aspect` $a^*(t_i)$ `and secondly,` a_j `into the goal stack.`

Generating a response turn, *A* has, firstly, to attack *B*´s argument concerning the aspect a_j and secondly, to stress the ´´title´´ aspect $a^*(t_i)$ of the current tactic t_i taking them off from the stack in the reverse order.

For another example, the general form of the rules of category II-2 is as follows:

`IF the current tactic is` t_i `AND` a_j `is on the top of the goal stack AND` $a^*(t_i)$ `lies under the top aspect in the goal stack AND there are utterances for decreasing` $w(a_j)$ `by` x `units AND there are utterances for increasing` $w(a^*(t_i))$ `by` y `units AND reasoning triggered` `by the determinant` $a^*(t_i)$ `on the changed partner model gives a decision ´´do` D`´´ THEN` **`choose`** `these utterances (and the corresponding dialogue acts) AND` **`eject`** a_j `and` $a^*(t_i)$ `from the goal stack.`

In such a case, *B* has presented the counter-argument against performing *D,* concerning the aspect a_j. *A* has, firstly, to attack this counter-argument and secondly, to stress the ´´title´´ aspect of the current tactic.

4 SOFTWARE DEVELOPMENT

Our software is implemented in two versions which differentiate from each other mainly by involvement of linguistic processing. In both variants, the computer plays *A*´s role.

In the first version, there are ready-made expressions for both the computer and the user, each of which represents an Estonian sentence. Consequently, the computer does not make any morphological and syntactic analysis or generation of texts and does not use any linguistic knowledge. Semantic analysis and generation are extremely simplified by classifying all the expressions. For example, sentences informing about the communicative goal (*The firm offers you a trip to Venice.*), sentences stressing/downgrading the pleasant/unpleasant/useful etc. aspects of an action (*The nature is very beatiful there.*, *You must pay the travel costs yourself.*, etc.), affirming sentences (*OK, I shall go*.), etc. Semantic analysis/generation of an expression means only determining its semantic class. The files of Estonian sentences can easily be substituted with their translations and thus interaction can take place in another language.

In the second version of the software, there are ready-made expressions only for the computer. The user can put in free text which will be analysed by the computer. Speech recognition and speech synthesis of Estonian are not included.

4.1 Interaction with the User

At the beginning of a communication process the computer (*A*) chooses tactics (of enticing, convincing or threatening) and generates (randomly) a model of the partner, according to which the corresponding reasoning procedure (*wish, needed* or *must*) yields a positive decision, i.e. the computer presupposes that its partner (user *B*) can be influenced this way. A dialogue begins by an expresssion of the communicative goal (this is the first turn tr_1^A of the computer). If the user refuses (after his reasoning: to perform *D* or not by implementing a normal human reasoning which we are trying to model here), the computer recognizes on the basis of the response (tr_1^B) the step in the reasoning procedure where the reasoning forked into the ´´negative branch´´, determines the aspect of *D* the weight of which does not match the reality and changes this weight so that a new model will give a negative result as before but it is an extreme case: if we increased this weight (in case of positive aspects of D) or decreased it (in case of negative ones) by one unit we should get a positive decision. We suppose that *A*´s each expression has a value (in the first version of software all the values are equal to 1) and will change a weight of the corresponding aspect of the action *D* by this value. The computer composes its turn tr_2^A choosing a sentence from the set of sentences for increasing/decreasing this weight and at the same time it increases/decreases this weight in the partner model by the value of the chosen sentence. A reasoning procedure based on the new model will yield a positive decision (i.e. the computer supposes that the user´s decision will be positive). Now the user must choose (or put in as free text) his response and the process can continue in a similar way.

A dialogue will be generated jointly by a user and the computer. The computer uses its communicative tactics. Let us suppose that the user - after the computer´s proposal to perform an action *D* - will create a model of herself, i.e. she will attribute values to all aspects of *D* and will do reasoning on the basis of this model. Still, creating this model is certainly inexplicit. In reality the user does not think that the usefulness of *D* is 5 and its harmfulness is 7 but she figures out that doing *D* would be more harmful than useful. In principle, this reasoning procedure may as well be considered creating a model of oneself. The task of the computer is, by implementing its communicative tactics, to try to influence the partner model this way that on the basis of the changed model the partner would make a positive decision. The only problem is that the computer does not ´´know´´ the real weights attributed to different aspects of *D* by the user. It can only make guesses from the user´s negative responses.

As said before, when starting a dialogue the computer randomly generates a user model. At the beginning, we have set only one restriction: we required that the initial model should satisfy the presumption(s) underlying the corresponding reasoning procedure. Thus, for enticing *w(pleasant) > w(unpleasant)*, for convincing *w(useful) > w(harmful)* and for threatening *w(obligatory) = 1*. But the experiments have shown that such an initial model has given relatively bad results. 65% of the dialogues were hopeless because after three pairs of turns the computer had reached such values in the user model that the continuation of the dialogue became meaningless: the weights of negative aspects had reached such a level compared with the positive aspects that it was hopeless to try to reach a model where the reasoning would yield a positive decision by the partner.

The situation improved considerably when we added another restriction to the initial model: we required that the chosen reasoning procedure should aim at getting a positive decision in this model. In real life this restriction is also meaningful: while making a proposal or request we suppose that our partner will agree and only when counter-arguments are put forward shall we try to refute them.

4.2 Updating the User Model

The following example demonstrates in more details how the partner model is used in interaction.

Let us suppose that the computer has chosen the tactics of enticing and has generated the following user model:

w^{AB} = *{w(are-resources)=1, w(pleasant)=9, w(unpleasant)=7, w(useful)=5, w(harmful)=0, w(obligatory)=1, w(prohibited)=0, w(punishment-do)=0, w(punishment-not)=1}*.

The reasoning procedure WISH in this model yields a positive decision since *w(are-resources)=1, w(pleasant) > w(unpleasant) + w(harmful), w(pleasant)+w(useful) > w(unpleasant) + w(harmful), w(prohibited)=0* (cf. Fig. 1). Let us suppose that the user chose a refusing sentence and indicated that *w(harm)* must be corrected. There are three possible negative outcomes when applying the procedure *wish* (Fig.1).

Let us suppose here that every sentence has the value 1. In this case: if *w(obligatory) = 1* we have *w(harmful) >= w(pleasant) - w(unpleasant) + w(useful) + w(punishment-not) = 8*. Thus, in the corrected model *w(harmful) = 8*. In this case the procedure *wish* will yield a negative decision as before but if we decreased the value of *w(harmful)* by 1 we should reach a positive decision soon.

If there is more than one possible non-empty domain of allowed values for correcting a weight we shall choose the domain with the greatest lower barrier (for negative aspects of *D*) or with the least upper barrier (for positive aspects), i.e. the worst case.

The following example (2) is an excerpt of an enticing dialogue where the goal of the agent is to reach the partner´s decision to travel to Venice (*A* - computer, *B* - user, ready-made sentences were used by both the computer and the user).

(2)

A: Would you like to travel to Venice? Our firm needs to conclude a contract there.

B: Why me?

A: You look very smart, this is important for making contracts.

B: Why do I suit better than Mary?

A: You have a talent for making such contracts.

/---/

B: When I am abroad my husband will be unfaithful.

A: Sorry, I could not convince you.

In the second version of the software, a database is used for identifying different key words and phrases in the user input (the input is checked against regular expressions). The database also includes an index of answer files and links to suitable answers, as well as files corresponding to different communicative tactics containing various arguments to present to the user.

The use of unrestricted natural language text as input is both an advantage and a disadvantage for the application as it helps in creating a more natural dialogue but at the same time, if the database is compiled poorly, it can turn the conversation unnatural in a few turns.

5 CONCLUSION

We are considering dialogues in natural language where one participant (initiator of interaction, *A*) has a communicative goal that the partner (*B*) will perform an action *D*. If *B* does not agree then in the following interaction, *A* tries to stress positive and down-grade negative aspects of *D* in order to direct *B*´s reasoning about performing *D* toward the positive decision. In the reasoning process, *B* is weighing different aspects of *D*. If the positive aspects weigh more than negative then the decision will be to do *D*. *A* can present different arguments for *D* in a systematic way, e.g. to stress time and again pleasantness of performing *D* (i.e. to entice *B*), to stress usefulness of *D* (i.e. to convince *B*), etc. *A* can also act passively, only averting the arguments presented by *B* and not stressing any positive aspects of performing *D*.

We have worked out a model of conversational agent which includes a reasoning model and implemented it as a computer program, which can be used for training argumentation skills. The user can interact with the computer in Estonian, playing the role of the participant *B*, either choosing ready-made sentences as counter-arguments against performing the action or putting in free texts. In the last case, cue words are used by the computer in order to analyse user sentences. So far, a limited number of voluntary testers have worked with the software. However, we believe that such software is useful when training the skills of finding arguments and counter-arguments for or against performing an action. The program can establish certain restrictions on argument types, on the order in the use of arguments and counter-arguments, etc (cf. [3]).

Our future work includes implementing a conversational agent that can also play *B*´s role and software evaluation in user studies.

ACKNOWLEDGEMENTS

This work is supported by the Estonian Science Foundation (grant 9124).

REFERENCES

[1] J. Allen, *Natural Language Understanding*. 2nd ed. The Benjamin/Cummings Publ. Comp., Inc. 1995.

[2] G. Boella. and L. van der Torre. ´BDI and BOID Argumentation´, *Proc. of CMNA-03* http://www.computing.dundee.ac.uk/staff/creed/-research/previous/cmna/finals/boella-final.pdf (2003).

[3] S. Bringsjord, J. Taylor, A. Shilliday, M. Clark, K. Arkoudas, and S. Khemlani.´Slate: ´An Argument-Centered Intelligent Assistant to Human Reasoners´, *Proc. of CMNA-08* http://www.cmna.info/-CMNA8/ (2008).

[4] D´Andrade, R. ´A Folk Model of the Mind´, *Cultural models of Language and though,* Eds. D. Holland and A. Quinn, London: Cambridge University Press, 112-148, (1987).

[5] M. Davies and T. Stone, *Folk psychology: the theory of mind debate.* Oxford, Cambridge, Massachusetts: Blackwell. 1995.

[6] J. Ginzburg and R. Fernández, ´Computational Models of Dialogue´, Clark, Fox, and Lappin (eds.) *The Handbook of Computational Linguistics and Natural Language Processing*, Wiley Blackwell Publishers, 429-481, (2010).

[7] K. Jokinen, *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. John Wiley & Sons Ltd, 2009.

[8] M. .Koit, ´Conversational Agent in Argumentation: Updating of Information States´, *Proc. of the International Conference on Knowledge Engineering and Ontology Development: KEOD 2011, Paris, France, 26-29 October, 2011. (Eds.) Joaquim Filipe and Jan L. G. Dietz.* Paris: SciTEC Publications Ltd, 2011, 375-378, (2011).

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

[10] H. Õim, ´Naïve Theories and Communictive Competence: Reasoning in Communication´, *Estonian in the Changing World,* University of Tartu, 211-231 (1996).

[11] D.Traum and S. Larsson. ´The Information State Approach to Dialogue Management´, *Current and New Directions in Discourse and Dialogue*, J. van Kuppevelt and R. Smith (Eds.), Kluwer, 325-353, (2003).

[12] T. Yuan, D. Moore, and Alec Grierson, ´Human-Computer Debate: a Computational Dialectics Approach´, *Proc. of CMNA-02.* http://www.csc.liv.ac.uk/~floriana/CMNA/YuanMooreGrierson.pdf (2002).

Questions, Arguments, and Natural Language Semantics

Adam Wyner[1]

Abstract. Computational models of argumentation can be understood to bridge between human and automated reasoning. Argumentation schemes represent stereotypical, defeasible reasoning patterns. Critical questions are associated with argumentation schemes and are said to attack arguments. The paper highlights several issues with the current understanding of critical questions in argumentation. It provides a formal semantics for questions, an approach to instantiated argumentation schemes, and shows how the semantics of questions clarifies the issues. In this approach, questions do not attack schemes, though answers to questions might.

1 Introduction

Computational models of argumentation can be understood to bridge between human and automated reasoning, for both represent, reason with, and evaluate valid arguments. Arguments can be proposed and *attacked* by counter-arguments; where an argument either is not attacked or is defended from attack, we may accept that argument. While abstract argumentation, e.g. [5], focusses on abstract arguments with no internal structure, other approaches provide formal analyses of the internal structure of arguments [15, 3], where the propositions of the arguments are drawn from a knowledge base that people might use. Argumentation schemes are even closer to human forms of reasoning [21], for they represent stereotypical, defeasible reasoning patterns about everyday activities or considerations. Arguments are also used dialogically in that two (or more) users may discuss a topic, presenting statements that instantiate argumentation schemes, and arguing for or against claims. As individuals have only partial, conflicting, or alternative information, people use arguments to fill in information, resolve conflicts, chose among alternatives, or at least provide an explicit, rational explanation of the precise nature of the dispute.

A central aspect of argumentation schemes are *critical questions* associated with it, which are said to identify how arguments can be attacked such as [23]:

Argument from Expert Opinion

- Premise: Source E is an expert in subject domain S containing proposition A.
- Premise: E asserts that proposition A is true (false).
- Claim: A is true (false).

- CQ2: Field Question. Is E an expert in the field that A is in?
- CQ4: Trustworthiness Question. Is E personally reliable as a source?

Answer *no* to any of these questions, the reasoning to the claim fails.

[1] Department of Computer Science, University of Liverpool, Liverpool, UK email: adam@wyner.info

As we discuss later, there are a variety of ways that critical questions are discussed in the literature. But, there is more at stake in an analysis of them than distinguishing these ways or choosing among them. First, in our view, critical questions as *questions* and as presented in the argumentation literature are not straightforwardly compatible with formalised approaches to instantiated argumentation, e.g. ASPIC+ [15] or Logic-based argumentation [4]. Yet, it only makes sense to provide an approach to questions that does straightforwardly suit instantiated argumentation. Second, the analysis of critical questions is not formalised, structured, or systematic in and of itself: What is the logical space of critical questions relative to a given scheme?; What are the relationships between the questions and the schemes; Why is it that a critical question of one scheme cannot serve to attack some other unrelated scheme? Simply giving a list of critical questions relative to a scheme does not explain them or enable productive use of them. Third, the analysis of critical questions is not related to a well-developed formal semantics of questions in natural language [10]. If argumentation is to be a medium of broadly applicable man-machine communication, then questions ought to be consonant with how questions are used by humans and analysed by linguists, not as something specially defined in argumentation. Moreover, we argue that there are advantages to making a theory of questions formally related to linguistic analysis, for it makes explicit information which has been otherwise left implicit.

In this paper, we begin to address these issues. The novelty of the paper is that it applies a well-developed, widely adopted formal analysis of the semantics of questions to the discussion of argumentation and critical questions, thereby establishing a *baseline* on the treatment of questions. Furthermore, our proposal distinguishes and modularises the roles of questions, answers, arguments, and dialogue which elsewhere appear to be conflated. However, this paper does not attempt a systematic reconstruction or reanalysis of prior proposals, which is beyond the scope of this paper; this is left for future work. By furthering the integration of argumentation with computational linguistics, we further the cause of argumentative human-machine communication.

In section 2, we set the discussion in the context of a typology of questions, selecting only those that are relevant for argumentation. We outline a formal semantic analysis of questions in section 3. A formalisation of instantiated argumentation is outlined in section 4. In section 5, we apply the theory of questions in the context of instantiated argumentation. Other approaches to argumentation schemes and critical questions are discussed in section 6 and compared to our proposal. Finally, we close with some future work and general observations in section 7.

2 Natural Language Analysis of Questions

In Linguistics and Computational Linguistics, the syntax, semantics, and pragmatics of questions have long been studied [10]. Syntax means here the grammatical analysis of the form of questions, semantics relates to the content to of the questions, and pragmatics to the question/answer speech acts in dialogue. Each of these subtopics itself is the object of extensive research. For our purposes, we focus on the semantics, presume the syntax, and leave dialogical aspects largely to future discussion. We first narrow the discussion, separating out from the spectrum of kinds of questions and their answers those that are most directly and immediately useful for argumentation.

There are a wide variety of questions, not all of which (yet) appear to be relevant to argumentation. To set the context, we briefly mention some key issues. One distinction is between unembedded questions (main clause) and embedded questions (subordinate clause), where the embedded questions appear after a variety of verbs, e.g.*indicate, know, believe, wonder*, and others.

- When will Jill arrive?
- Bill knows [when Jill will arrive].

We look only at the main clause questions, for while subordinate clause questions may appear in an argumentation scheme, they do not serve as critical questions about a scheme. See [13] for more on this important distinction.

One particular sort of questions are yes/no questions:

- Is your mother at home?

The answers to such questions can be taken as elliptical for the corresponding declarative that gives a full answer to the question; that is, answering *Yes* is an elliptical form for *My mother is at home*, while *No* is *My mother is not at home*. Yes/no questions are restrictive in the sense that they only represent a *one* literal and its negation.

There are a range of other sorts of questions and issues about them. For example, WH-questions contain a *wh-word*, e.g. *who, what, when, where, why, how*.

- What did John buy?

This can be answered with a *short answer*, e.g. *War and Peace*. Alternatively, this can be answered with a *propositional answer*, e.g. *John bought War and Peace*. Here we can take the short answer as elliptical of the propositional answer. There are a range of additional issues about the syntax and semantics of questions, but yes/no questions serve as good starting point into a formal analysis of questions in argumentation.

3 A Formal Semantic Theory of Questions

In this section, we briefly outline a well-developed, widely adopted formal semantic analysis of questions [10]. A range of interpretations of questions are reviewed, particularly the success and satisfaction conditions of the illocutionary act of asking a question. It is argued that dialogical interpretations of interrogatives *presuppose that questions have a distinct type of semantic object*. Thus, the key task is to define this semantic object and to know what it means to answer a question. A simple, yet explanatory analysis is provided, which then helps us better understand the role of questions in argumentation.

A core claim is that there is a semantics of questions related to a semantics of propositions, and that we can provide a *static* analysis, which then can provide the basis of a *dynamic* (e.g. dialogic) analysis, where other issues arise such as processing a question, selecting an optimal answer, shifting roles of the discussants, and so on. As with a semantics of indicatives, the two most important criteria of adequacy for a theory of questions are that it specify a notion of equivalence between two questions (*semantical identity*) and of entailment (*meaning inclusion*). In the following, we provide the background intuitions to such a theory, followed by a formalisation in intensional propositional logic.

The analysis is based on leading intuitions from [11], called *Hamblin's picture*:

i An answer to a question is a sentence.
ii The possible answers to a question form an exhaustive set of mutually exclusive possibilities.
iii To know the meaning of a question is to know what counts as an answer to that question.

Postulate [i] focuses on *propositional* meanings for answers, where sentences are represented as propositions in a logic. Postulate [ii] means that the set that is the union of the answers exhaustively and completely fills the logical space of the question so that no possible answers are left out. The logical space is the space of possibilities that the world could be like. Consequently, one answer to a question *excludes* other answers. Where we leave aside the issue of presuppositions that we assume are fulfilled, the possible answers to a question *partition* the logical space; consequently, the answer to a question in a context is the unique proposition that is true in the context from amongst the possible answers. Postulate [iii] identifies the meaning of the question with the partition itself, so while questions are related to propositions, they are not reduced to them. Where no partitions are possible, e.g. propositions that are true at every world in the model, we suppose a Gricean explanation for the absence of the question since its answer is uninformative.

The formal semantic theory for Hamblin's picture is set within *Propositional Intensional Logic*. A fundamental notion is *possible world*, which is a notion of *an alternative way that things could be* [19]. In an *extensional theory*, a model M specifies the denotations of the terms, relations, and complex expressions; it can be understood as a singular specification of a world. In intensional logic, we have several such worlds; a model M in intensional logic is a set of possible worlds. The meaning of an indicative ϕ is the extension relative to a model M and a world w, $[\phi]_{M,w}$, which is a function from M, w, and ϕ to the truth value (indicated with 0 or 1) assigned in w to ϕ. The intension of ϕ in the model M, called a proposition, is the *set of worlds in the model where ϕ is true*: $[\phi]_M$ = {w in M $\|$ $[\phi]_{M,w}$ = 1}. In a complete model, every proposition is exclusively either true or false in every world; the intersection of $[\phi]_M$ and $[\neg\phi]_M$ is empty, while their union is the set of worlds in the model. Logical equivalence and entailment can be defined set-theoretically relative to a model.

The intensional interpretation of interrogatives is constructed from the intensional interpretation of propositions. Syntactically, an interrogative is indicated with a question mark prefixed to the proposition - $?\phi$. We have the meaning of an interrogative in a world, which is equivalent to the propositional answer in that world:

Def (Interpretation at M,w). $[?\phi]_{M,w} = \{w' \in M | [\phi]_{M,w'} = [\phi]_{M,w}\}$

The extension of a yes/no-interrogative in a world is the meaning of the proposition in that world, providing a complete and precise answer to the question posed by the interrogative. The intension of $?\phi$ in a model M is the set of its extensions in M:

Def (Interpretation in M). $[?\phi]_M = \{[\phi]_{M,w} | w \in M\}$

Since the subsets of meaning of $?\phi$ do not intersect, the question has *partitions*.

A proposition is an *answer* to a question where the meaning of the proposition (a set of worlds) is a subset of the meaning of the question:

Def (Answerhood). $\phi| =?\psi$ iff $\forall M \exists w : [\phi]_M \subseteq [?\psi]_{M,w}$

For our purposes, saying *Yes* in answer to a yes/no question is to accept that the indicative form of the question is true, while *No* is to accept the indicative is false. For example, the question *Did Bill leave?* and answered *Yes* means, in the context where the question is answered, that *Bill left* is true, while *No* means *Bill left* is false. The question abstracts over these contexts, thus, *Did Bill leave?* denotes the partition of propositions {*Bill left*, *Bill did not leave*}.

This analysis corresponds to Hamblin's picture. An answer to a question is a proposition (derived from a sentence); for yes/no interrogatives, the answers to the question are the propositions that are mutually exclusive and that exhaust the logical space consisting of all possible worlds in M; the meaning of the question is just the partition of answers. The theory is formally adequate as logical equivalence and entailment can be defined in set-theoretically, much like the indicatives. This theory is an initial, formal basis for the analysis of questions in argumentation.

4 Argumentation

In this section, we discuss instantiated argumentation with respect to a knowledge base (of literals and rules). For our purposes, we work with the *logic-based* approach of [3], which represents arguments in terms of classical propositional logic. We review how arguments are constructed and how attacks between them are identified in a logic-based approach. In particular, in instantiated argumentation, we have positive and negative literals. Where the literals are semantically interpreted as in intensional logic, we can use them to form questions as outlined above. Argumentation schemes can be represented in an instantiated argumentation theory.

In a logic-based approach, statements are expressed as atoms (lower case roman letters), while formulae (greek letters) are constructed using the logical connectives of conjunction, disjunction, negation, and implication. The classical consequence relation is denoted by $\vdash$. Given a knowledge base Δ comprised of formulae and a formula α, $\Delta \vdash \alpha$ denotes that Δ entails α. Δ can be inconsistent, containing contradictory propositions. We assume a set of formulae Δ from which arguments are constructed. Where $\perp$ denotes inconsistency, $\Delta \vdash \perp$ denotes that Δ is inconsistent. An argument is an ordered pair $< \phi, \alpha >$, where $\phi \subseteq \Delta$, ϕ is a minimal set of formulae such that $\phi \vdash \alpha$, and $\phi \not\vdash \perp$. ϕ is said to support the claim α. For example, where p and q are atoms, and where the knowledge base is comprised of p and $p \rightarrow q$, then $< \{p, p \rightarrow q\}, q >$ is an argument, where $p, p \rightarrow q$ is the support for the claim q. For our purposes, argumentation schemes *are* arguments in a logic-based approach, presuming that the rule from which we draw the claim is implicit in the argumentation schemes, but explicit in the argument. Arguments in this approach are *defeasible* because it is possible for one argument to attack the support or the claim of another argument.

With contradictory propositions from Δ, we can construct arguments in *attack* relations, where the propositional claim of an argument is contradictory to the propositional claim of another argument or is contradictory to some proposition in the support of another argument. These are the *attack* relations between arguments $< \Psi, \beta >$ and $< \Phi, \alpha >$ such as *undercutter* and *rebuttal*; attacking arguments are referred to as *counterarguments*. $< \Psi, \beta >$ is an undercutter for $< \Phi, \alpha >$ where β is $\neg(\phi_1 \wedge \ldots \wedge \phi_n)$ and $\{\phi_1 \ldots \phi_n\} \subseteq \Phi$; in essence, the claim of one argument is the negation of a set of formulae in the support of another argument. $< \Psi, \beta >$ is a rebuttal for $< \Phi, \alpha >$ if and only if $\beta \leftrightarrow \neg\alpha$ is a tautology; the claims of the arguments are inconsistent. For example, supposing the following in a knowledge base (from [4]): p, $p \rightarrow \neg q$, r, $r \rightarrow \neg p$, $\neg p \rightarrow q$. From this, we can construct an argument to support the claim $\neg q$: $< \{p, p \rightarrow \neg q\}, \neg q >$. With respect to this argument, we have an undercutter $< \{r, r \rightarrow \neg p\}, \neg p >$ and a rebuttal $< \{r, r \rightarrow \neg p, \neg p \rightarrow q\}, \neg p >$.

In this theory, a yes/no question is expressed as $?\Phi$, which denotes the partition that is the denotation of Φ and the denotation of $\neg\Phi$. In a Logic-based approach, as with ASPIC+ [15], there is one knowledge base which is used to instantiate the argumentation schemes. Dialogical models, where there are different participants, may be defined as subsets of this knowledge base, and because of this the analysis of questions in a dialogical setting is defined with respect to the union of each participants' knowledge base. In the next section, we give several points that hold of our semantic theory of questions with instantiated arguments.

5 Questions and Argumentation Schemes

To this point, we have reviewed a semantic theory of questions in section 3 and instantiated arguments with attacks in section 4. In this section, we apply our theory of questions to the approach to instantiate arguments.

Models in intensional propositional logic may be used to represent inconsistent knowledge bases, as contradictory propositions denote distinct sets of possible worlds. This approach to questions appears to be all that is required by logic-based argumentation, for questions denote the partition of contradictory propositions. This is a very straightforward result. Following [10], it leaves aside issues bearing on the *dialogical issues* of questions in the context of argumentation.

Several key points hold of this analysis of questions and instantiated arguments.

1. Questions denote partitions of answers, where answers are propositions. Such partitions reflect conflicts of information in the model; questions arise where ever such conflicts exist.
2. Questions are answered with respect to a world, and the answer indicates what holds in that world.
3. Answers, as propositions, may be *justified* just as with any other proposition. In this theory, questions *cannot* be justified.
4. Questions reflect the model in that there can only be yes/no questions if there are contradictory propositions in the model.
5. Only propositions can introduce attack relations between arguments since attack is defined in terms of contradictory propositions, and only propositions can be negated. Questions do not bear truth values and cannot themselves be negated; thus, it is a category mistake to say that a *question attacks an argument*. However, the *answer* to a question may give rise to an attack.
6. To ask a question with respect to an argumentation scheme implies that *the model can represent the meaning of the question (i.e. the propositions)*. Moreover, it implies that *the argumentation scheme itself represents the relevant proposition (either in a positive or negative form)*. This follows from the meaning of a question, instantiated arguments, and attack; if this were not the case, the question would be *irrelevant* with respect to the argumentation scheme.

7. The number and kind of questions is entirely dependent on the number and kind of propositions that (possibly implicitly) specify the scheme.
8. Given a model and an argumentation scheme with all premises explicit, yes/no questions could generated, so would be formally redundant.

In this approach, the knowledge base represents domain knowledge, lexical semantic information, and so on. With respect to the knowledge base, argumentation schemes are instantiated. As the knowledge base is inconsistent, questions can be generated. The analysis is abstract, as the possible worlds analysis provides a static view on all alternatives. It makes no claims about changes of the knowledge base can be changed, growth of knowledge, extensions of argumentation schemes, or dialogical issues. The approach also makes no claims about the necessary or sufficient conditions for an argumentation scheme; rather, if it is felicitous to ask a question with respect to an argumentation scheme, then one of the answers to the question is a premise of the scheme. However, the approach outlined above is proposed as a basis for such dynamics, following a similar trajectory dialogue [7] and dynamic semantics [12].

In the next section, we mention previous approaches to critical questions

6 Comparison

We are not aware of previous research that relates a formal theory of instantiated argumentation with a formal theory of questions that is based on a formal linguistic analysis. However, there has been a body of work that discusses critical questions, which we may take as representative, e.g. [22], [8], [2], and [18].

First, it is important to reiterate a point made in section 3, where it is claimed that there is a semantics of questions that is *presumed* by dialogical/discourse approaches to questions [10]. There are dialogical approaches to argumentation [17], [14], [9], and [1]), among others. And questions are discussed in these contexts. However, it is our view that modularising the analysis, e.g. separating out questions from their dialogical function, such as is done in the formal semantics of questions, helps to isolate and clarify the overall analysis. The dialogical analysis should be seen to overlay or apply to the semantics of questions. The same move is made in the analysis of the semantics of sentences in static and dynamic modes. In the literature that we have reviewed, the semantics of questions seems often to be conflated with their dialogical role.

In [22], several approaches to critical questions are reviewed - [20] and [9]. The proposal is made that critical questions can be understood as implicit premises of an argument. As we have discussed in section 5, in formally representing the knowledge of argumentation schemes, we make all information as explicit and overt as possible. This applies as well to the various subtypes of questions proposed in [20]. We have also discussed that argumentation schemes only have propositions in premises and claim, which precludes questions: an *answer to a question (or its negation) may be a premise of an argumentation scheme, but not the question itself*. Yet, as we discuss in the conclusion, there are interesting topics about questions in arguments. In addition, there is a discussion about how critical questions are tied to shifts in burdens of proof and to proof standards, which we have not discussed in this paper, but see [16]; whether all argumentation schemes are associated with burdens of proof and proof standards is an open question in our view. Dialogical aspects are discussed, e.g. sorting the premises according to their role in dialogical investigation of the acceptability of the argument.

In [8], we find an overview of philosophically oriented analyses of argumentation schemes and critical questions. They consider the role of critical questions in the evaluation of schematic arguments, the correct number and kind of critical questions accompanying a scheme, and burdens of proof and proof standards. We have addressed some of these issues in 5. We agree that questions can be *used* to test three aspects of argument cogency: relevance, acceptability and sufficiency. However, it is the *answer*, not the question, that plays the crucial role. Moreover, just how these aspects are to be defined remains an open issue. A general topic is raised about whether argumentation schemes are intrinsically *open textured* in the sense that we cannot define the necessary and sufficient conditions for them. This is a general problem for the representation of human knowledge and arises in analysing language, vision, and other higher cognitive functions. For our purposes, we can take schemes as *prototypes* subject to *refinement*. Our proposal about questions makes no claim on these matters.

A different approach to critical questions is taken in [2] and [18], concerning the *Practical Reasoning with Values* argumentation scheme of the form:

> In the current circumstances R, we should perform action A, which will result in new circumstances S, which will realise goal G, which will promote value V.

A semantic model is provided with a domain of actions, agents, states, and values as well as relations and constraints. The scheme is an abstraction with respect to the model, where the variables can be instantiated. We do not have the space here present the formal analysis, but sketch the treatment of critical questions.

The core idea is that in posing a critical question, an opponent attacks an the element of the instantiated scheme. The scheme has 16 associate critical questions, among them:

- CQ1: Are the believed circumstances true?
- CQ5: Are there alternative ways of realising the same consequences?

Answering *no* to the first or *yes* to the second is said to attack a presumption of the scheme such that the presumed claim does not follow. However, the presumptions are not represented in the scheme itself, but are incorporated into the meaning and function of the answer to the question. That is, if we answer *yes* to CQ5, this means that relative to the way of realising the consequences given by the instantiation of the scheme, and relative to what is available in the semantic model, there are alternative ways of realising the same consequences. Moreover, having such alternatives implies that we cannot presume the proposed action should be done. In [18], we have formalisations of all 16 critical questions, where each is presented as an argument instantiation that attacks the target scheme.

This approach is not consistent with our proposal concerning the relationship between questions and argumentation schemes. Questions are represented as arguments, for which there is no justification or evidence. The attack on the instantiated scheme is "directly defined", but not with respect to Logic-based argumentation or ASPIC+, since there is no component of the instantiated scheme that is attacked. Furthermore, it allows that that an arbitrary question could be defined so as an attack on the argument. An alternative approach would be to take the semantic information encoded in a critical question and make it specifically part of the argumentation scheme as a premise. Then the answer to the question serves as an attack on the scheme, consistent with the semantics of questions and Logic-based argumentation and ASPIC+. Arbitrary attacks cannot be defined in this approach since there can only be attacks on premises that are part of the presumptive reasoning of the scheme. The dialogical aspect could still be overlain the questions.

7 Conclusions

The paper discusses the role and representation of questions with respect to argumentation schemes. In contrast to research in argumentation *per se*, the formal semantics of questions does not treat questions as *attacks*, but as partitions of answers. It is the answers, not the questions, from which we derive argument attack. The semantics of questions is compatible with current approaches to instantiated argumentation. The analysis clarifies the role of questions in identifying auxiliary premises of schemes, which would be best made explicit. It also separates the semantics of questions from their dialogical role. We compared our analysis against extant analyses, showing how questions, attacks, premises, and dialogue are conflated.

There are many issues that remain to be investigated. First, the existing critical questions ought to be converted into explicit premises, leaving aside the issue of implicit representation. The formal semantics of questions ought to be integrated into a dialogical system. It would be worth investigating the nature of the questions that can be asked about schemes, what type and range they may have. Finally, we should consider *Erotetic Logic*, where questions can be the premises of rules, in the context of argumentation since they challenge fundamental assumptions both of the semantics of questions and of argumentation [6]. As part of the investigation, we should see how such questions are related to those from which attacks are derived.

ACKNOWLEDGEMENTS

The author was supported by the FP7-ICT-2009-4 Programme, IMPACT Project, Grant Agreement Number 247228. The views expressed are those of the author.

REFERENCES

[1] Katie Atkinson, Trevor Bench-Capon, and Peter Mcburney, 'A dialogue game protocol for multi-agent argument over proposals for action', *Autonomous Agents and Multi-Agent Systems*, **11**(2), 153–171, (2005).

[2] Katie Atkinson and Trevor J. M. Bench-Capon, 'Practical reasoning as presumptive argumentation using action based alternating transition systems', *Artificial Intelligence*, **171**(10-15), 855–874, (2007).

[3] Philippe Besnard and Anthony Hunter, *Elements of Argumentation*, MIT Press, 2008.

[4] Phillip Besnard and Anthony Hunter, 'Argumentation based on classical logic', in *Argumentation in Artificial Intelligence*, eds., Iyad Rahwan and Guillermo Simari, 133–152, Springer, (2009).

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

[6] Jonathan Ginzburg, 'Questions: Logic and interaction', in *Handbook of Logic and Language*, eds., Johan van Benthem and Alice ter Meulen, Elsevier, second edition edn., (2010).

[7] Jonathan Ginzburg and Ivan Sag, *Interrogative Investigations: the form, meaning, and use of English Interrogatives*, CSLI, 2000.

[8] David Godden and Douglas Walton, 'Advances in the theory of argumentation schemes and critical questions', *Informal Logic*, **27**(3), 267–292, (2007).

[9] Thomas Gordon, Henry Prakken, and Douglas Walton, 'The Carneades model of argument and burden of proof', *Artificial Intelligence*, **171**, 875–896, (2007).

[10] Jeroen Groenendijk and Martin Stokhof, 'Questions', in *Handbook of Logic and Language*, eds., Johan van Benthem and Alice ter Meulen, 1059–1124, Elsevier, 2nd edn., (2010).

[11] Charles Hamblin, 'Questions', *Australasian Journal of Philosophy*, **36**, 159168, (1958).

[12] Hans Kamp and Uwe Reyle, *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language: Formal Logic and Discourse Representation Theory*, Springer, 1993.

[13] Lauri Karttunen, 'Syntax and semantics of questions', *Linguistics and Philosophy*, **1**(1), 3–44, (1977).

[14] Henry Prakken, 'Formal systems for persuasion dialogue', *Knowledge Engineering Review*, **21**(2), 163–188, (2006).

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

[16] Henry Prakken, 'On the nature of argument schemes', in *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument*, eds., Chris Reed and Christopher Tindale, 167–185, College Publications, London, (2010).

[17] Henry Prakken and Giovanni Sartor, 'A dialectical model of assessing conflicting arguments in legal reasoning', *Artificial Intelligence and Law*, **4**(3-4), 331–368, (1996).

[18] Luke Riley, Katie Atkinson, Terry Payne, and Elizabeth Black, 'An implemented dialogue system for inquiry and persuasion', in *Theory and Applications of Formal Argumentation - First International Workshop, (TAFA 2011)*, Lecture Notes in Computer Science (LNCS), pp. 67–84, Berlin, Germany, (2011). Springer.

[19] Robert Stalnaker, *Ways a World Might Be*, Oxford University Press, 2003.

[20] Bart Verheij, *Virtual Arguments: On the Design of Argument Assistants for Lawyers and Other Arguers*, T.M.C. Asser Press, 2004.

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

[22] Douglas Walton and Thomas Gordon, 'Critical questions in computational models of legal argument', in *Argumentation in Artificial Intelligence and Law*, eds., Paul Dunne and Trevor Bench-Capon, IAAIL Workshop Series, p. 103111, Nijmegen, The Netherlands, (2005). Wolf Legal Publishers.

[23] Douglas Walton, Chris Reed, and Fabrizio Macagno, *Argumentation Schemes*, Cambridge University Press, 2008.

Towards Bridging Between Natural Language and Logic-Based Representations of Natural Arguments

Helmut Horacek[1]

Abstract. Representations of natural language arguments and logic-based representations of arguments, such as argumentation frameworks, are two backbones of the field of computational models of natural arguments. Still, the task of relating these two forms of representation in sufficient degrees of formal rigour is a widely untouched activity in the field and known to be a great challenge. First results developed in the context of controled online discussion forums look quite promising, but the mapping between natural language and logic-based representations is a bit coarse-grained yet. Consequently, the potential benefits of using argumentation frameworks and logical reasoning can by far not be exploited. In this paper, we propose an interactive process in which the user and the system cooperatively attempt to make the logical representation of a new user argument in a natural debate more accurate – through some recasting operations, the originally proposed representation is made more explicit and uniform in the embedding context. The attainment of such precise representations enables an increased functionality on the side of the logical components, such as checking the validity of arguments, updating their status, and proposing anchoring points for further arguments, for instance by referring to argumentation schemes.[1]

1 INTRODUCTION

For a long time, two major branches in the field of computational models of natural argument, natural language methods and logic-based reasoning techniques, lived side by side without serious interaction. To some extent, this is hardly surprising, since the demand on natural language processing techniques is considerably high for argumentation purposes, and many of the techniques and results of logic-based methods in the field were enabled by abstractions made in the underlying representations, thus making the gap to natural language-based representations even wider. In the long run, however, a tighter connection between the two areas appears to be urgent and desirable in order to develop techniques and build systems that exhibit both, natural language communication and logical reasoning capabilities to a certain extent, thus enabling increased overall functionalities.

In a number of papers, Wyner and his co-authors [7, 8] have shown how state-of-the-art natural language processing methods can be applied to build abstracted representations to be used by argumentation frameworks [2] under some simplifications – the restriction to controled English, and user cooperation to specify the role and scope of newly introduced arguments. The overall system enables the argumentation framework to compute extensions in terms of sets of consistent arguments, which correspond to possible viewpoints adopted by participants in the underlying debate. However, it is unclear how this enviornment could enable more advanced functions of the logical system, such as checking validity of arguments and finding anchor points for new potential arguments. In our view, further progress is mainly inhibited by some lack of uniformity and explicitness in the logical representations. In order to increase the rigour in mapping from natural language to logical representations, we propose an interactive process in which the user and the system cooperatively attempt to make the logical representation of a new user argument in a natural debate more accurate – through some recasting operations, the originally proposed representation is made more explicit and uniform in the embedding contexts.

This paper is organized as follows. We first illustrate some of the logical representation deficits in Wyner's running example. Then we outline a method that aims at building more accurate logical representations through interaction between the system and a user who introduces a new argument in a debate. We illustrate this idea by a walk-through of Wyner's running example. Finally, we discuss the state of affairs and future prospects.

2 SOME DEFICITS IN BUILDING ARGUMENTATION FRAMEWORKS OUT OF NL

In this section, we examine Wyner's running example from the perspective of how adequately the assertions to be ultimately incorporated into an argumentation framework are categorized and anchored in the incrementally constructed argument graph (see Figure 1 for the list of assertions, and Figure 3 for the argument graph built out of them). When a user raises a new argument, he also specifies the argument to which the new one is related and the category of that relation. Since humans generally tend to be sloppy in their formulations, express pieces of information in limited degrees of explicitness, especially in inference-rich discourse, and may find it hard to precisely identify semantic relations in a given context, we can expect a number of problems associated with user specifications of this kind.

By analysing the running example, we aim at pointing to a variety of formal inaccuracies, such as ontological discrepancies and duplications, uncover relations between assertions, identify misplaced relations and implicit information. It is utterly important to avoid such representational deficits to increase the formal rigour of the logical representation built. Achieving a high level of formal accuracy is an indispensible prerequisite for running logical reasoning capabilities over portions of the argument graph with reasonable success.

[1] Research Center for Artificial Intelligence (DFKI), Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany. Email: helmut.horacek@dfki.de

1. Every householder should pay tax for the garbage which the householder throws away.
2. No householder should pay tax for the garbage which the householder throws away.
3. Paying tax for garbage increases recycling.
4. Recycling more is good.
5. Paying tax for garbage is unfair.
6. Every householder should be charged equally.
7. Every householder who takes benefits does not recycle.
8. Every householder who does not take benefits pays for every householder who does take benefits.
9. Professor Resicke says that recycling reduces the need for new garbage dumps.
10. A reduction of the need for new garbage dumps is good.
11. Professor Resicke is not objective.
12. Professor Resicke owns a recycling company.
13. A person who owns a recycling company earns money from recycling.
14. Supermarkets create garbage.
15. Supermarkets should pay tax.
16. Supermarkets pass the taxes for the garbage to the consumer.

Figure 1. The set of arguments In Wyner's running example

In our analysis, we adopt a terminology inspired by logic/deductive systems, which is entirely consistent with argumentative frameworks, but does not always suit well linguistic categories. An argument that attacks (resp. supports) another argument (q) consists of a *fact* (p) and a *rule* ($p \rightarrow \neg q$ resp. $p \rightarrow q$). The fact component is also termed *proper argument*, because raising it frequently implies the rule component, either through the conversational context or through background knowledge. The *rule* component is also termed *justification*. Hence, *fact* and *justification* correspond to *data* and *warrant* in Toulmin's terminology [4]. We have identified the following categories of deficits:

- *Assertions may bear varying roles*
  Some of the assertions in Figure 1 represent facts, some others represent rule-like pieces of information; for example, the assertion labeled (3) is a rule and the assertion labeled (4) is a fact, which is almost identical to the conclusion of (3). Rules and facts representing their conclusion are typically mentioned as alternatives in raising an argument, in accordance with the "modus brevis" presentations in explanations [1, 3]. Distinguishing among them is useful for reasons of uniformity, and for making relations between several arguments explicit (in particular, redundancy; arguments expressed differently may be identical from a logical perspective).
- *Assertions may elaborate/exemplify others*
  Some of the assertions may be related to others by adding semantic details, which does not play a proper argumentative role; for example, the assertion labeled (6) describes a state which elaborates on what fairness/unfairness in this context is supposed to mean, referring to the argument labeled (5). Treating such assertions not as arguments per se is essential.
- *Assertions suggest reference to argumentation knowledge*
  Some of the assertions may be related to implicit reasons, which are known from argumentation methods; for example, the assertion labeled (9) suggests that its justification is grounded in an argumentation scheme [5, 6], appeal to expert opinion. Recognizing such a situation may put a system in a position to consult relevant background knowledge, thus enabling it to check details of user specifications and anticipating possible follow-up arguments in the ongoing debate.
- *Assertions miss implicit contextual properties or restrictions*
  Some of the assertions may be formulated in a too strong manner, where the intended context or restrictions are left implicit; for example, the assertion labeled (11) is unlikely to mean that Prof. Resicke is not objective in general, but only in the specific context and circumstances where his opinion is referred to as an argument. It is fairly obvious that inaccuracies of this kind are quite likely to inhibit correct logical inferences over portions of the argument graph.

This analysis is not intended to be an exhaustive list; it merely aims at illustrating some of the discrepancies between the normal interpretation of assertions in the given context and the plausible intentions underlying these assertions. In order to enable reasoning about portions of the argument graph, such as changes in the state of arguments and the resulting impacts, and proposing promising attacking points for further arguments, it is overly important to avoid discepancies of this kind, such as redundancies, misplacements and inaccurate expressions of arguments. Consequently, the envisioned reasoning demand by the system motivates some dedicated recasting operations prior to incorporating abstracted representations built out of these assertions in a to-be-built logical representation.

3 OBTAINING MORE ACCURATE LOGICAL COUNTERPARTS TO NL

The aim is to find a more adequate structure of arguments than the one that was collectively built by the users, that is, the contributors to the argument graph for a case at hand. The crucial question is, how this can be achieved, and who the major agents involved in this enterprise should be. We think that neither the users nor the system alone can come very far. It can be assumed that the users did their best in formulating their arguments and anchoring them in the argument graph at its current state. Conversely, expecting a system to automatically analyze a new argument in the context of a given argument graph and to find a better representation would be extremely overoptimistic, given the limited capabilities of natural language interpretation techniques, in particular discourse parsers, despite respectable recent progress obtained. Hence, we pursue the following idea: since the discourse and argumentation reasoning capabilities of a system, though certainly limited, can be expected to be in some sense complementary to those of the user, a system produces a small list of suggestions for potential argument representation improvements, which the user is supposed to verify and adopt, if he gets convinced. Suggestions come as questions (see Figure 2).

1 . "Is the current argument equivalent to <argument>?"
(typically a sister argument to the proposed claim)
2 . "Is the current argument a direct support for <argument>?"
(its ancestor, one level up in the argumentative chain)
3 . "Is the current argument an elaboration of <argument>?"
(typically the argument it is supposed to support)
4 . "Is the current argument to be contextualized?"
(on the basis of argumentation knowledge)
5 . "Does the current argument fit into <argumentation scheme>?"
(on the basis of argumentation knowledge)

Figure 2. Categories of questions to induce an argument reorganisation

In order for a system to produce meaningful instantiations of these questions, the current argument needs to be analyzed and compared to some of the arguments in the present state of the argument graph. As a first step, the form of the argument is made more explicit. The user typically raises an argument (p and $p \rightarrow \neg q$ resp. $p \rightarrow q$) by stating either the fact (p) or the rule ($p \rightarrow \neg q$ resp. $p \rightarrow q$) perspective only, and the system needs to make both forms explicit; it may prove useful to paraphrase this explicit variant to the user, asking for confirmation.

Once this is done, several comparisons to related arguments in the argument graph are made. One of these comparisons aims at finding out, whether the current argument proposed is similar or even identical to another argument already raised. For this comparison, the content of the current argument is compared to that of the others. If a sufficient degree of similarity is diagnosed, for example by consulting an entailment method, the respective argument becomes a candidate for instantiating the meta-argument in the question labeled (1) in Figure 2. Both perspectives of the argument (fact and rule) are subject to this comparison. Arguments to be compared are preferably close to those in the topological vicinity of the place where the user has proposed the current argument for attachment, prominently "sister" arguments, that is, arguments already attacking resp. supporting the claim related to the current argument. Especially comparisons across perspectives of arguments may be helpful to uncover similarities or even duplications overlooked otherwise.

A further comparison concerns the position of the current argument, as proposed by the user. It may sometimes be the case that the support resp. attack relation can more accurately be assigned to an argument that resides higher in the argumentative chain where the proposed position is a leaf node. Hence, there are a few distinct positions to consider only: if the current argument is proposed as a support or attack for some argument q, and argument q, in turn, supports argument r, then the current argument p logically supports argument r, and it is worth testing whether a direct support from p to r is ontologically preferable. Similarly, if argument q attacks argument r, and r attacks argument s, the relation proposed between p and q may potentially be placed between p and s as well. Whether the direct relation is at least worth to be considered as an alternative to the variant proposed by the user can be checked by consulting a discourse parser. If that system considers the relation proposed by the user and the alternative one being of competitive plausibility, it seems to be reasonable to confront the user with this alternative. Question labeled (2) in Figure 2 exemplifies this constellation.

Another comparison stresses the ontological role of the relation proposed by the user. The scope of the arguments to be worth for being considered is the same as that defined in the previous paragraph, including also the argument originally proposed by the user as destination of a support relation. However, the ontological relation to be tested, again by a discourse parser, is different: in case the discourse parser prefers the interpretation of an elaboration relation to the proposed support relation, this gives rise to an instance of the question labeled (3) in Figure 2.

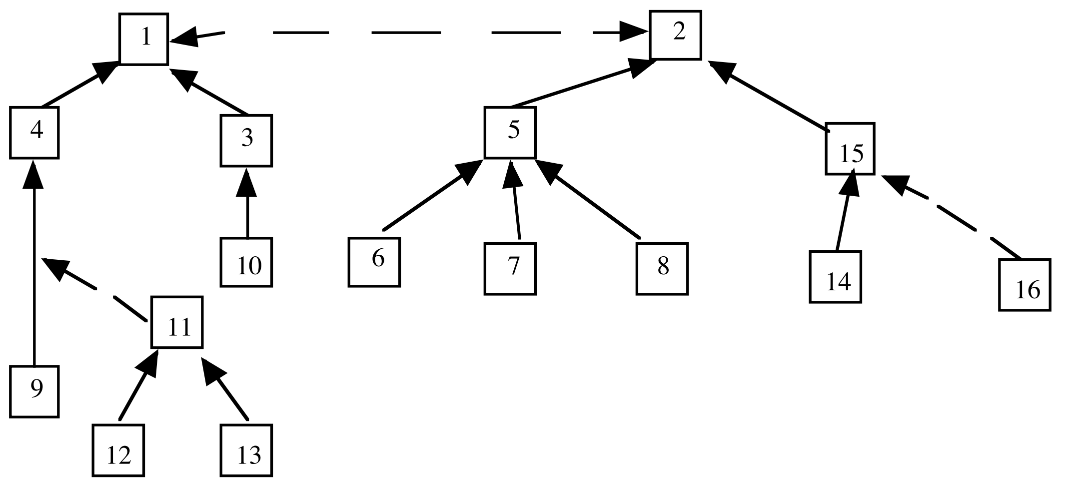


Figure 3: The original argument graph by Wyner

In the comparisons made so far, logical concerns and discourse knowledge play central roles in determining the plausibility of alternative interpretations. In the remaining questions labeled (4) and (5) in Figure 2, issues in argumentation constitute the essential background knowledge, notably argumentation schemes [5, 6]. In case there are linguistic indications for the potential relevance of a specific argumentation scheme, the user should be asked whether this is indeed the force behind the argument proposed (see question labeled (5) in Figure 2). If this is confirmed, additional knowledge can be invoked, so that the content of the arguments related, in particular the current one raised by the user, can be checked against descriptions associated with the argumentation scheme recognized as being relevant. Such a check may, for instance, give rise to follow-up questions such as the one labeled (4) in Figure 2.

These comparisons may give rise to asking several questions, but it is not advisable to present the complete set of questions to the user in each case, if there are many of them. A reasonable strategy seems to be to ask at most three questions in sequence, skipping the remaining part in this sequence if the user answers a question positively; this may only be followed by a follow-up question associated with the positively answered question. This way, there seems to be a fair chance that the user is not bored by the extended interaction and that he develops some trust and confidence in the system, if the questions occasionally lead the user to improve the specifications he originally has provided.

Altogether, we are still in the stage of exploration in the elaboration of our method. This concerns the selection among candidate tools (entailment reasoners, discourse parsers, etc.), the concrete details how to apply them, as well as how to interpret the results and orchestrate the user interaction on that basis. Nevertheless, we feel that the general idea has a significant potential, which we illustrate by the discussion of the example set of arguments in the following section.

4 WALKING THROUGH AN EXAMPLE

In this section, we illustrate the envisioned effects of our method, exemplified by Wyner's running example. We sketch the incremental building of a new argument graph, taking into account optimal effects of the questions generated to induce argument reorganization: we assume the user to make the best change induced by a question, abstracting from the proper dialog. The original argument graph relating the assertions in Figure 2 is given in Figure 3. The first two assertions, the point of the debate, remain unchanged; impacts on the other assertions are:

3. *Paying tax for garbage increases recycling.*
   This assertion is categorized as a rule. Its conclusion is paraphrased as "increased recycling", assessed as a positive result. The rule is considered adequate in its granularity/explicitness, hence no changes in representation are proposed.
4. *Recycling more is good.*
   This assertion is categorized as a fact, and it is considered equivalent to the fact perspective of argument 3 above, that is, its conclusion. Through recasting assertion 3, arguments 3 and 4 are essentially unified in their extended forms.
5. *Paying tax for garbage is unfair.*
   This assertion is taken unchanged as a fact perspective, the associated rule perspective "no tax, because of unfairness" is considered adequately in terms of granularity/explicitness.
6. *Every householder should be charged equally.*
   This assertion is better considered as an elaboration to assertion 5, not a separate argument; it should therefore be interpreted as an enhancement to the description of that assertion.
7. *Every householder who takes benefits does not recycle.*
   This assertion is considered as a fact that extends on the previous argument of unfairness. The granularity of the underlying rule, "it is unfair, because some householders do not recycle" is unsatisfactory. Due to lack of detailed knowledge, that rule stands for the moment.

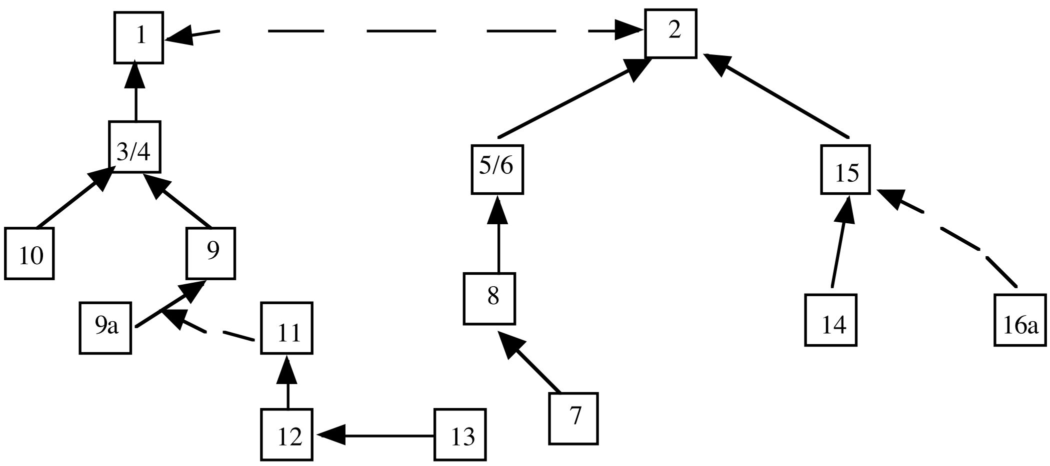


Figure 4: The revised argument graph

8. *Every householder who does not take benefits pays for every householder who does take benefits.*
This assertion is a supporting argument for the unfairness argument labeled (5/6). Since the immediately preceding assertion (7) is better conceived as an argument in favor of the current assertion (8) rather than for assertion (5/6) directly, argument (8) is adjoined between arguments (5/6) and (7). As a consequence, the deficit of the granularity of the rule built when directly relating assertion (7) to assertion (5/6), as done in a tentative version built before, gets remedied.
9. *Professor Resicke says that recycling reduces the need for new garbage dumps.*
This assertion suggests (through the reference "Professor") that the justification for this argument may potentially originate from a specific argumentation scheme, namely appeal to expert opinion. This scheme is therefore attached to the argumentation graph (labeled as 9a).
10. *A reduction of the need for new garbage dumps is good.*
This assertion is almost identical to the previous one. However, it must be seaparated from that assertion, due to the explicit mentioning of the expert in assertion (9) (this seems to be some kind of a challenge for entailment tools).
11. *Professor Resicke is not objective.*
The formulation of this assertion is too general, as induced by the argumentation scheme attached (9a). This flaw can be remedied according to the critical question in the argumentation scheme that assertion refers to (as a follow-up question).
12. *Professor Resicke owns a recycling company.*
This assertion instantiates a specification in the critical question referred to by assertion (11), so that it simply adds on to the knowledge
13. *A person who owns a recycling company earns money from recycling.*
This assertion explains the role of the previous one. It could also be subsumed under the semantics associated with assertion (12) (similarly to the pair of assertions (5) and (6)). However, this categorization depends on how the argumentation scheme appeal to expert opinion is formalized.
14. Supermarkets create garbage. and
15. Supermarkets should pay tax.
These assertions remain unchanged as arguments.
16. *Supermarkets pass the taxes for the garbage to the consumer.*
This assertion is conceived as a rule, with the conclusion that customers pay for the garbage. The ontological recasting is indicated by labeling this assertion (16a) in Figure 4.

Figure 4 illustrates the differences obtained by these reorganization operations. Through the increased uniformity by explicitly distinguishing between facts and justifications, it was possible to unify some otherwise identical arguments that originally were specified from different perspectives. In addition, there is more ontological rigour and explicitness in the modified graphical representation, and the graph has a more chained appearance (as opposed to several, partially related sister arguments). All these properties should make this representation better suited for logical reasoning about the status of arguments and the potential of new attacks and support than the original representation.

6 CONCLUSION AND DISCUSSION

In this paper, we have described a method that aims at increasing the rigour in mapping from natural language to logical representations: an interactive process is initiated in which the user and the system cooperatively attempt to make the logical representation of a new user argument in a natural debate more accurate – through some recasting operations, the originally proposed representation is made more explicit and uniform in the embedding contexts, Admittedly, the degree of elaboration is still on some kind of anecdotal level. Some of the procedures involved may turn out to be difficult to built when it comes down to details: assessing semantic relations, building instantiations, checking subsumption, and even categorizing into facts or rules.

To a large extent, the method relies on user cooperation, both in terms of the increased effort and the hoped for improved assessment. The extra effort may be considered acceptable if the number of questions to look at is small, and if they demonstrate the system's competence – the user was able to better categorize and position a new argument being guided by a system question in several cases. Nevertheless, users may not find it easy to recognize the superiority of a more accurate alternative, since humans tend to have problems with producing explicit and accurate descriptions; the hope is that they might do better when being confronted with explicit questions and comparisons

In the future, we intend to investigate a variety of questions that carry the elaboration of our method forward, comprising technological, architectural, but also psychological questions. Technological questions concern how much NLP is required, which methods are useful to check the ontological and structural accuracy of assertions in argument graphs. Architectural questions concern how to apply/parameterize these methods, and how to convert the results into reorganization inducing questions. Finally, psychological questions concern how to encourage humans to cooperate, and which questions and how many of them to confront the user with, how to formulate and precisely when to present them, so that the user can make the best out of this material in the attempt of improving his argument formulation .

REFERENCES

[1] R. Cohen. 'Analyzing the Structure of Argumentative Discourse'. *Computational Linguistics* 13(1-2), 11-24, (1987).
[2] P.M.Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games'. *Artificial Intelligence* 77(2), 321–358, (1995).
[3] J. Sadock. 'Modus Brevis: The Truncated Argument'. In *Papers from the 13th Regional Meeting, Chicago Linguistic Society*, 545-554, (1977).
[4] S. Toulmin, 'The Use of Argument', 1958.
[5] D. Walton. 'Argumentation Schemes for Presumptive Reasoning'. Erlbaum, Mahwah, N.J. (1996).
[6] D. Walton, C. Reed, and F. Macagno. 'Argumentation Schemes'. Cambridge Univ. Press, (2008).
[7] A. Wyner, T. van Engers and A. Hunter. 'Working on the Argument Pipeline: Through Flow Issues between Natural Language Argument, Instantiated Arguments, and Argumentation Frameworks', The ECAI 2010 workshop on *Computational Models of Natural Argument*, Lisbon, Portugal, August (2010).
[8] A. Wyner, T. van Engers, and K. Bahreini. 'From policy-making statements to first-order logic'. In K. Normann Andersen, E. Francesconi, A. Gronlund, and T. M. van Engers, editors, *EGOVIS*, volume 6267 of Lecture Notes in Computer Science, 47–61, Springer, (2010).

Natural language argumentation in face of AI models

Leila Amgoud and **Henri Prade**[1]

Abstract. Formal AI models of argumentation define arguments as reasons that support claims. Such arguments may be attacked by other arguments. The main issue is then to identify the accepted ones. Works in linguistics rather focus on understanding the notion of argument, identifying its types, and describing different forms of counter-argumentation. They distinguish mainly between four forms of argumentations: two of them are arguments (as defined in AI models) and two others are rejections of arguments. Four modes of counter-argumentation are also distinguished.
This paper advocates that such typologies are instrumental for capturing real argumentations. It shows that some of the forms cannot be handled properly by AI models. Namely, rejections of arguments are partially captured. The main problem comes from the fact that AI models build on logical representations of knowledge and thus, are simple. Finally, the paper shows that the use of square of oppositions (a very old logical device) illuminates the interrelations between the different forms of argumentation.

1 Introduction

Argumentation is a social activity of reason in which a proponent agent tries to convince an opponent one that a certain statement is true (or false) by putting forwards *arguments*. While reasoning looks for the truth of a statement, argumentation looks only for persuading agents. Indeed, the proponent may succeed to persuade the opponent even if himself is not convinced by the statement.

Argumentation is an interdisciplinary topic. It is studied by philosophers like Hamblin [10], Rescher [19], Perelman and Olbrechts-Tyteca [15] and Toulmin [25]. Patterns of argumentation are studied in a pedagogical perspective for identifying fallacies in reasoning and avoiding them [6]. Argumentation also becomes an Artificial Intelligence keyword since early nineties. It is particularly used for nonmonotonic reasoning (e.g. [8, 22]) and for modeling dialogues between agents (e.g. [3, 17, 23]). Whatever the application is, the same kind of argumentation model is considered. It consists of a set of arguments supporting statements and attacks among those arguments. Acceptability semantics are then applied in order to evaluate the arguments and to decide on which statements to rely on. In all existing "concrete" models (i.e., non abstract ones), arguments are built from a knowledge base whose elements are encoded in a particular logical language. They have mainly two parts: a *conclusion* and a set of premises (called *support*). The conclusion follows logically from the support. In other words, the support is seen as a logical proof for the conclusion.

Besides, argumentation is largely studied by linguists like Salavastru [21] and Apothéloz [4]. The main focus here is on the notion of argument and its different types in real dialogues. In [4], four argumentative types are defined. Two of them are arguments and two others are rejections of arguments. In addition, Apothéloz defined four modes of counter-argumentation. Each of them may be divided into at least two distinct cases.

Our aim in this paper is to analyze the typologies of arguments and the four modes of counter-argumentation proposed in [4], and to point out the main differences with AI works on argumentation. Comparing the works of the two communities (computer scientists and linguists) is important since it allows to better understand their works and also may lead to the development of richer models of argumentation in both sides.

The paper is organized as follows: We start by presenting and analyzing the notion of argument as defined by Apothéloz in [4]. In the definition, not only the reason and the conclusion of an argument are represented but also the functions of reason and conclusion are considered. We show how this may lead to four argumentative forms where only two of them are arguments. In a subsequent section, we present in detail the four modes of counter-argumentation proposed by Apothéloz in [4]. We analyze them through several examples. We show that the notion of a counter-argument in [4] takes into account the *intention* behind the counter-argument. The next section is devoted to AI formalizations of arguments and counter-arguments. It shows how arguments are defined using an underlying logic. In this paper, we do not focus on a particular logic. We assume a general and abstract logic in which negation is encoded. We show that the notion of argument is richer in linguistics than in AI. Then, we show that some of the modes of counter-argumentation cannot be handled properly by AI models. There are two reasons for that: The first one is due to the fact that in AI models, rejections of arguments are not modeled. The second reason is related to the fact that linguists encode intentions behind arguments when defining counter-arguments while this is not possible in AI models. Finally, we show that the use of square of oppositions (a very old logical device) illuminates the interrelations between the different forms of argumentation.

2 Argumentative Forms in Linguistics

In [4], an argument is a pair $\mathcal{C}(x) : \mathcal{R}(y)$ where $\mathcal{C}$ is the *function of concluding* and x its content, $\mathcal{R}$ is the *function of reason* and y its content. The argument is read as follows: y is a reason for concluding x. We say that y is *argumentatively oriented* toward x. The contents x and y may either be premises (propositions) or arguments as we will see in the next section. Moreover, an argument is an *enthymeme*, i.e., an incomplete syllogism. Indeed, some generic rules relating y to x are left implicit. For instance, the argument "Mary will miss her exams (me) since she did not work hard (wh)"

[1] amgoud, prade@irit.fr

is written as $\mathcal{C}(me) : \mathcal{R}(\neg wh)$. Thus, the rule stating that "not working hard leads to missing exams" is not made explicit in the reason part of the argument. This is not surprising since linguists are concerned by natural language arguments, which are very often enthymemes.

In AI works on argumentation, the functions of conclusion and reason are implicit in the formal definition of an argument. However, we will see that making explicit these functions is of great importance in 'natural language' counter-argumentation. Besides, the two contents x and y are formally defined. They are generally propositions, except in [12, 26] where they may be arguments. Finally, in AI models the link between x and y is defined whereas in the work of Apothéloz, it is not.

Due to the presence of functions and contents, Apothéloz argues that there are two forms of negation: one for refuting a function and one for refuting its content. Refuting a function does not mean that its content is also refuted. The difference between the two negations is similar to the difference between $\vdash \neg p$ and $\not\vdash p$ (where p is a propositional formula and $\vdash$ stands for the classical consequence relation). Let $-$ denote both types of negation. These double negations give birth to four basic argumentative forms:

c_1	$\mathcal{C}(x) : \mathcal{R}(y)$	y is a reason for concluding x
c_2	$\mathcal{C}(x) : -\mathcal{R}(y)$	y is not a reason for concluding x
c_3	$-\mathcal{C}(x) : \mathcal{R}(y)$	y is a reason against concluding x
c_4	$-\mathcal{C}(x) : -\mathcal{R}(y)$	y is not a reason against concluding x

The contents x and y can themselves be replaced by their negation, leading to a combinatorics of 16 distinct argumentative forms, which includes $\mathcal{C}(-x) : \mathcal{R}(y)$ (y is a reason for concluding 'not x'), or $\mathcal{C}(x) : \mathcal{R}(-y)$ ('not y' is a reason for concluding x). It is worth noticing that only the forms c_1 and c_3 are arguments. The forms c_2 and c_4 are rejections of arguments. The form c_1 allows the representation of two epistemic states: one in which x is true and one in which x is false (i.e., $\neg x$ is true). However, the form c_3 encodes *ignorance* wrt. x. It expresses the fact that the conclusion x cannot be made but this does not mean neither that $-x$ is true. Let us illustrate the four forms by a dialogue between agents A, B, C, D.

A: Clara is at home (h). There is light from her window (l).

B: The fact that there is light from the window does not mean that she is at home.

C: But, she is on vacation! (v)

D: The fact that she is on vacation does not mean that she cannot be at home.

Agent A presents the argument $\mathcal{C}(h) : \mathcal{R}(l)$ which is of form c_1. Agent B rejects this argument. Note that B is not refuting l (i.e., he is not saying that there is no light from Clara's window). He is neither saying that the conclusion h is false, but he is refuting the fact that l may play the function of reason in favor of h. This move is written as $\mathcal{C}(h) : -\mathcal{R}(l)$, that is of the form c_2. Apothéloz argued that this rejection aims at refuting $\mathcal{C}(h)$, thus it can be considered as an argument, $-\mathcal{C}(h) : \mathcal{R}(\mathcal{C}(h) : -\mathcal{R}(l))$, which is read as follows: the fact of rejecting the argument $\mathcal{C}(h) : \mathcal{R}(l)$ gives a reason for suspending the conclusion $\mathcal{C}(x)$. The agent C does not know whether Clara is at home or not, but thinks that he has a good reason for suspending the conclusion h. Indeed, since Clara is on vacation, then one cannot confirm that she is at home. The argument of c is encoded as $-\mathcal{C}(h) : \mathcal{R}(v)$, i.e., it has the form c_3. Note that the negation is on the function $\mathcal{C}$ and not on the content h since $\neg h$ would mean that c thinks that Clara is not at home while this is not the case. Agent D thinks that the fact that Clara is on vacation is not a sufficient reason for suspending the conclusion h. This move is then encoded as $-\mathcal{C}(h) : -\mathcal{R}(v)$.

3 Counter-Argumentation in Linguistics

Some linguists studied the different ways of defining a counter-argumentation, i.e., how to attack a given argument. A prominent work is done by Apothéloz [4]. Indeed, Apothéloz identified four modes of arguing against a given argument $\mathcal{C}(x) : \mathcal{R}(y)$:

1. Disputing the *plausibility* or the truth of the propositions used in y.
2. Disputing the *completeness* of the reason y. This is done by providing a new reason that is anti-oriented to the conclusion x, and that is presented as being more decisive than the reason y.
3. Disputing the *relevance* of the reason with respect to the conclusion x.
4. Disputing the *argumentative orientation* of the reason, by stating that the reason considered is rather in favor of $-x$, or is at least not in favor of x.

Throughout the paper, $\mathcal{K}$ stands either for $\mathcal{C}(-x)$ or for $-\mathcal{C}(x)$.

3.1 Disputing the Plausibility of a Reason (DPR)

Disputing the plausibility of the reason of an argument $\mathcal{C}(x) : \mathcal{R}(y)$ amounts to prove that y is false. Apothéloz argued that there are three ways for doing that:

1. By asserting an argument of the form $\mathcal{K} : \mathcal{R}(-y)$. In this case, no reason is given in favor of $-y$. Let us consider the following example.

   a_1: Clara will miss her exams (me). She did not work hard ($-wh$).

   a_2: Clara? She did not stop working!

   The argument a_1 is written as $\mathcal{C}(me) : \mathcal{R}(-wh)$. The counter-argument a_2 intends blocking the conclusion me and is thus encoded as $-\mathcal{C}(me) : \mathcal{R}(wh)$. Recall that this does not mean that $-me$ is true or even supported.
2. By asserting an argument $\mathcal{K} : \mathcal{R}(\mathcal{C}(-y) : \mathcal{R}(z))$, that is by providing a reason against y as illustrated below.

   a_3: No, she worked hard. Her eyes are encircled (ee).

   Here, not only the premise $-wh$ is denied but it is also supported by a reason, that is $\mathcal{C}(wh) : \mathcal{R}(ee)$. This argument gives a reason for not concluding me, thus the following argument: $-\mathcal{C}(me) : \mathcal{R}(\mathcal{C}(wh) : \mathcal{R}(ee))$.
3. By asserting an argument of the form $\mathcal{C}(\mathcal{C}(x) : -\mathcal{R}(y)) : \mathcal{R}(-y)$. Here, the fact of denying y is considered as a reason for rejecting the whole argument $\mathcal{C}(x) : \mathcal{R}(y)$. This is illustrated by the following example:

   a_4: Clara works hard (wh) because she is ambitious (am).

   a_5: It is not by ambition that Clara works hard. She is not ambitious.

The argument a_4 is written as $\mathcal{C}(wh) : \mathcal{R}(am)$. The intention behind a_5 is not to suspend (or to deny) the conclusion wh as in the two previous cases. The agent providing this argument seems agree on wh but not on am. His intention then, is to reject the whole argument a_4. Thus, a_5 is defined as $\mathcal{C}(\mathcal{C}(wh) : -\mathcal{R}(am)) : \mathcal{R}(-am)$. Note that the conclusion of a_5 is a rejection of an argument.

To sum up, by denying the reason y of an argument $\mathcal{C}(x) : \mathcal{R}(y)$, one intends either blocking the conclusion x (cases 1 and 2) or rejecting the whole argument $\mathcal{C}(x) : \mathcal{R}(y)$ (case 3). Moreover, $-y$ may be supported or not by another reason.

3.2 Disputing the Completeness of a Reason (DCR)

Unlike the previous case where the reason y of an argument $\mathcal{C}(x) : \mathcal{R}(y)$ is false, here it is accepted but it is not sufficient to conclude x. This is due to the existence of a stronger argument which is anti-oriented toward the conclusion x. In [4], it is argued that this task can be achieved in two ways:

1. By asserting an argument of the form $\mathcal{K} : \mathcal{R}(z)$ where z is anti-oriented toward x. The following example illustrates this case:

   a_1: Clara will miss her exams (me). She did not work hard (wh).

   a_6: Clara will not miss her exams. She is very smart (sm).

   Here the agent who uttered the argument a_6 may agree that the premise $-wh$ is true, but thinks that it is *not sufficient* to conclude me. Indeed, there is a *stronger* reason which prevents this conclusion. Thus, the argument a_6 is given as $\mathcal{C}(-me) : \mathcal{R}(sm)$.

   Let us consider now the following alternative reply to a_1 in the previous dialogue.

   a_7: But Clara is very smart.

   In this case, the agent does not know whether Clara will miss or not her exams but he provides an argument against concluding that she will miss them. Thus, a_7 is as follows: $-\mathcal{C}(me) : \mathcal{R}(sm)$. It is worth noticing that this example is similar to the following one provided by Pollock in [16].

   a_8: This object is red (or) since it looks red (lr).

   a_9: But the object is illuminated by a red light (irl).

   The argument a_8 is written as $\mathcal{C}(or) : \mathcal{R}(lr)$ while the argument a_9 is defined as $-\mathcal{C}(or) : \mathcal{R}(irl)$ and its role is to prevent concluding or.
2. The second possibility is more tricky. It consists of giving a reason that is in favor of y but which is anti-oriented toward the conclusion x. The counter-argument has the form: $\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$. Let us illustrate this form of counter-argumentation by a simple example:

   a_{10}: Paul is in his office (of) because his car is in the carpark (pa).

   a_{11}: But the car is in the carpark because it is broken down (br).

   According to the argument a_{10}, written as $\mathcal{C}(of) : \mathcal{R}(pa)$, the fact that Paul's car is in the carpark is a reason to think that Paul is still in his office. The reply a_{11} gives an explanation why the car is in the carpark: thus an argument $\mathcal{C}(pa) : \mathcal{R}(br)$. However, this explanation is anti-oriented toward the conclusion of, i.e., it blocks this conclusion. The argument a_{11} is thus defined as follows: $-\mathcal{C}(of) : \mathcal{R}(\mathcal{C}(pa) : \mathcal{R}(br))$.

   It is worth mentioning that in AI works on bipolar argumentation systems, namely the work by Cayrol and Lagasquire [7], an argument whose conclusion is a premise of another argument is considered as a support for this latter. Namely, they would consider the argument $\mathcal{C}(pa) : \mathcal{R}(br)$ as *supporting* the argument a_{10} (since its conclusion is exactly a premise of a_{10}). A support relation has a positive flavor and the conclusion of an argument which is supported by other arguments is strengthen. Unfortunately, the previous dialogue shows clearly that this is not always the case, and the conclusion "Paul is in his office" (of) is suspended after receiving the argument a_{11}.

3.3 Disputing the Relevance of a Reason (DRR)

The third way of attacking an argument $\mathcal{C}(x) : \mathcal{R}(y)$ is by disputing the relevance of the reason y with respect to the conclusion x. What is denied is neither x nor y but the fact that y may constitute a reason for x. This can be done in three ways:

1. By giving an argument of the form $\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$ showing that y is irrelevant for x. This is exactly the case of the previous dialogue where the fact that Paul's car is broken down explains why the car being in a car park is not a reason for concluding that Paul is in his office. Note that in this case it is both a matter of irrelevance and incompleteness of the reason.
2. By blocking the conclusion x via a rejection of the argument as follows: $-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$. Let us illustrate this case by considering the argument a_1 and with the reply a_{12}.

   a_1: Clara will miss her exams (me). She did not work hard ($-wh$).

   a_{12}: Indeed, she did not work hard, but not working hard is not a reason to necessarily miss her exams.

   The intention behind such an argument is clearly to suspend the conclusion me by rejecting the fact that $-wh$ may play the role of a reason in favor of me. Note that in this reply, it is admitted that Clara does not work hard (i.e., the reason y is true).
3. By rejecting the argument, i.e., by uttering $\mathcal{C}(x) : -\mathcal{R}(y)$. An example would be:

   a_{13}: She will not miss her exams because she did not work hard, but rather because of the stress (st).

   In this example both x and y are recognized as true, but y is not the real reason for x being true. The real reason is st, that is $\mathcal{C}(me) : \mathcal{R}(st)$. Note that $\mathcal{C}(me) : \mathcal{R}(st)$ alone does not express the fact that the first argument is attacked or rejected. The rejection is expressed by $\mathcal{C}(me) : -\mathcal{R}(-wh)$.

3.4 Disputing the Argumentative Orientation of a Reason (DOR)

The fourth mode of counter-argumentation in [4] consists of disputing the argumentative orientation of the reason. The idea is that the reason y is not in favor of the conclusion x as stated in the argument $\mathcal{C}(x) : \mathcal{R}(y)$ but in favor of the opposite conclusion, that is $\mathcal{C}(-x) : \mathcal{R}(y)$. Let us illustrate this idea by the following example borrowed from [5].

a_{14}: 'A World Apart' is not a good film ($-gf$). It does not teach us anything new about apartheid ($-ta$).

a_{15}**:** That's precisely what makes it good.

The argument a_{14}, written as $\mathcal{C}(-gf) : \mathcal{R}(-ta)$, supports $-gf$ with the premise $-ta$. The counter-argument a_{15}, $\mathcal{C}(gf) : \mathcal{R}(-ta)$, supports the opposite conclusion with the same premise.

4 Argumentative Forms in AI

In the previous section, we have shown how arguments are defined by linguists. The definition is semi-formal since the link between the support and the conclusion is not specified, and the properties of the two functions are not clear. From the multiple examples given in [4] and from other works on natural language argumentation (e.g., [18, 20]), arguments are enthymemes. Thus, the content of the reason function leaves generic rules aside. For instance, the argument stating that Clara will miss her exams since she did not work hard ($\mathcal{C}(me) : \mathcal{R}(-wh)$) is based on an implicit generic rule which is 'not working hard leads to missing exams'.

Besides, in AI models of argumentation, arguments are defined from a knowledge base whose elements are formulas of a logical language. Arguments are logical proofs for their conclusions. They are defined using the consequence operator that is associated with the logical language. This logic (i.e., the language and the corresponding consequence operator), called *base logic* in [11], is monotonic (e.g. [1, 2, 22]). In this section, we show the type of arguments that can be modeled, and analyze how to encode the different modes of counter-argumentation defined in linguistics, namely in [4].

Throughout this section, we assume a logical language $\mathbb{L}$ in which two sets are distinguished: a set $\mathbb{F}$ of *facts* and a set $\mathbb{R}$ of *generic rules*. Facts concern particular instances, like 'Tweety is a bird', whereas generic rules concern classes of instances, like 'Generally birds fly' and 'Penguin do not fly'. This distinction is important for recovering some of the previous modes of counter-argumentation. Apart from this distinction, the only requirement that is imposed on $\mathbb{L}$ is that it contains a connector of negation, denoted by $-$. Thus, $\mathbb{L}$ may be any language, for instance, a propositional language or the rule-based language used in the ASPIC argumentation system [2]. In ASPIC, certain generic rules (like 'Penguin do not fly') are encoded by strict rules whereas defeasible ones (like 'Generally birds fly') are encoded by defeasible rules. Finally, facts are literals gathered in a knowledge base.

Let $\mathtt{CN}$ be a consequence operator, that is $\mathtt{CN} : 2^{\mathbb{L}} \to 2^{\mathbb{L}}$. We do not assume particular requirements on $\mathtt{CN}$. Finally, from the logic $(\mathbb{L}, \mathtt{CN})$, a notion of *consistency* is defined as in [24], that is a set $X \subseteq \mathbb{L}$ is consistent iff $\mathtt{CN}(X) \neq \mathbb{L}$. Propositional logic is used in some places only to illustrate issues. An argument is defined as follows:

Definition 1 (Argument) *An* argument *is a pair* (x, y) *s.t.*

- $y \subseteq \mathbb{L}$
- *y is consistent*
- $x \in \mathtt{CN}(y)$
- $\nexists y' \subset y$ *s.t.* $x \in \mathtt{CN}(y')$

x is the conclusion *of the argument whereas y is its* reason/support.

In this definition, the function of reason and that of conclusion are not explicit. However, their contents are clearly defined. These contents cannot be arguments, thus arguments of the forms $\mathcal{K} : \mathcal{R}(\mathcal{C}(-y) : \mathcal{R}(z))$, or $\mathcal{C}(\mathcal{C}(x) : -\mathcal{R}(y)) : \mathcal{R}(-y)$ cannot be expressed in our formal setting. Another key difference with the definition of linguists is that arguments are not entymemes. Assume that $(\mathbb{L}, \mathtt{CN})$ is propositional logic, then the argument a_1, $\mathcal{C}(me) : \mathcal{R}(-wh)$, is written as follows in the previous definition: $(me, \{-wh, -wh \to me\})$. The generic rule $-wh \to me$ is left implicit in $\mathcal{C}(me) : \mathcal{R}(-wh)$. Finally, remember that Apothéloz defined four basic argumentative forms: $\mathcal{C}(x) : \mathcal{R}(y)$, $-\mathcal{C}(x) : \mathcal{R}(y)$, $\mathcal{C}(x) : -\mathcal{R}(y)$ and $-\mathcal{C}(x) : -\mathcal{R}(y)$. Only the two first ones are arguments and the two others are rejections of arguments. The above definition only captures one form of arguments: $\mathcal{C}(x) : \mathcal{R}(y)$. Indeed, it allows to provide a reason either for x or for $-x$, but it does not block conclusions, i.e., does not express *ignorance* wrt x. Thus, $-\mathcal{C}(x) : \mathcal{R}(y)$ cannot be expressed in Definition 1. Note that this drawback is shared by those argumentation systems that reason about arguments [12, 26], i.e., where arguments may support other arguments.

Let us now analyze how an argument (x, y) may be attacked. Four different ways are distinguished:

1. *By building a new argument in favor of the opposite conclusion*, i.e., $(-x, z)$. This relation is known as *rebuttal* in [9]. Indeed, an argument rebuts another iff they have opposite conclusions. Note that this form of counter-argumentation corresponds to the first way of disputing the completeness of a reason in [4]. Thus, the argument a_6 (written as $(-me, \{sm, sm \to -me\})$ under propositional logic) rebuts the argument a_1. This relation captures also the fourth mode of counter-argumentation, that is disputing the argumentative orientation of a reason. For instance, the arguments a_{14} and a_{15} are encoded respectively as $(-gf, \{-ta, -ta \to -gf\})$, $(gf, \{-ta, -ta \to gf\})$. Note that in this case, the disagreement comes from the generic rules. From the same information $-ta$, one of them leads to gf while the other concludes $-gf$. This situation may be more complicate. Imagine the two following arguments: $(x, \{y, y \to x\})$ and $(-x, \{y, y \to z, z \to -x\})$. From y and following different paths, the two arguments lead to opposite conclusions.
2. *By disputing a fact in the support y*. This amounts to build an argument (x', z) where x' is $-t$ and $t \in \mathbb{F} \cap y$. This relation is known in argumentation literature as *assumption attack* [9]. At a first glance, it seems to correspond exactly to disputing the plausibility of a reason in [4], especially since arguments are enthymemes in that work, thus the content of the reason is facts. However, this is not always the case. Indeed, since Definition 1 does not allow neither blocking conclusions nor supporting arguments, the intentions behind the three cases of disputing the plausibility of a reason cannot be encoded. Let us revisit the examples presented before. The two arguments a_1 and a_2 are encoded as follows: $a_1 = (me, \{-wh, -wh \to me\})$ and $a_2 = (wh, \{wh\})$ while in [4], $a_2 = -\mathcal{C}(me) : \mathcal{R}(wh)$. The reply a_3 is defined as $(wh, \{ee, ee \to wh\})$ while Apothéloz writes $-\mathcal{C}(me) : \mathcal{R}(\mathcal{C}(wh) : \mathcal{R}(ee))$. Finally, the two arguments a_4 and a_5 are defined respectively as: $(wh, \{am, am \to wh\})$, $(-am, \{-am\})$ while a_5 is written as $\mathcal{C}(\mathcal{C}(wh) : -\mathcal{R}(am)) : \mathcal{R}(-am)$ by Apothéloz.
3. *By disputing the applicability of a generic rule t in the support y*, i.e., $t \in y \cap \mathbb{R}$. The idea is that the rule t is true in general but not applicable in a certain situation. This relation, called *undercut*, was defined in [16]. Several cases discussed by Apothéloz

fall into this relation. The first way of disputing the completeness of a reason can be captured by this relation. Indeed, the argument $a_7 = -\mathcal{C}(me) : \mathcal{R}(sm)$ is against applying the generic rule $-wh \rightarrow me$ when a person is smart (sm). The argument $a_9 = -\mathcal{C}(or) : \mathcal{R}(irl)$ aims at blocking the application of the rule ('when an object looks red the it is red' ($lr \rightarrow or$)) when the object is illuminated by a red light (irl). Similarly, the argument a_{11} blocks the applicability of the generic rule saying that if Paul's car is in the carpark, then Paul is in his office ($pa \rightarrow of$). It is important to notice that the phenomenon of blocking a generic rule raises in *default reasoning*. Indeed, a rule is blocked in presence of an *exception*.

4. *By disputing a generic rule*, that is by asserting that it is false. This is typically what happens in the second way of refuting the relevance of a reason. Let us consider the argument a_{12}. It says that just because Clara did not work hard is not a reason to miss her exams'. Here the agent recognizes that Clara did not work hard. So what is disputed is the plausibility of the rule $-wh \rightarrow me$. This is again captured by assumption attack which consists of undermining an element of the support of an argument.

Table 1 summarizes the four modes of attacking an argument $\mathcal{C}(x) : \mathcal{R}(y)$ as defined in [4] as well as the ways of capturing them in an AI model. It shows that most of the modes of counter-argumentation are only partially modeled in our logical formalism. Indeed, the intention behind each attack is not captured. Moreover, at a formal level we do not make any difference between the four cases of applying assumption attack. Similar comment holds for undercut and rebut. While the differences may be crucial for evaluating arguments. Indeed, disputing a fact is not like disputing a generic rule and refuting a fact by providing a new reason is not like rejecting the fact without justification. Moreover, from a dialogical point of view, it is important to be able to represent accurately the moves of the agents. In our formalism, the rejection of an argument (DRR3) is not possible while such a move is very common in dialogues.

5 Organizing Argumentative Statements in a Square of Opposition

A key point in the categorization introduced by Apothéloz in [4] is the presence of two kinds of negation, one pertaining to the contents x or y, and the other to the functions $\mathcal{R}$ or $\mathcal{C}$. It has been observed that such a double system of negations gives birth to a formal logical structure called *square of opposition*, which dates back Aristotle's time (see, e.g., [14] for a historical and philosophical account). We first briefly recall what this object is, since it has been somewhat neglected in modern logic.

5.1 Classical Squares of Opposition

It has been noticed for a long time that a statement (A) of the form "every a is p" is negated by the statement (O) "some a is not p", while a statement like (E) "no a is p" is clearly in even stronger opposition to the first statement (A). These three statements, together with the negation of the last statement, namely (I) "some a is p", give birth to the square of opposition in terms of quantifiers $A : \forall a\, p(a)$, $E : \forall a\, \neg p(a)$, $I : \exists a\, p(a)$, $O : \exists a\, \neg p(a)$, pictured in Figure 5.1. Such a square is usually denoted by the letters A, I (affirmative half) and E, O (negative half). The names of the vertices comes from a traditional Latin reading: **A**ff**I**rmo, n**E**g**O**). Another standard example of the square of opposition is in terms of modalities: $A : \Box r$, $E : \Box\neg r$, $I : \Diamond r$, $O : \Diamond\neg r$. As can be seen from these two examples, different relations hold between the vertices, which give birth to the following definition:

Definition 2 (Square of opposition) *Four statements A, E, O, I make a* square of opposition *if and only if the following relations hold:*

1. *A and O are the negation of each other, as well as E and I;*
2. *A entails I, and E entails O;*
3. *A and E cannot be true together, but may be false together, while*
4. *I and O cannot be false together, but may be true together.*

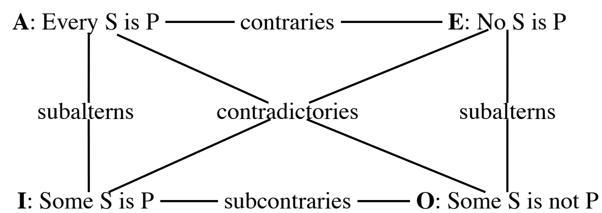


Figure 1. Square of opposition

Note that A entails I pressupposes in the example of the Figure 5.1 that $\{s \mid p(s)\ is\ true\} \neq \emptyset$, otherwise A cannot entail I since there is no s. Similarly $r \not\equiv \bot$ is assumed in the modal logic case.

5.2 A Square of Opposition for Argumentation

The observation that two negations are at work in the argumentative statements classified by Apothéloz [4] has recently led Constantin Salavastru [21] to propose to organize the four basic statements into a square of opposition; see also [13]. However, his proposal may be discussed on one point, as we are going to see. Indeed, taking $\mathcal{C}(x) : \mathcal{R}(y)$ for vertex A, leads to take its negation $\mathcal{C}(x) : -\mathcal{R}(y)$ for O. Can we take $-\mathcal{C}(x) : \mathcal{R}(y)$ for E? This first supposes that A and E are mutually exclusive, which is clearly the case. Then, we have to take the negation of E for I, i.e., $-\mathcal{C}(x) : -\mathcal{R}(y)$. We have still to check that A entails I and E entails O, as well as condition (4) above. If y is a reason for not concluding x, then certainly y is not a reason for concluding x, so E entails O; similarly y is a reason for concluding x entails that y is not a reason for not concluding x, i.e., A entails I. Finally, y may be a reason neither for concluding x nor for not concluding x. This gives birth to the argumentative square of opposition of Figure 2.

It can be checked that the contradiction relation (1) holds, as well as the relations (2), (3), and (4) of Def. 2.

Proposition 1 *The four argumentative forms $A = \mathcal{C}(x) : \mathcal{R}(y)$, $E = -\mathcal{C}(x) : \mathcal{R}(y)$, $O = \mathcal{C}(x) : -\mathcal{R}(y)$, $I = -\mathcal{C}(x) : -\mathcal{R}(y)$ make a square of opposition.*

Note that we should assume that $\mathcal{C}(x) : \mathcal{R}(y)$ is not self-contradictory (or self-attacking) in order that the square of opposition makes sense. In propositional logic, this would mean that $x \wedge y \neq \bot$.

This square departs from the one obtained by Salavastru in [21] where vertices A and I as well as E and O are exchanged: in other

DPR1	$\mathcal{K} : \mathcal{R}(-y)$	Assumption attack on facts
DPR2	$\mathcal{K} : \mathcal{R}(\mathcal{C}(-y) : \mathcal{R}(z))$	Assumption attack on facts
DPR3	$\mathcal{C}(\mathcal{C}(x) : -\mathcal{R}(y)) : \mathcal{R}(-y)$	Assumption attack on facts
DCR1	$\mathcal{C}(-x) : \mathcal{R}(z)$	Rebut
DCR2	$-\mathcal{C}(x) : \mathcal{R}(z)$	Undercut
DCR3	$\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$	Undercut
DRR1	$\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$	Undercut
DRR2	$-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$	Assumption attack on rules
DRR3	$\mathcal{C}(x) : -\mathcal{R}(y)$	?
DOR	$\mathcal{C}(-x) : \mathcal{R}(y)$	Rebut

Table 1. The four modes of counter-argumentation and attack relations

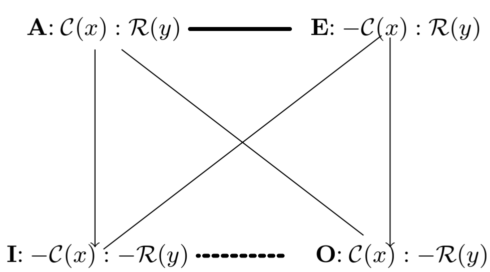


Figure 2. An argumentative square of opposition

words the entailments (2) are put in the wrong way. This may come from a misunderstanding of the remark made in [4] that the rejection $\mathcal{C}(x) : -\mathcal{R}(y)$ is itself a reason for not concluding x, which can be written $-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$. But this does not mean that $\mathcal{C}(x) : -\mathcal{R}(y)$ entails $-\mathcal{C}(x) : \mathcal{R}(y)$ since it may be the case, for instance, that $\mathcal{C}(-x) : \mathcal{R}(y)$. Salavastru made another mistake regarding the link between A and I. He assumed that I entails A. Let us show through a simple example that this implication is false, but it is rather in the other way around.

a_{16}**:** The fact that Paul is a French citizen fr is not a reason to not conclude that he is smart st.

This is clearly a statement of form c_4, i.e., $-\mathcal{C}(sm) : -\mathcal{R}(fr)$. The question now is: does this statement entails the argument $\mathcal{C}(sm) : \mathcal{R}(fr)$ (i.e., the fact that Paul is french is a reason to conclude that he is smart)? The answer is certainly no. However, the converse is true. That is $\mathcal{C}(sm) : \mathcal{R}(fr)$ implies $-\mathcal{C}(sm) : -\mathcal{R}(fr)$.

6 Conclusion

This paper reported a very interesting work by linguists on argumentation theory, and analyzed it from an AI perspective. Indeed, we have shown how linguists define the notion of argument by making explicit two functions: a function of conclusion and a function of reason. This allows also to have two types of negation: one for refuting a function and another one for disputing its content. These double negations give birth to four argumentative forms: two of which are arguments and two others are only rejections of arguments. We have shown through examples that the four forms are meaningful and very frequent in natural language dialogues. We have then shown the four modes of counter-argumentation proposed by Apothéloz in [4]. Each mode can itself have various cases. We have then defined the notion of argument and counter-argument in a more formal way as it is done in AI. We have shown that the formal definition captures only one argumentative form among the four proposed by Apothéloz. As a side effect, the different modes of counter-argumentation cannot all be captured. Moreover, the ones which are captured are only encoded partially. The last contribution of this paper consists of showing that the proposal of Apothéloz makes sense since it obeys the properties of a square of opposition. Indeed, we have shown that the four argumentative forms constitute a square of opposition. A future work would be then to develop a rich argumentation system that captures the various modes of argumentation and counter-argumentation.

REFERENCES

[1] L. Amgoud and Ph. Besnard, 'Bridging the gap between abstract argumentation systems and logic', in *International Conference on Scalable Uncertainty Management (SUM'09)*, pp. 12–27, (2009).

[2] L. Amgoud, M. Caminada, C. Cayrol, MC. Lagasquie, and H. Prakken, 'Towards a consensual formal model: inference part', in *Deliverable D2.2 of ASPIC project*, (2004).

[3] L. Amgoud, Y. Dimopoulos, and P. Moraitis, 'A unified and general framework for argumentation-based negotiation', in *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*, pp. 963–970. ACM Press, (2007).

[4] D. Apothéloz, 'Esquisse d'un catalogue des formes de la contre-argumentation', *Travaux du Centre de Recherches Sémiologiques*, **57**, 69–86, (1989).

[5] D. Apothéloz, P. Brandt, and G. Quiroz, 'Champ et effets de la négation argumentative : contre-argumentation et mise en cause', *Argumentation*, **6**, 99–113, (1992).

[6] P. Blackburn, *Logique de l'Argumentation*, Editions du Renouveau Pédagogique, Saint-Laurent, Québec, 1989.

[7] C. Cayrol and M. Lagasquie, 'On the acceptability of arguments in bipolar argumentation frameworks', in *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'2005)*, pp. 378–389, (2005).

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

[9] M. Elvang-Goransson, J. Fox, and P. Krause, 'Dialectic reasoning with inconsistent information', in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pp. 114–121, (1993).

[10] C. L. Hamblin, *Fallacies*, Methuen, London, UK, 1970.

[11] A. Hunter, 'Base logics in argumentation', in *Proceedings of COMMA'10*, pp. 275–286.

[12] S. Modgil and T. Bench-Capon, 'Metalevel argumentation', *Journal Logic and Computation*, **21**(6), 959–1003, (2011).

[13] A. Moretti, 'Argumentation theory and the geometry of opposition (abstract)', in *7th Conference of the Inter. Soc. for the Study of Argumentation (ISSA'10)*, (2010).

[14] T. Parsons, ‘The traditional square of opposition’, *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, (2008).
[15] C. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: a Treatise on Argumentation.*, Notre Dame Press, University of Notre Dame, 1969.
[16] J. Pollock, ‘How to reason defeasibly’, *Artificial Intelligence Journal*, **57**, 1–42, (1992).
[17] C. Reed, ‘Dialogue frames in agent communication’, in *Proceedings of the 3rd International Conference on Multi Agent Systems*, pp. 246–253, (1998).
[18] C. Reed, ‘The role of saliency in generating natural language arguments’, in *International Joint Conference on Artificial intelligence (IJCAI’99)*, pp. 876–881, (1999).
[19] N. Rescher, *Dialectics: A controversy-oriented approach to the theory of knowledge*, State University of New York Press, 1977.
[20] J. Sadock, ‘Modus brevis: The truncated argument’, in *13th Regional Meeting of the Chicago Linguistics Society*, pp. 545–554, (1977).
[21] C. Salavastru, *Logique, Argumentation, Interprétation*, L’Harmattan, Paris, 2007.
[22] G.R. Simari and R.P. Loui, ‘A mathematical treatment of defeasible reasoning and its implementation’, *Artificial Intelligence Journal*, **53**, 125–157, (1992).
[23] K. Sycara, ‘Persuasive argumentation in negotiation’, *Theory and Decision*, **28**, 203–242, (1990).
[24] A. Tarski, *On Some Fundamental Concepts of Metamathematics*, Logic, Semantics, Metamathematic. Edited and translated by J. H. Woodger, Oxford Uni. Press, 1956.
[25] S. Toulmin, *The Uses of Argument*, Cambridge University Press, 1958.
[26] M. Wooldridge, P. McBurney, and S. Parsons, ‘On the meta-logic of arguments’, in *ArgMAS*, pp. 42–56, (2005).

Some Aspects of a Preliminary Analysis of Argumentation in Western Tonal Music

Patrick Saint-Dizier [1]

Abstract. In this paper we present some aspects of argumentation as it is realized in a non-verbal system: western tonal music. We show via examples (Bach's organ Passacaglia and Beethoven's piano sonatas) that argumentation is very much developed in music, but very abstract, leaving space for various interpretations. We investigate the means used by these two composers and analyze their psychological impact on the mental state of the listener.

1 Introduction

It may seem at first glance that music has limited relation with argumentation. It is commonly admitted that arguments essentially have a language support with a clear contents, coming from either written documents (e.g. news) or oral communication (e.g. dialogue, political speech). Argumentation is a rational process whose goal is to convince someone of a certain statement or view. Nevertheless, non-verbal media, which may seem less rational, such as sound, images and video may also play a major role in argumentation.

The very preliminary analysis presented here originates from the fact that music is a natural, but very abstract, language, with a well-defined syntax for low level constructions (melody, harmony) as well as for high level constructions (musical rhetorics). The structure of a piece of music cannot be accounted for using the tools used in computational linguistics: any attempt to find close analogies would probably fail, but there are some principles which can be considered [3, 6].

Instead of developing their ideas and feelings via language, composers used the means offered by music, via its emotional and symbolic dimensions, to characterize mental states, which originate from both melodies and harmony but also from its very structure. A number of composers had internal debates about essential points in their lives, which can be very clearly identified as arguments. These arguments can get a very strong emotional strength, rarely reached in language. These emotional states can then lead to rational attitudes.

The work we present here remains largely exploratory. Considering a well-identified period of musical production (western tonal music), we aim at elaborating a model that accounts for the means developed by composers to develop arguments, argumentation and the associated rhetorical relations. Identifying and modelling these means is a useful challenge and an interesting extension to argumentation analysis and development. It is also a set of means which could be re-used in concrete communication situations. In parallel with the linguistic (semantics and pragmatics) aspects of argumentation and related schemes (e.g. [9]), which are now relatively well developed, we argue that it is possible and of interest to investigate more abstract modes or psychological approaches that give a more comprehensive and global analysis to argumentation.

The relations between music and language and reasoning have not been much investigated. Let us note the very insightful investigation of Generative Theory applied to music production [1, 3], which essentially addresses syntactic aspects. To the best of our knowledge, no work has been carried out to model the rhetoric and argumentation aspects of music following the analysis principles applied to language. There is a relatively abundant literature on musical rhetorics, but oriented towards musical analysis or production, which serves very different aims.

2 Music as a science of numbers and proportions

Music is basically a science based on numbers and proportions. Pythagoras was probably the first, in our western world, to initiate this view, with the well-known definition *music are numbers made audible*. Till the Renaissance, music was part of the *Quadrivium* together with geometry, arithmetics and astronomy. Saint Augustine (in the confessions and De Musica) and Boece (470-525, in the consolation) justify that music is a science, with rational knowledge based on numbers, that manages the harmony of movements. Roughly, music is not only a mathematical object that accounts for harmony and rhythm, but it is also a form of abstraction, with a strong explicative power, that reflects creativity and perfection. The Medieval period developed a very strong view of music via a metaphysics of sound organization: music becomes a part of theology. In the Gregorian song, music is viewed as a 'perfect sound with a unified view of body movements, pitch, metrics and text', it is an art of the orator (jubilus).

From the Renaissance, music was associated with a more analytical vision, with, among others, the following major points of investigations:

- analysis of proportions and their 'psychological' effects, e.g.: proportions between notes (intervals in a melody), between durations, leading to rhythms.
- analysis of the facets of the tension-resolution mechanism in tonal harmony, which allows the introduction of contrasts,
- from the two points above points, development of polyphony techniques together with their symbolism and the analysis of their communicative dimension on the listeners, culminating in the late baroque period,
- analysis of the numerous types of metaphors introduced in the construction of melodies and in harmony: orientational (moving up is positive, moving down is negative), spatial (ambitus, distance between voices or notes, etc.) and metaphors based on colours

[1] IRIT-CNRS Toulouse France, email: stdizier@irit.fr

(keys and chords are associated with colours, modulations and sequences of chords draw colour ondulations).

- development and analysis of musical structures and their underlying symbolism, with the emergence of typical forms (e.g. ricercar for scholarly music, forms of dance such as the minuet for more popular forms such as ternary forms) and techniques of thematic development. From these elements emerged a rhetorics of music, with a first stage culminating in the late-baroque with J. Mattheson treaty on musical rhetorics (circa 1722).

3 Argumentation in Tonal Music

3.1 Some epistemic considerations

In the examples below, we sketch some basic elements in musical discourse which are proper to argumentation:

- it is a system that uses all the means of the 'language' at stake: argumentation is not an abstract process. Arguments and argumentation are constructed from the means offered by the language: sounds, pitch, harmony, duration at a low level, and formation of themes, musical development and rhetorics at a higher level. The analysis of their effects on the listener is then a central issue.
- it is clearly a system designed for communication, where the speaker prevails.
- it is a form of interaction, which, given a question, presents various views or reactions, positive or negative. It is therefore capable of presenting contrastive views, in particular via theme elaboration, opposition between themes, and variation techniques.
- it is normalized and quite generic in the sense that it follows rules recognized by a certain group of listeners, in a certain context or historical period,
- it is figurative in the sense that it uses forms, largely symbolic, which have a clear impact on listeners with the goal to help him/her to deepen himself the initial question, with the help of the musical support.

To summarize, these points argue that music and musical structure are potential means that can support listeners internal debates about a question raised by the composer (or similar questions proper to the listener). Those means, while being symbolic, do define the main trends of a debate, its importance, its strength or even its violence. Musical elements can be extremely powerful in terms of persuasion.

3.2 The emergence of argumentation in tonal music

The period of tonal or modal music that is considered here starts roughly at the end of Renaissance or the early baroque (1600, with e.g. H. Schütz) till the dissolution of tonality into various systems in the early 20th century. In the next sections, we concentrate on the end of the 18th century and the beginning of the 19th century, where argumentation is the easiest to characterize in a few sentences. During the baroque period, ending 1750 (death of JS Bach), musical works were essentially constructed around a unique theme, which underwent a number of variations and amplifications. Even the most achieved form of the baroque period, the fuga, was constructed around a unique theme (or subject) with a response and one or more counter-subjects, but the root (or the head, in linguistic terms) remains the unique seed.

From J. Haydn works (sonatas, symphonies and string quartets) emerged the idea of the confrontation of two very different themes, with very elaborated forms of symbolic contrasts or 'fights' between them. Although themes of a given musical work had major differences in rhythm, melody and harmony, these often had a lot of musical material in common. This is comparable to opposite views in language which also share a number of prerequisites, topics, etc. This preserves the cohesion of a debate.

This was the beginning of the 'Sturm und Drang' period where the authors' feelings dominated their musical production, in contrast with the baroque period which had more general purpose considerations (religion, ceremonies, etc.). This view, typical of the German and Austrian Romanticism, was initially developed by a number of authors, including Mozart, Schubert and Beethoven. Then, this confrontation became more shallow and more complex, with the emergence of cyclic themes in the late Romantic period (J. Brahms, C. Franck, G. Mahler, etc.). In this short article, to illustrate this point, we will concentrate on a few works by L. van Beethoven, which are readily accessible, and will make our approach more clear. It is interesting to note that, given a certain question or statement, related arguments are found in a number of different works, not a single piece like a sonata of a string quartet. Argumentation was indeed related to very foundational questions that Beethoven raised and which he elaborated over several years. We then developed a different view of argumentation from JS Bach's work.

4 Beethoven and his *'Muss es sein?'* question

In Spring 1801, Beethoven, who had fragile health, felt the first effects of deafness. He also encountered a large number of personal difficulties, including financial. At that period, he seriously thought about committing suicide, as witnessed in the well-know *Heiligenstadt Testament* [4]. In this document, Beethoven raised essential questions about destiny in general, and how to behave: rebellion (and how) or acceptance. Beethoven was very close to several poets of that period (including Goethe, Schiller, Brentano and Eichendorff) who were very idealistic about society and people. Beethoven has always been a 'humanitarian idealist', following Kant's views: *the sky with stars above us, and morality in us*. All these aspects were of much interest to Beethoven and subject of several questions which were immediately transposed into his musical works.

Questions about destiny were the most crucial. These are often realized by means of very recurrent, powerful, if not brutal or violent, musical means [2]. Questions are followed by responses of very different natures, also based on quite recurrent musical means. Let us present here very briefly a few relevant examples taken from his piano sonatas.

Typical forms of questions appeared as early as 1801, in the **8th sonata** (op. 13, 'pathetic'). They all have approximately the same melodic profile and rhythm, that globally follow the natural intonation used in language. The introduction of the 8th sonata is abrupt and has the typical intonation profile of painful and doubtful questions: gradually ascending melody, minor mode, with a typical rhythm (example 1)[2]: long duration followed by a short one, repeated a number of times, in a quite obsessive way, quite typical of question intonation. After a climax (bar 4), the sequence goes on via an elaboration (bars 5 to 8), the question ends by a long descent (bars 9-10), anticipating the response. The main part of the first movement is then composed of a first theme in C minor which is very dynamic and abrupt, symbolizing fight to maintain this idealism in spite of the corruption of the society in which he had to live (and survive). It is

[2] Examples are given at the end of this paper. These score extracts are free of any reproduction rights, see http://imslp.org/)

composed of 4 bars with an ascending melody staccato followed by a 4 bars descending melody with long values. The second theme, in A flat major tone, sounds like a folk song, it introduces a form of rest or relaxation which reinforces the strength of the first theme. The question appears again twice (4 bars long each time). The movement ends abruptly, with a portion of the first theme, unfinished, leaving little space for optimism.

The next major ’step’ is **sonata nb. 23, op. 57** (Appassionata in F minor, a particularly ’dark’ tonality). It was inspired from ’The Tempest’ by Shakespeare with the opposition between young people undergoing a dramatic wreck (literally and metaphorically) and the wise Prospero. The question was again about destiny: acceptance or rebellion ? The first movement starts by this question, with a melodic profile and a rhythm close to the op.13, but darker and more violent, and with more contrast between long and short durations, making the atmosphere very distressing (example 2). This first theme (bars 1 to 11) is followed by 2 variations (12 bars each), based on specific fragments of the theme. These variations are meant to reinforce the question, outlining its main features. The second theme (from bar 36), in A flat major, has the same rhythm, but develops a luminous melody, very enthusiastic (example 10). Via this theme, Beethoven expresses his faith in positive aspects of destiny leading to positive conclusions of his life (e.g. marriage with Josephine de Brunswick, which unfortunately never happened). These two themes, which seem so antagonistic, are based on the same rhythmic structure and melody, but with reversed orientations (called mirror in musical analysis) and opposite modes (minor versus relative major): the question with a very negative colour, and this very positive second theme realizes a symbolic form of an argumentation scheme. This movement concludes in a chaotic way, with thematic material borrowed to the two themes, in the lower, ’dark’, part of the keyboard, leaving the listener with negative feelings about the initial question. The underlying, somewhat symbolic, message in this argumentation scheme is that it is difficult and painful to reach the positive conclusion: ’accept destiny’, and that numerous difficulties and failures pave this way, as suggested by the music.

Due to a lack of space, we will not discuss the **sonata nb. 29** (HammerKlavier) which is a crucial step in Beethoven’s production. The last works we want to briefly investigate here are the last two **sonatas, nb. 31 (op. 110) and 32 (op 111)**, composed around 1821. The question is raised in a very brutal way in the sonata nb. 32, with a global shape (example 3), once again, quite similar to the op. 13, which is about 20 years younger. In Sonata nb. 31, the question is raised by totally different means: a very slow and short sequence, in minor tone, a kind of *recitativo* as found in baroque cantatas, characterizing the emotional state associated with the question.

Let us now consider the responses. In the case of the sonata nb. 32, it is a set of variations in C major which roughly progresses in a way quite similar to the variations of the op. 57. From a relatively melodic and peaceful start (with incursions in A minor) the melody becomes more and more luminous. The initial rhythm of the question is now used as a support for a kind of folk dance (as in symphony nb. 7). The variations end by an extensive use of trills which have a symbolic role, that of liberation and of the evocation of Heaven and angels. The cycle of Beethoven’s sonata therefore ends by a positive conclusion, after a number of arguments supporting the argument ’acceptance’, based on different musical language means, opposed to others supporting ’revolt’ found in previous works.

Sonata nb. 31 offers the same conclusion but more gradually, and with more symbolic means. After the question, there is an *arioso dolente*, very painful.Then starts a fuga, with a positive, ascending theme (example 5). The fuga is in general felt to be the ’ideal’ musical form, thus reflecting perfection. In Beethoven’s view, a fuga always means a form of liberation of pain and oppression, which are by nature imperfect. However, the second part of this fuga becomes darker and darker, with more violence in the theme, which is now in the minor mode, in octaves, with syncopas in the counter subject suggesting anger and pain. The fuga ends in a negative mood, suggesting a negatively oriented response to the question. Then follows again an *arioso dolente*, in the same manner as the previous one. However, this arioso progressively ’climbs’ up in the upper part of the keyboard. It ends by several very luminous chords in G major. Then follows another fuga, using the same theme as the previous fuga, but in mirror (example 6, melodic intervals are symmetric to the original theme as a mirror). The fuga becomes more and more luminous, using rhythmic variation effects expressing joy and happiness (rhythmic structures are also inverses to those of the question: roughly a short duration followed by a longer one). It ends by a kind of choral that symbolizes victory. This form is very close to the last movement of the 9th symphony: the celebration of victory after several failures.

Obviously our comments are very short and informal, they nevertheless suggest the non-verbal, in a rather symbolic form, arguments for or against a certain attitude that traversed Beethoven’s life. The rhetoric symbols he used are powerful tools with an immediate impact on listeners, with strong persuasion effects.

From a rhetoric and argumentation point of view, this analysis shows:

- analogy with language forms, e.g. question intonation, stress and rhythm,
- numerous musical elements shared by the various pros and cons arguments showing strong interactions,
- alternations between major and minor modes, ascending-descending melodies, etc. to illustrate pros and cons,
- proto-rhetorical forms such as mirror effects, which suggest opposition or resistance to the initial figure,
- use of highly symbolic forms such as fugas, trills, etc.

5 Symbolism and argumentation in JS Bach C Minor Passacaglia: ’Quaerendo invenietis’

The late works of JS Bach are extremely symbolic and follow very strict rhetoric schemas. The citation above from Bach’s Musical Offer (quaerendo invenietis: investigate and you’ll understand, an analogy to the Biblical ’ask and you’ll receive’) indicates the needs to explore the structure of his works to understand the topics addressed and how they are addressed. This late baroque period was very much influenced by works in philosophy and mathematics. Bach made an extensive use the symbolism of numbers (there is a large literature on this topic). Leibnitz in particular (he also lived in Leipzig) had an important influence on Bach.

Let us investigate here a rather accessible work, the Passacaglia in C. Minor, BWV 582 (1716), for organ, which was written when Bach was 31 years old [8, 10]. It had a major influence on composers from the 19th and 20th centuries. The work is based on a powerful theme, used as an ostinato (repeated theme), mainly at the bass, of 3/4 times over 8 bars. It is repeated a total of 21 times (a symbolic figure: 3 x 7: here trinity and perfection). The three voices above the bass are counterpoint elaborations that embody the messages and here the arguments Bach wants to push forward in this work. Each of these 21 sequences is associated with a very precise symbolism that forms a coherent network of signs. Let us very briefly present them below.

When analyzing the rhetorical structure of this work and the symbols in the various sequences, it becomes clear that there is a strong symbolic dimension associated with the structure of the work:
- numerical proportions in melodies and rhythms: 3 for the Trinity, 7 for the seven days of the creation, etc.
- use of symbolic forms in melodies, such as the cross, formed from the notes: B,A,C,A (or equivalently: B A C H, which is Bach's signature),
- use in each sequence of Lutheran choral fragments borrowed from the Orgelbuchlein (e.g. *Nun komm' der Heiden Heiland, Christ lag im Todesbanden*) that make more clear the underlying contents of each sequence.

Radulescu [5] shows that this work is an argumentation in favor of the necessity of crucifixion. See also: http://www.davidrumsey.ch/Passacaglia.pdf. The 21 sequences reflect that debate, the conclusion being acceptance.

The organist MC Alain divides the 21 sequences of this work into groups of 3, each contributing to an organized form of debate, related to crucifixion and redemption, and each with a specific melodic symbolism and a specific choral. The different facets of the debate, each sequence, is an argument, the whole piece being the argumentation leading to the acceptance of crucifixion. The global rhetorical architecture of the work follows the classical Renaissance rhetorics, with two views which are apparently in opposition, but which can be merged into a conclusion. Very briefly:

- **Expositio and Elaboratio**: the three first sequences introduce the atmosphere and the general problematics: perversion of humanity, very introspective counterpoint of the melody, which is gradually descending with numerous syncopas suggesting pain; the organ registration must be sober.
- **Exordium**: introduces the facets of the debate (God is unhappy with humans), sequences 4 to 6, descending intervals in canonical form in sequence 5 (example 7), or voices going in opposite directions. The contents is suggested by choral short extracts and for sequence 6 by a rather suggestive rhythm at the bass (2 anacroustic fourths followed by a quaver with an accent). XXXXX
- **Propositio**: general statement (humanity needs sacrifice), necessity of crucifixion becomes clear with very recurrent symbols, giving a strong persuasion force to these statements. This is realized in sequences 7 to 10, (with the use of groups of 4 notes: examples 8 (a) and (b) symbolizing the cross, and the inclusion of choral fragments in quarter note), reference to God's son (choral: *Vom Himmel kam der Engel Schar*).
- **Confutatio**: counterarguments: sequences 11 to 13, sequence 12 is in general analyzed as the climax of the work, where the theme progressively disaggregates at the bass and climbs to the upper part (sequence 13), in contrast with the other sequences, as a large complain (example 9). These three sequences express doubts (theme disaggregation) and anger (dramatic use of the theme on the upper part, no pedal), they therefore constitute a kind of schema for a symbolic counter-argument, furthermore a contrast is introduced by the lack of pedal which was so far present, suggesting a very stable athmosphere.
- **Confirmatio**: going beyond the two views for or against, reinforcement of the initial proposal, crucifixion is accepted. Sequences 14 to 16 contain intertwined melodic fragments from the two previous rhetoric structures.
- **Peroratio**: conclusion of argumentation, crucifixion entails redemption, sequences 17 to 19, with a new dynamics in rhythm symbolizing happiness and a reference to Easter period chorals.
- **Final conclusion (coda)**: redemption and celebration, sequences 19 to 21, using ascending intervals, organ registration must be brilliant.

It is clear that Bach's music is not as direct as Beethoven's: symbols are deeper, more abstract and more complex to perceive. However, at his period, chorals and their main melodic characteristics were known by almost everyone, facilitating understanding. From reports and comments of this period, it seems that understanding such a work was accessible to a wide audience.

6 Conclusion

In this short paper we have informally presented some very preliminary aspects suggesting abstract forms of argumentation in western tonal music. These aspects remain largely symbolic or psychological, but this is a constitutive part of argumentation. However, music is at least as complex as language, using more abstract means. Obviously there is always a rational, almost textual, contents which is implicit, and which can be retrieved in the composer's life (Beethoven) or in the liturgy (Bach).

In this paper, we have presented two composers, with very different profiles and culture. We have also attempted to show how musical themes are treated and transformed using a model based on alternations, but with very different processes. We also aim at analysing the different musical means deployed by composers in terms of pre- or proto-rhetoric forms (e.g. mirror, theme fragment amplification, fugatos), as we could have in language proto rhetoric relations. We feel some form of **argumentative signature** could be defined.

There are many other composers worth considering to investigate argumentation, in particular from the 19th century. If some of them are rather flat in terms of internal debate, there are other composers which are of much interest. Besides R. Wagner, a particularly interesting case is R. Schumann, who himself created two characters, Eusebius (quiet, dreaming, pessimistic) and Florestan (noisy, optimistic, unpredictable, etc.), to characterize his personal debates. In his work, these two characters correspond to different musical moods, types of melodies, rhythms, etc. His questioning concerned several aspects of his artistic creation, ending in a suicide in the river Rhein, not at night, but on a gloomy, dull morning, as 'predicted' in his *Gesange der Frühe*.

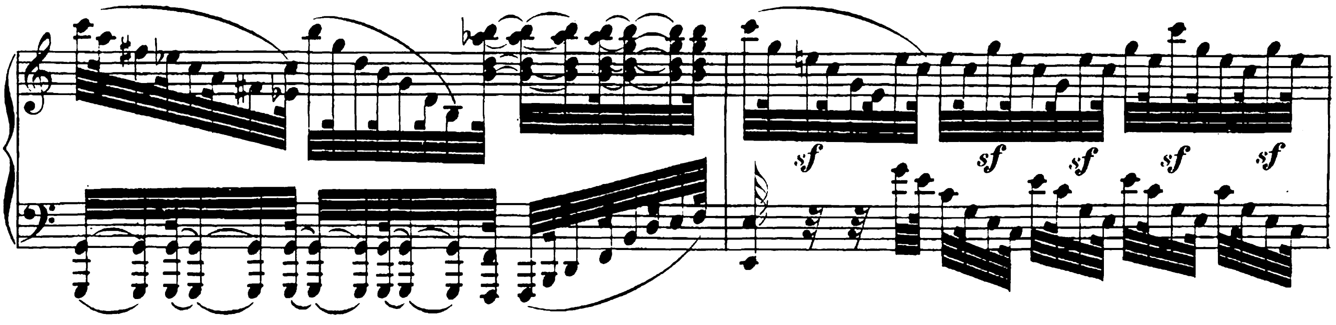
REFERENCES

[1] J. Katz, D. Pesetzky, The Identity Thesis for Language and Music, research report, MIT, 2009.
[2] H.A. Harding, Analysis of Form in Beethoven's Sonatas, General Books, 2010.
[3] F. Lerdahl, R. Jackendoff, A Generative Theory of Tonal Music, MIT Press, 1996.
[4] J. Massin, B. Massin, *Beethoven*, Fayard, 1967.
[5] Martin Radulescu. On the form of Johann Sebastian Bach's Passacaglia in C minor, The Organ Yearbook, 1980, 95103.
[6] R. Jackendoff, Parallel and non parallel between language and music, *Music Perception*, vol. 26-3, 2009.
[7] R. Jackendoff, F. Lerdahl, The capacity for music: What is it, and whats special about it?, *Cognition*, 2005.
[8] P. F. Williams, *The Organ Music of J. S. Bach*, Cambridge University Press, 2003.
[9] D. Walton, C. Reed, F. Macagno, *Argumentation Schemes*, Cambridge University Press, 2008.
[10] M. Dupré, *Integral of Bach organ works*, 12 annotated volumes, Leduc, 1980.

Example 1

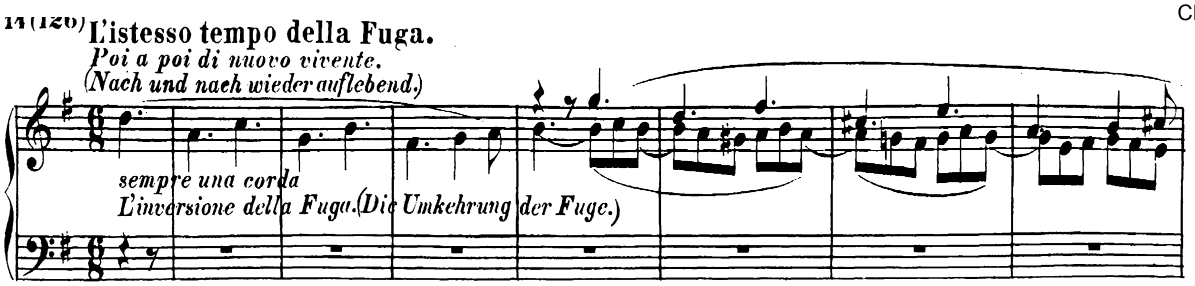
Example 2

Example 3



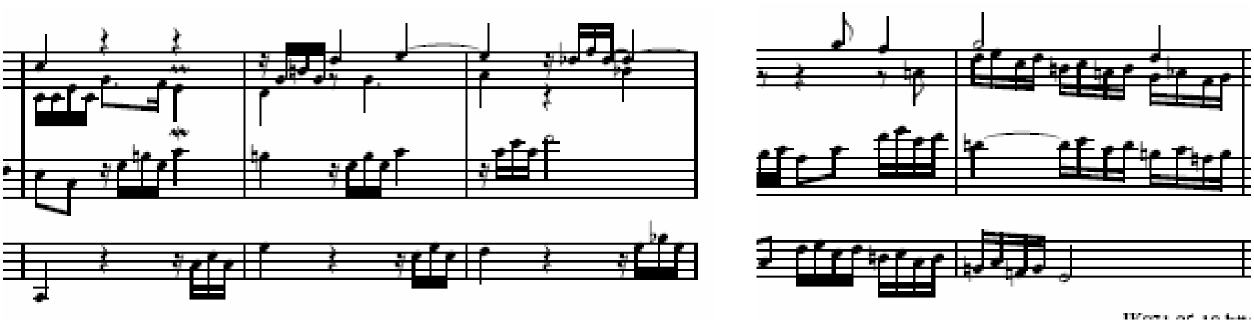
Example 4

Example 5

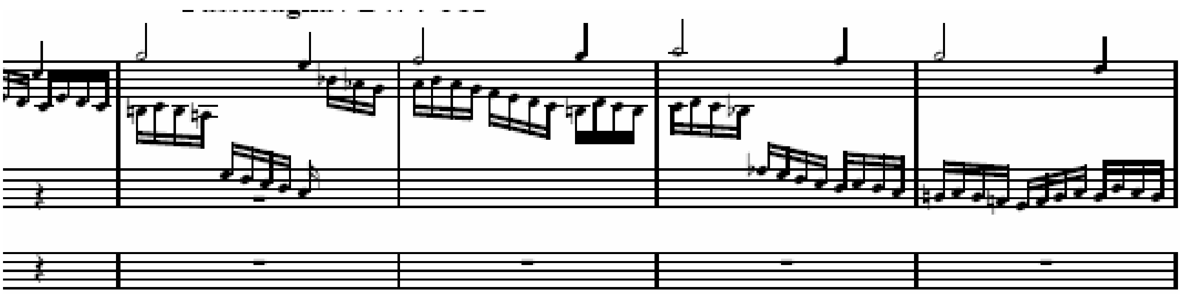


Example 6

Example 7



Example 8 (a) and (b)



Example 9

Example 10

An analysis of critical-link semantics with variable degrees of justification

Bin Wei[1] and **Henry Prakken**[2]

Abstract. The main aim of this paper is to critically examine Pollock's critical-link semantics with variable degrees of justification. We point out some possibly counterintuitive consequences of Pollock's definition of degrees of justification and propose a modified definition which avoids these consequences. We then modify the *ASPIC*$^+$ framework to allow for variable degrees of justification and then apply our modified way to compute these degrees.

1 Introduction

In most current AI approaches on modeling Argumentation, the justification status of arguments and conclusions is an all-or-nothing affair, but in many realistic applications, such as legal reasoning about evidence or other applications of epistemic reasoning, it is natural to regard them as justified to variable degrees. Pollock moddelled this in his so-called critical-link semantics in [1] and [2].

Pollock introduced variable justification degrees to account for the so-called "diminishing" effect of attempted defeaters that are weaker than their target. In such cases Pollock wanted to model that the attempted defeaters can still weaken the degree of justification of their target. The present paper aims to contribute to such a study by critically examining Pollock's proposal. In particular, we will argue that Pollock's approach in some cases gives counterintuitive outcomes, then modify his account in a way that avoids these outcomes. At the end, we will briefly discuss how Pollock's ideas and our modifications can be incorporated in the *ASPIC*$^+$ framework for structured argumentation recently proposed by [3].

This paper is organized as follows. In Section 2, we first summarize Pollock's semantics. In Section 3, we then discuss some arguably counterintuitive outcomes, present our revised definitions and show that they avoid these outcomes. In section 4, we discuss how to transfer the revised semantic into *ASPIC*$^+$ framework. Finally, we conclude in Section 5.

2 Semantics

In this section we present Pollock's critical-link semantics with variable degrees of justification, preceded by a brief overview of his [4] multiple-assignment semantics.

2.1 Basic features

In Pollock's account of defeasible reasoning, reasoning proceeds from a knowledge base of classical-logic formulas by chaining reasons into inference graphs, where all reasons are either deductive or defeasible. Only applications of defeasible reasons can be defeated, and there are two kinds of defeaters: *rebutting* defeaters attack the conclusion of a defeasible inference by favoring a conflicting conclusion, while *undercutting* defeaters attack the defeasible inference itself, without favouring a conflicting conclusion.

More precisely, Pollock assumes as given a knowledge base of first-order formulas and two sets of deductive and defeasible reasons, which technically are inference rules. Pollock then considers arguments, which are sequences of *argument lines*. The strength of an element φ of the knowlege base is below written as $\delta(\varphi)$ while the strength of a reason r will be written as $\rho(r)$.

Definition 2.1. *An argument line is a tuple (φ, r, L, s), where φ is a proposition, r is the reason applied to infer φ, L is the set of preceding lines from which φ is inferred, and s is the line's strength*[3]*.*

Definition 2.2. *An argument line (φ, r, L, s) defeats an argument line (φ', r', L', s') iff r' is a defeasible rule, and $s \geq s'$, and $\varphi = \neg\varphi'$ or $\varphi = \neg r'$ (here $\neg r$ is shorthand for saying that the antecedents of rule r do not support its consequent).*

Definition 2.3. *For any argument line $l = (\varphi, r, L, s)$ (where $L = \{l_1, \ldots, l_n\}$) the strength $s(l)$ is inductively defined as follows:*

- *If l takes φ from the knowledge base, then $s(l) = \delta(\varphi)$.*
- *Otherwise, $s(l) = \min\{\rho(r), s(l_1), \ldots, s(l_n)\}$.*

With respect to accrual of arguments for the same conclusion, Pollock proposed that if we have two separate undefeated arguments for a conclusion, the degree of justification for the conclusion is the maximum of the strengths of the two arguments.

2.2 Multiple assignment semantic

In [4] Pollock considers inference graphs, where the nodes represent the propositions inferred from which they are inferred, support-links tie nodes to the nodes, and defeat-links indicate defeat relations between nodes. These links relate their roots to their targets. The root of a defeat-link is a singe node, while the root of a support-link is a set of nodes. He then proposes a labeling approach to define the justification status of nodes and propositions.

Definition 2.4. *A node of the inference-graph is* initial *iff its node-basis and list of node-defeaters is empty, where*

- *The* node-basis *of a node is the set of roots of its support links.*

[1] The Institute of Logic and Cognition, Sun Yat-Sen University, Guangzhou, China. email: srsysj@gmail.com

[2] Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands

[3] Below the strength of argument line l will sometimes be written as $s(l)$.

- *The* node-defeaters *are the roots of the defeat links having the node as their target.*

Definition 2.5. *An assignment σ of defeated and undefeated to a subset of the nodes of an inference-graph is a* partial status assignment *iff:*

- *σ assigns* undefeated *to any initial node;*
- *σ assigns* undefeated *to a non-initial node α iff σ assigns* undefeated *to all the members of the node-basis of α and σ assigns* defeated *to all node-defeaters of α ;*
- *σ assigns* defeated *to a non-initial node α iff either σ assigns* defeated *to a member of the node-basis of α or σ assigns* undefeated *to a node-defeater of α.*

Definition 2.6. *Assignment σ is a* status assignment *iff σ is a partial status assignment and σ is not properly contained in any other partial status assignment.*

Definition 2.7. *A node α of an inference graph is* undefeated *iff every status assignment to the inference graph assigns* undefeated *to α; otherwise α is* defeated*.*

2.3 Critical-link semantics with variable degrees of justification

The core idea of critical-link semantics [1, 2] is to build new inference-graphs as subgraphs of the original inference graph and assign various statuses to initial nodes in different cases. This idea is formally defined as follows:

Definition 2.8. *An* inference/defeat-path *from a node φ to a node θ is a sequence of support-links and defeat-links such that (1) φ is a root of the first link in the path; (2) θ is the target of the last link in the path; (3) the root of each link after the first member of the path is the target of the preceding link; (4) the path does not contain an internal loop, i.e., no two links in the path have the same target.*

Definition 2.9. *A node θ of an inference graph is* φ-dependent *iff there is an inference/defeat-path from φ to θ.*

Definition 2.10. *A* circular inference/defeat-path *from a node φ to itself is an inference/defeat-path from φ to φ via a defeater of φ.*

Definition 2.11. *A defeat-link is* φ-critical *iff it is a member of some minimal set of defeat-links such that removing all the defeat-links in the set suffices to cut all the circular inference/defeat-paths from φ to φ.*

Definition 2.12. *If φ is a node of an inference graph G, then G_φ is the inference graph that results from (1) deleting all φ-critical defeat-links from G and (2) making all members of the node-basis of φ initial nodes in G_φ and (3) making all φ-independent nodes initial-nodes in G_φ with stipulated defeat-statuses the same as their defeat-statuses in G.*

We next discuss how Pollock uses his critical-link semantics to define variable degrees of justification. A main motivation of the idea that propositions should have variable degrees of justification is Pollock' notion of a *diminisher*. A diminisher is a defeater of a node that is weaker than its target, which is able to diminish the *degree* of justification of its target.

For the sake of the mathematics of diminishers, Pollock proposed that there exists a function $\diamond$[4] such that given two argument lines that rebut one another, if their strengths are x and y, the degree of justification for the conclusion of the former is $x \diamond y$, while the degree of justification for conclusion of y is $y \diamond x$. He assumed that "the degree of justification can be measured using real numbers, possibly augmented with ∞, i.e., 'the extended real numbers'. More precisely, the degrees of justification fall in some interval $[o, \theta]$, where $0 \leq o \leq \theta \leq \infty$. o corresponds to no justification, and θ to perfect justification, presumably only possible for necessary truths.". Then Pollock defined mathematical properties of $\diamond$ as follows:

Definition 2.13. *[Mathematics of $\diamond$]*
(A1) $\diamond$ is continuous on the interval $[o, \theta]$.
(A2) If $\theta > \alpha > \beta > o$, then $\alpha > \alpha \diamond \beta > o$.
(A3) If $\theta > \alpha > \beta > \gamma > o$, then $\alpha \diamond \beta < \alpha \diamond \gamma$ and $\alpha \diamond \gamma < \beta \diamond \gamma$.
(A4) If $\theta \geq \alpha \geq \beta > o$, then $\beta \diamond \alpha = o$.
(A5) If $\theta \geq \alpha > o$, then $\alpha \diamond o = \alpha$.
(A6) If $\theta > \alpha$ and β and γ are in $[o, \theta]$, then $(\alpha \diamond \beta) \diamond \gamma = (\alpha \diamond \gamma) \diamond \beta$.

Pollock proved that if (A1) $-$ (A6) hold, then $\diamond$ has a very simple representation as follows:

Definition 2.14. *[Representation of $\sim$]*

$$x \sim y = \begin{cases} x - y & \text{if } y < x < \infty \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Definition 2.15. *[Computation of degree of justification]*
(DJ) If P is inferred from the basis $\{B_1, \ldots, B_n\}$ in an inference-graph G in accordance with a reason of strength ρ, $D_1, \ldots, D_k$ are the P-independent defeaters for P , and $D_{k+1}, \ldots, D_m$ are the P-dependent defeaters of P, then $J(P, G) = \min\{\rho, J(B_1, G), \ldots, J(B_n, G)\} \sim \big[\max\{J(D_1, G), \ldots, J(D_k, G)\} + \max\{J(D_{k+1}, G_P), \ldots, J(D_m, G_P)\}\big]$.

DJ is a computation for "collaborative defeat", where the nodes are defeated by both node-dependent and node-independent defeaters.

3 Problem cases and modifications

In this section, we discuss some possible problems of Pollock's critical-link semantics with variable degrees of justification, by analyzing some problem cases.

3.1 Problem case on diminishers

The first problem concerns some arguably counter-intuitive consequences of the mathematical properties and representation of the function $\diamond$. We present an example and discuss why the outcomes may be counter-intuitive, and then modify some properties of $\diamond$ and choose another definition for $\sim$ to represent $\diamond$.

Consider rebutting defeaters in Figure 1. Let P be "Jones says that it is not raining", R be "Smith says that it is raining", and Q be "It is raining". Let us first assume that Smith and Jones as equally reliable. Then according to Pollock both Q and $\neg Q$ should be defeated. If we apply Definition 2.15 and again assume that the degrees of justification of the initial nodes are at least as great as the strengths of reasons, then we have $J(Q, G) = J(\neg Q, G) = 0$. Assume next that Smith is much more reliable than Jones: then Q defeats $\neg Q$ while $\neg Q$ diminishes Q: by Definition 2.15 we have $J(\neg Q, G) = 0$ and $J(Q, G) = J(R, G) \sim J(P, G) > 0$.

The arguably counter-intuitive consequence is that node $\neg Q$ has in both cases the same degree of justification, namely, 0, while yet in the

[4] Pollock added the mathematical analysis in his extended version, http://oscarhome.soc-sci.arizona.edu/ftp/PAPERS/Degrees.pdf.

second case the degree of justification of Q is higher than in the first case. Thus intuitively, although node $\neg Q$ is in the first case not accepted, it is still much more reliable than in the second case. Thus the degrees of justification of nodes in cases of symmetric defeat should be greater than the ones in cases of asymmetric defeat. Moreover, the first case is similar to "zombie arguments"[5]: although the arguments are defeated, they can still affect another arguments. In other words, the node $\neg Q$ in the first case still has ability to attack or support other nodes, but the node $\neg Q$ in the second case does not. So it is necessary to make a difference between the degrees of justification of nodes in these two cases.

3.2 Problem case on "presumptive defeat"

The previous point can be further developed in a discussion of ambiguity blocking vs. ambiguity propagating (by Pollock called "presumptive defeat"in [1]). Consider again Figure 1 but let now Q stand for "Rain was predicted by the morning weather forecast", P for "Jones says that no rain was predicted by the morning weather forecast", R for "Smith says that rain was predicted by the morning weather forecast", S for "It will rain"and A for "rain was predicted by the afternoon weather forecast". Suppose again that the reason strengths are at least as great as those of the initial nodes and suppose that P and R are equally strong. Then according to Pollock's new approach the degree of justification of all of Q, $\neg Q$ and $\neg S$ equals 0, so that $\neg S$ cannot diminish or defeat S. However, according to Section 3.1 the degrees of justification of Q and $\neg Q$ should be greater than 0, and this has the consequence that $\neg S$ potentially has the force to diminish or even defeat S.

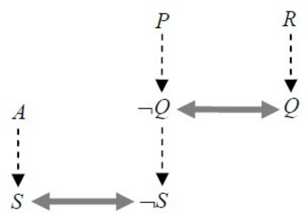


Figure 1. Presumptive defeat

3.3 Problem case on undercutters

Next we discuss a problem of the computation principle *DJ* by arguing that it gives an unnatural treatment of the effect of undercutters on the degree of justification of their target. Consider an inference graph with undercutter, let P be "Jones says that it is raining" and Q be "It is raining", R be "Smith says that John always lies" and $P \otimes Q$ be "John is lying" means "P does not guarantee Q". Note that node $P \otimes Q$ attacks the connection between node P and node Q, so the strength of node $P \otimes Q$ should arguably directly weaken the strength of the reason from P to Q and only indirectly weaken the strength of node Q. In other words, the strength of an undercutting node should be in comparison with the strength of the reason it undercuts rather than with the strength of the node it attacks. However, in Pollock's definitions this is not the case.

3.4 Modified definition of representation

In his final paper [6], Pollock reconsidered the problem of degrees of justification. He measured degrees of justification using numbers in the interval $[0, 1]$, for which reason we henceforth choose the scale as $[0, 1]$. From assumptions (A2) and (A4) it's clear to show that Pollock meant to design the function to capture the diminishers diminish nodes without completely defeating and diminishers diminish nodes with completely defeating. However, some assumptions of mathematical properties of operator are counter-intuitive and should be revised in order to avoid the above problems.

Firstly, according to the above analysis on diminisher and "presumptive defeat". Assumption $(A4)$ should be modified as follows:

(A4) If $\theta > \alpha > \beta > o$, then $\beta \diamond \alpha = o$.

(A4') If $\theta > \alpha = \beta \geq o$, then $\beta \diamond \alpha \geq o$.

These two revised assumptions that the degrees of justification of nodes in defeat cycles should be greater than 0.

Secondly, according to the above analysis of diminishers, the degree of justification of diminished node reduces to real number 0 when the strength of the diminishing node with completing defeating is approaching to the strength of the diminished node. However, the degree of justification of the diminished node would be definitely greater than 0 in accordance with (A4') if the strengths of the rebutting defeaters are equal. Therefore, the representation is not continuous on the whole interval $[0, 1]$, since any point (x_0, y_0) that satisfies $x_0 = y_0$ would be a discontinuous point. But Pollock wanted that diminishing nodes without completely defeating and diminishing nodes with completely defeating are, respectively, continuous. Therefore, we use $f(x, y)$ to present a diminishing node with degree y that completely defeats a diminished node with degree x, and use $g(x, y)$ to present a diminishing node with degree y that does not completely defeat a diminished node with degree x. We replace assumption (A1) by saying that $f(x, y)$ and $g(x, y)$ are continuous.

Thirdly, the degree of justification for a diminished node should be the strength of this node decremented by an amount determined by the strength of the diminishing node. Moreover, the strength of a node as conclusion is determined by the strength of its reason and the strength of its node as premise. Rebutting defeaters or undercutting defeaters can both act as diminishers but their influences on diminished nodes are different. Undercutting defeaters weaken the strength of the reason they attack, while rebutting defeaters directly weaken the strength of the node as conclusion. Therefore, the order in which undercutting defeaters and rebutting defeaters as diminishers are applied to an argument makes a difference to the degrees of justification, and this in turn means that A(6) is invalid.

In sum, our analysis in Sections 3.1-3.3 makes that assumption (A4) must be modified while assumptions (A1) and (A6) cannot hold. We now define a new representation $\sim$ for operator $\diamond$, which matches the above-revised assumptions. Let us define:

$$x \sim y = \begin{cases} x(1-y) & \text{if } y \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

It's easy to prove that the new function satisfies the revisions of Pollock's conditions:

(A1) $f(x, y) = x(1-y)$ and $g(x, y) = 0$ are continuous on the interval $[0, 1]$

(A2) If $1 > x > y > 0$, then $x > x \sim y > 0$.

(A3) If $1 > x > y > z > 0$, then $x \sim y < x \sim z$ and $x \sim z > y \sim z$

(A4) If $1 > x > y > o$, then $y \sim x = 0$.

(A4') If $1 > x = y \geq 0$ then $x \sim y \geq 0$

(A5) If $1 \geq x > 0$, then $x \sim 0 = x$

3.5 Modified definition of variable degrees of justification

The revised idea for the problem case of undercutters is that the degree of justification of node P equals the minimum of the strength of reason after being diminished and the degrees of justification of its premises. Then the computation for nodes not in a circular path can be modified as follows: If P has P-independent defeaters $D_1,\ldots,D_k$ in G and has no P-dependent defeaters, then $J(P,G) = \min\{(\rho \sim \max\{J(D_1,G),\ldots,J(D_k,G)\}), J(B_1,G),\ldots,J(B_n,G)\}$.

We next discuss the case where a node P is defeated by both P-dependent defeaters and P-independent defeaters. We propose that these two kinds of defeaters can unite to defeat node P with a double counting, but computing it with P-independent defeaters firstly and then continue to compute it with P-dependent defeaters. The final computation can be modified as follows:

Definition 3.1. *[Modified Computation]*
If P is inferred from the basis $\{B_1,\ldots,B_n\}$ in an inference-graph G in accordance with a reason of strength ρ, $D_1,\ldots,D_k$ are the P-independent defeaters for P, and $D_{k+1},\ldots,D_m$ are the P-dependent defeaters of P, then $J(P,G) = \min\big\{(\rho \sim \max\{J(D_1,G),\ldots,J(D_k,G)\}), J(B_1,G),\ldots,J(B_n,G)\big\} \sim \max\{J(D_{k+1},G_P),\ldots,J(D_m,G_P)\}$

For instance, in Figure 2, node $\neg S$ is S-dependent, node S is $\neg S$-dependent and node $Q \otimes S$ is S-independent. Let $J(P,G) = 0.15$, $J(Q,S) = J(R,G) = 0.8$ and the reasons are equally strong: $\rho = 0.9$. Then $J(Q \otimes S,G) = 0.8$, $\rho \sim J(Q \otimes S,G) = 0.18$, $J(\neg S,G_S) = 0.15$, $J(S,G) = \min\{\rho \sim J(Q \otimes S,G), J(Q,G)\} \sim J(\neg S,G_S) = 0.18 \sim 0.15 = 0.153$, and $J(S,G_S) = 0.18$, $J(\neg S,G) = \min\{\rho, J(P,G)\} \sim J(S,G_{\neg S}) = 0.15 \sim 0.18 = 0$.

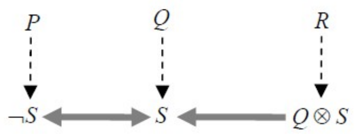


Figure 2. Inference graphs with collaborative defeaters

3.6 Solution to the problem cases

We now show that the new definition avoids the arguably counter-intuitive outcomes we described above. We do this by analyzing the example of presumptive defeat, which includes the problem case of diminishers. Consider again the example in Figure 1. In the multiple-assignment semantics in [4], $\neg Q$ has the ability to support $\neg S$ if $\neg Q$ is assigned undefeated in the partial status assignment and $\neg Q$ has no ability to support S if $\neg Q$ is assigned defeated in the other partial status assignment. With our new definition of $\sim$ the outcome is different. For simplicity, we again assume that the strengths of reasons are at least as great as the degrees of justification of the initial node. Then the computation of $J(\neg S,G)$ can be concluded as follows: $J(\neg S,G) = \min\{\rho, J(\neg Q,G)\} \sim J(S,G_{\neg S}) = J(\neg Q,G) \sim J(A,G) = \big(J(P,G) \sim J(R,G)\big) \sim J(A,G)$.

We discuss the possible degrees of justification of $\neg Q$ and $\neg S$. $\neg Q$ has ability to support $\neg S$ iff $J(P,G) \geq J(R,G)$. Hence, $J(\neg S,G) > 0$ iff $\neg Q$ has ability to support $\neg S$ and $J(P,G)(1 - J(R,G)) \geq J(A,G)$. Otherwise, $J(\neg S,G) = 0$. For instance, let $J(P,G) = J(R,G) = 0.8$, $J(A,G) = 0.1$ and the reason-strengths are equally strong: $\rho = 0.9$, then $J(\neg Q,G) = \min\{\rho, J(P,G)\} \sim J(Q,G_{\neg Q}) = J(P,G) \sim J(R,G) = 0.16$; $J(Q,G) = \min\{\rho, J(R,G)\} \sim J(\neg Q,G_Q) = J(R,G) \sim J(P,G) = 0.16$; $J(\neg S,G) = \min\{\rho, J(\neg Q,G)\} \sim J(S,G_{\neg S}) = J(\neg Q,G) \sim J(A,G) = 0.144$.

Apparently, $\neg Q$ has the power to support $\neg S$ and $\neg S$ therefore has the ability to defeat or support another nodes. Moreover, if we let $J(P,G) < J(R,G)$ or $J(P,G)(1 - J(R,G)) < J(A,G)$, the justification of $\neg S$ equals 0.

4 Variable degrees of justification in the *ASPIC*$^+$ framework

The idea of critical-link semantics with variable degrees of justification is a general theory and can be applied in other argumentation formalisms as well. We will discuss the computation of degrees of justification combined with *ASPIC*$^+$, using the new notion of an argument graph. We regard the degree of justification of an argument[5] as the variable degree for accepting or rejecting the argument from a cognitive perspective. We next give some new definitions that are useful in our modification associated with *ASPIC*$^+$.

Definition 4.1. *[Argument strength] $\mathcal{V}$ is a function to evaluate the strength of an argument with conditions as follows:*

- *if $A \in K$, then $\mathcal{V}(A) = \eta(A)$, where η is a function that assigns the degrees of acceptability of the premises in an argument, modeled as $\eta(A) : 2^{Prem(A)} \rightarrow [0,1]$.*
- *if A is the form $A_1,\ldots,A_n \rightarrow/\Rightarrow \varphi$, then $\mathcal{V}(A) = \min\big\{\mathcal{V}(A_1),\ldots,\mathcal{V}(A_n), \nu\big(\texttt{Conc}(A_1),\ldots,\texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi\big)\big\}$, where ν is a function assigns the degree of support from antecedent to consequent in a strict or defeasible inference, modeled as: $\nu(\delta) : \delta \rightarrow [0,1)$, where $\delta \in \mathcal{R}_s$ and $\nu(\delta) = 1$, where $\delta \in \mathcal{R}_d$.*

Definition 4.2. *[Maximal proper subargument] Argument A is a* maximal proper subargument *of B iff A is a subargument of B and there does not exist any proper subargument C of B such that A is a proper subarugment of C.*

Definition 4.3. *[Direct attacking] Argument A* directly attacks *argument B iff A rebuts or undercuts B on B; otherwise A indirectly attacks B.*

Definition 4.4. *An* argument graph *G is a labeled, finite, directed, bipartite graph, consisting of argument nodes and attacking links indicating attacking relationships between argument nodes and proper subargument links indicating connecting subargument relationships between an argument and its proper superarguments.*

The attacking links relate their roots to their targets and the root of an attacking link is an attacker in the graph, while the proper subargument links relate their roots to their targets and the root is the proper subargument of its target or the target is the proper superargument of its root in graph. In the diagrams of argument graphs, argument are

[5] We assume that the degree of justification of one argument equals the degree of justification of its conclusion.

displayed as dots, attacking links are indicated using ordinary arrowheads, while proper subargument links are indicated using closed-dot arrowheads. The initial arguments in G can be defined as follows:

Definition 4.5. *An argument is initial in G iff it is not the target of any attacking link or proper subargument link.*

Consider and Pollock's inference graph in Figure 1. We assume arguments in *ASPIC*$^+$ framework as $B : B_1 \Rightarrow \neg S$; $B_1 : B_2 \Rightarrow \neg Q$; $B_2 : P$; $C : C_1 \Rightarrow Q$; $C_1 : R$; $D : D_1 \Rightarrow S$; $D_1 : A$. We show the arguments in Figure 3. Note that C directly rebuts B_1 and indirectly rebuts B, B directly rebuts D. Moreover, nodes B_2, C_1 and D_1 are initial arguments.

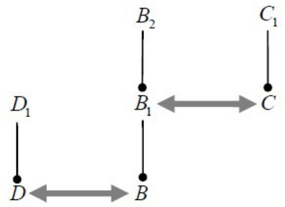


Figure 3. Argument graph

Definition 4.6. *An $argument\ path$ $P(A, B)$ from argument A to argument B in graph G is a sequence of attacking links and proper subargument links $\langle L_1, \ldots, L_n \rangle$, such that*

1. *Argument A is the initial argument that there is no argument in graph G attacks A;*
2. *there exists arguments $B_1, \ldots, B_{n-1}$, such that $L_1 = (A, B_1)$, $L_{i+1} = (B_i, B_{i+1})$, and $L_n = (B_{n-1}, B)$, where (A, B) means the attack link or proper subargument link from A to B.*

Next we will make our approach simpler than Pollock's by defining the notions of a basic set and its extension instead of the notions of node-dependent and node-critical links.

Definition 4.7. *The notions of basic set and critical extension can be defined as follows:*

1. *A set of attack links is a* basic set *of argument A in graph G iff removing all members of the set suffices to cut all cycles from A to A.*
2. *A set of attack links is a* critical extension *of argument A in graph G iff it is a minimum basic set of argument of A in graph G.*

Proposition 1. *For any argument A in a circular path, there exists at least one basic set of A.*

Proposition 2. *For any attack link L in a circular path P, there exists at least one critical extension containing L.*

Corollary 1. *If an attack link does not occur on any circular path, then it does not belong to any critical extension.*

Definition 4.8. *Given a graph G, the new graph G_A is the argument-graph that results from removing all members in all critical extensions in graph G and making all arguments $B_1, \ldots, B_n$ which are not in a defeat cycle initial with $J(B_i, G_A) = J(B_i, G)$.*

Definition 4.9. *[Justification computation]*

1. *If A is initial in G, then $J(A, G) = \mathcal{V}(A, G)$.*
2. *If A is initial in G_A, and $B_1, \ldots, B_n$ are direct rebutters of A or undermining attackers in cycles from A to A, then $J(A, G) = \mathcal{V}(A, G) \sim \max\{J(B_1, G_A), \ldots, J(B_n, G_A)\}$.*
3. *If A is not initial in G, and $A_1, \ldots, A_n$ are the maximal proper subarguments of argument A, and ρ is the strength of $Toprule(A)$, $B_1, \ldots, B_i$ are direct undercuters of A and $B_{i+1}, \ldots, B_m$ are direct rebutters of A or undermining attackers in cycles from A to A, then $J(A, G) = \min\{(\rho \sim \max\{J(B_1, G), \ldots, J(B_i, G)\}), J(A_1, G), \ldots, J(A_n, G)\} \sim \max\{J(B_{i+1}, G_A), \ldots, J(B_m, G_A)\}$.*

We define $x \sim y = x(1 - y)$, if $y \leq x < 1$, otherwise, $x \sim y = 0$ and $\max\{\emptyset\} = 0$. The computation is for argument attacked both by direct undercutters and direct rebutters or underminers in cycles. It unites and double counts the computation for arguments only attacked by direct undercutters and the computation for arguments only attacked by direct rebutters or underminers in cycles.

Finally, we illustrate the new definition by computing the degree of justification of argument B in Figure 3. Let $J(B_2, G) = 0.8$, $J(C_1, G) = 0.8$, $J(D_1, G) = 0.1$ and the reasons are equally strong: $\rho = 0.9$. It is clear that C directly rebuts B_1, then from (DJ), it follows $J(B_1, G) = \min\{\rho, J(B_2, G)\} \sim J(C, G_{B_1}) = J(B_2, G) \sim J(C_1, G) = 0.16$; we also have B directly rebuts D, then from (DJ), it follows $J(B, G) = \min\{\rho, J(B_1, G)\} \sim J(C, G_B) = J(B_1, G) \sim J(D_1, G) = 0.144$; Similar, $J(B, G_D) = 0.16$; $J(D, G) = \min\{\rho, J(D_1, G)\} \sim J(B, G_D) = J(D_1, G) \sim J(B, G_D) = 0$.

5 Conclusion

In this paper we studied the modelling of variable degrees of justification in argumentation. We pointed out some arguably counter-intuitive consequences of Pollock's critical-link semantics with variable degrees of justification and then presented some modifications that avoid these outcomes. Moreover, to illustrate the generality of Pollock' approach and our modifications, we also discussed how they can be combined with the *ASPIC*$^+$ framework. In future work we aim to investigate the properties of our definitions and to study their application to realistic examples, including problems of legal reasoning with evidence.

ACKNOWLEDGEMENTS

Bin Wei was supported by the Chinese MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (12JJD720006).

REFERENCES

[1] J. L. Pollock. Defeasible reasoning with variable degrees of justification, *Artificial Intelligence*, 133: 233-282, 2002.
[2] J. L. Pollock. A recursive semantics for defeasible reasoning, *Argumentation in Artificial Intelligence*, eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 173-197, 2009.
[3] H. Prakken, An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2): 93-124, 2010.
[4] J. L. Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA,1995.
[5] D. Makinson and K.Schlechta, Floating conclusions and zombie paths, *Artificial Intelligence* 48: 199-209, 1991.
[6] J. L. Pollock, Defeasible reasoning and degrees of justification, *Argument and Computation*, 1(1): 7-22, 2010.

Interleaved Argumentation and Explanation in Dialog

Ioan Alfred Letia and **Adrian Groza**[1]

Abstract. Our goal is to provide computational models for natural arguments for the concepts of argument and explanation studied in the informal logic literature. Apart from distinguishing explanations from arguments we show our approach for modeling them. We describe the communicative acts of the agents by representing their different views on the topics of the dialog. By using description logics to define the differences, its reasoning is used to distinguish arguments from explanations.

1 INTRODUCTION

Argument and explanation are considered distinct and equally fundamental [8], whose complementary relationship [9] is a central issue for identifying the structure of natural dialogs. Considering the costs or arguing [13], the thesis of this research states that in the majority of natural dialogs people prefer to explain things and not just arguing on them.

In this study we also investigate the relation between knowledge, argument, and explanation. The role of knowledge in argumentation have been stressed out by Walton [17]. In natural dialogs knowledge interleaves with argumentation. When performing reasoning tasks on available knowledge, agents perform better if the reason is argumentative [11]. On the one hand, knowledge of agents is exploited when generating, conveying, and assessing arguments. On the other hand, argumentation can be an efficient tool for knowledge acquisition or collaborative knowledge construction.

The complementarity between argument and explanation is best characterized by the fact that humans tend to take decision both on knowledge and understanding [18]. For instance, in judicial cases, circumstantial evidence needs to be complemented by a motive explaining the crime, whilst the explanation itself is not enough without plausible evidence [9]. In both situations the pleading is considered incomplete if either argumentation or explanation is missing.

The following section stresses out the differences between argument and explanation as they already have been addressed in the current schools of thought in philosophical sciences. Section 3 illustrates how the distinguishing features of arguments and explanation can be modeled in description logic. Section 4 analysis the situation when parties differently interpret reasons as argument and explanation. Section 5 approaches the specific communicative acts from the perspective of differentiating between argument and explanation and shows the dynamics of these two interpretation in a natural dialog. After browsing related work in section 6, section 7 concludes the paper.

[1] Technical University of Cluj-Napoca, Romania, email: {letia,adrian}@cs-gw.utcluj.ro

2 DISTINGUISHING ARGUMENT FROM EXPLANATION

The role of argument is to establish knowledge, whilst the role of explanation is to facilitate understanding. Thus, to make an instrumental distinction between argument and explanation, one has to distinguish between knowledge and understanding. One legitimate question would be: does understanding represent more knowledge? From the perspective of [citation needed], knowledge represents awareness of information, whilst understanding represents the awareness of the connections between pieces of information. In the simplest computational model, understanding of a concept can be quantified in terms of the number of relations an agent is aware in a given context regarding that concept. A supplementary constraint would impose these relations to include causal, and other types of roles among them, in order to assign a meaning to concept. Note that both concepts are defined in terms of the epistemic notion of awareness. From an operational or behavioral viewpoint, understanding allows the knowledge to be put in practice. In this line, understanding represents a deeper level than knowledge.

The interaction between argument and explanation is the basic mechanism for augmenting an agent's knowledge and understanding. We consider the following distinctive features of argument and explanation:

1. Argumentation starts with a conflict. Explanation starts with non-understanding.
2. In explanation the roles are usually asymmetric: the explainer is assumed to have more understanding and wants to transfer it to the explainee. In argumentation, both parties start the debate from equal positions, thus initially having the same roles. Only at the end of the debate the asymmetry arises when the winner is considered to have more relevant knowledge on the subject.
3. In explanation one party supplies information. There is a linguistics indicator which requests that information. Because in argumentation it is assumed that all parties supply information, no indicator of demanding the information is required.

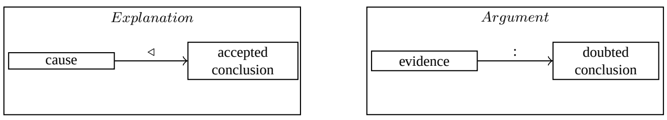


Figure 1. Distinguishing argument from explanation.

Regarding the first topic, for an argument, premises represent evidence supporting a doubted conclusion. For an explanation, the conclusion is accepted and the premises represent the causes of the consequent (see figure 1). The explanation aims to understanding the explanandum by indicating what causes it, whilst an argument aims to persuade the other party about a believed state of the world. An argument in considered adequate in principle if there is at least one agent who justifiably believes that the premises are true but who does not justifiably believe this about the consequent [7]. An explanation is adequate in principle if all the agents accepting the premises would also accept the consequent. The function of argument is to "transfer of justified belief", whilst the role of explanation is to "transfer of understanding".

Regarding the second topic above, consider the dialog between a teacher an a sophomore student which is almost entirely explicative. The ontology of the student regarding the specific scientific field is included in the ontology of the teacher. As the ontology of the student increases, resulting in different perspectives on the subject, exchanging arguments may occur.

The above scenario helps us to extract several knowledge conditions for arguments. Firstly, a doubted conclusion arises from different knowledge bases. Assuming the same reasoning capabilities, the precondition states that the agents should have different ontologies in order to be able to rise arguments. Formally, the intersection between agents ontologies shouldn't be empty ($\mathcal{O}_i \cap \mathcal{O}_j = \mathcal{O}_{ij} \neq \emptyset$), such that the agents can communicate, but the differences should be consistent enough to generate arguments ($\mathcal{O}_i \setminus \mathcal{O}_j \neq \emptyset$ and $\mathcal{O}_j \setminus \mathcal{O}_i \neq \emptyset$). The arguments are constructed based on knowledge in the symmetric difference of the agents ontology $\mathcal{O}_i \Delta \mathcal{O}_j = \mathcal{O}_i \setminus \mathcal{O}_j \cup \mathcal{O}_j \setminus \mathcal{O}_i$. Depending on the granularity of the common ontology $\mathcal{O}_{ij}$, one agent should convey more abstract or more concrete arguments in order to adapt them to the audience.

Regarding the third topic, the easiest way to distinguish between explanation and argument is to compare arguments $for\ F$ and explanations $of\ F$. The mechanism should distinguish between whether F is true and why F is true. In case F is a normative sentence, the distinction is difficult [18]. If F is an event, the question why F happened is clear delimited by the whether F happened.

3 REPRESENTING ARGUMENTS AND EXPLANATION

After browsing the technical instrumentation provided by description logics, this section models the distinguishing features of arguments and explanation in description logics (DL).

This section assumes that the reader is familiar with the basic concepts of description logics and the main idea of the Argument Interchange Format (AIF) ontology. Given that $Reason \sqsubseteq RuleScheme$ in the AIF ontology, we have:

Definition 1 *An argument is a reason in which the premises represent evidence in support of a doubted conclusion.*

$$Argument \sqsubseteq Reason \sqcap \forall hasPremise.Evidence \quad (1)$$

$$Argument \sqsubseteq (=1) hasConclusion.DoubtedStatement \quad (2)$$

Definition 2 *An explanation is a reason in which the premises represent a cause of an accepted fact.*

$$Explanation \sqsubseteq Reason \sqcap \forall hasPremise.Cause \quad (3)$$

$$Explanation \sqsubseteq (=1) hasConclusion.Fact \quad (4)$$

We define a doubted statement as a statement that is challenged by one agent:

$$DoubtedStatement \sqsubseteq \exists challenge.Statement \quad (5)$$

where the $challenge$ role has the concept $Agent$ as domain:

$$\exists rejects.Statement \sqsubseteq Agent \quad (6)$$

No $challenge$ relation should exist for a statement accepted as a fact, given by:

$$Fact \sqsubseteq Statement \sqcap \forall challenge.\perp \quad (7)$$

Both pieces of evidence and causes represent statements:

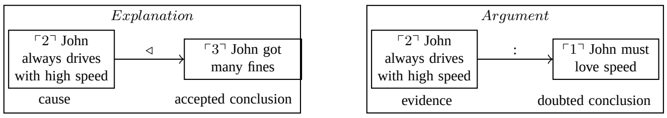
$$Evidence \sqsubseteq Statement, Cause \sqsubseteq Statement \quad (8)$$

We can refine this top level ontology by classifying evidence (in shortcut notation Ev), in direct or circumstantial evidence:

$$DirectEv \sqsubseteq Ev \sqcap \exists directsupport.DoubtedStatement \quad (9)$$

$$CircumstantialEv \sqsubseteq Ev \sqcap \exists indirectsupport.DoubtedStatement \quad (10)$$

where the practice in law treats a motive as circumstantial evidence: $Motive \sqsubseteq CircumstantialEvidence$.



$\ulcorner 2 \urcorner : Cause$	$\ulcorner 2 \urcorner : Evidence$
$\ulcorner 3 \urcorner : Fact$	$\ulcorner 1 \urcorner : DoubtedStatement$
$e : Reason$	$a : Reason$
(e $\ulcorner 2 \urcorner$): hasPremise	*(a $\ulcorner 2 \urcorner$): hasPremise*
(e $\ulcorner 3 \urcorner$): hasConclusion	*(a $\ulcorner 1 \urcorner$): hasConclusion*

Figure 2. Argument-explanation pattern: the same statement acts as a cause for an accepted statement and as an evidence for a doubted statement.

Example 1 (Argument-explanation pattern) $\ulcorner 1 \urcorner$ *John must love speed.* $\ulcorner 2 \urcorner$ *He drives with high speed all the time.* $\ulcorner 3 \urcorner$ *That's why, he got so many fines.* $\ulcorner 2 \urcorner$ *represents the support of argument* $\ulcorner 1 \urcorner$*, but also it acts as an explanation for* $\ulcorner 3 \urcorner$ *(see figure 2).*

Given the assertions in figure 2, e is classified by the agent p possessing the above knowledge as an explanation, whilst a as an argument.

Assuming that his partner o has the following assertion: *(related o* $\ulcorner 3 \urcorner$ *rejects)*. It means that the agent o classifies the statement $\ulcorner 3 \urcorner$ as doubted, and thus it does not treat the reason e as an explanation.

Each agent can have different interpretation functions of the same chain of statements.

Example 2 $\ulcorner 1 \urcorner$ *Heloise and Abelard are in love.* $\ulcorner 2 \urcorner$ *Heloise and Abelard are getting married.*

One agent can interpret $\ulcorner 1 \urcorner$ as a cause for the accepted fact $\ulcorner 2 \urcorner$, treating the reason as an explanation. An agent with a different interpretation function $\mathcal{I}$ will assert $\ulcorner 2 \urcorner$ as evidence for the doubted conclusion $\ulcorner 1 \urcorner$, therefore rising an argument.

If one does not have any assumptions or contextual clue about the acceptance status of the other agent regarding the consequent, does

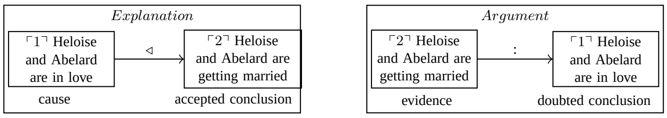
Agent A ($\mathcal{O}_A$)
$u : GoodUniversity$
$GoodUniversity \sqsubseteq \exists hasGood.ResearchFacility$

Agent's A view on agent B ($\mathcal{O}_{AB}$)
$u : GoodUniversity$
$GoodUniversity \equiv \exists hasGood.(ResearchFacility \sqcap TeachingFacility)$

Agent B view on agent A. ($\mathcal{O}_{BA}$)
$u : ResearchInstitute$
$ResearchInstitute \sqsubseteq \exists hasGood.ResearchFacility$

Agent B ($\mathcal{O}_B$)
$u : GoodUniversity$
$GoodUniversity \equiv \exists hasGood.ResearchFacility \sqcup \forall hasGood.TeachingFacility$

Figure 4. Subjective views of the agents.



$\ulcorner 1 \urcorner : Cause$	$\ulcorner 2 \urcorner : Evidence$
$\ulcorner 2 \urcorner : Fact$	$\ulcorner 1 \urcorner : DoubtedStatement$
$e : Reason$	$a : Reason$
(e ⌜1⌝): hasPremise	*(a ⌜2⌝): hasPremise*
(e ⌜2⌝): hasConclusion	*(a ⌜1⌝): hasConclusion*

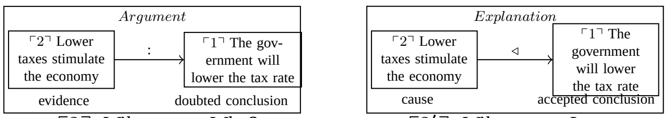
Figure 3. Different interpretation functions on the same statements. ⌜1⌝ acts as a premise in the first interpretation and as a conclusion in the second one.

the agent consider it as a fact or as doubted conclusion? We argue that, human agents tend to explain things instead of arguing on them, if no reason to argue or conflict have been previously identified. The usual reluctance of the human agent to argue is supported by the fact that the involvement in an argument may lead to more costs than benefits [13], in many quotidian scenarios. It means that, in our model, by default, agents convey explanations instead of arguments. If an agent accepts the conclusion according to its interpretation function, it treats the reason as an explanation.

One question regards how the agents can exploit the information that the given dialog is interpreted as an explanation by one party and as an argument by the other one, in order to eliminate the ambiguity?

Consider the example in [5]:
⌜1⌝: Bob says, The government will inevitably lower the tax rate.
⌜2⌝: Wilma says, Why?
⌜3⌝: Bob says, Because lower taxes stimulate the economy.

It is presented as an argument with the consequent ⌜1⌝ supported by the premise ⌜3⌝. Assume the Wilma's reply is slightly modified, given by:
⌜2′⌝: Wilma says, I agree. Why do you consider this?



⌜2⌝: *Wilma says, Why?*

⌜2′⌝: *Wilma says, I agree. Why do you consider this?*

Figure 5. The dialog provides indicators of the status of the consequent: accepted or doubted.

By accepting the statement ⌜1⌝, it becomes a fact in the system represented by the two agents Bob and Wilma. Consequently, the reason becomes an explanation in which the cause "lower taxes stimulate the economy" may explain the government decision (figure 5). Under the assumption that an agent accepts a statement only if it has a level of understanding of that sentence[2], one can infer that Wilma has own explanation regarding the fact ⌜1⌝, but she wants to find out the explanation of her partner.

Another issue regards the distinction between evidence and cause. Cognitive experiments [4] have shown difficulties when distinguishing between them, only 74% have correctly classified pieces of information as evidence or cause. Moreover, human agents are able to build a strategy of substituting explanation in case evidence is not available [4].

4 SUBJECTIVE VIEWS

The agents construct arguments and explanations from their knowledge bases which do no completely overlap. In the same time, each party has a model about the knowledge of his partner. Consider the partial knowledge in figure 4. Here the agent A sees the individual u as a good university, where a good university is something included in all objects for which the role $hasGood$ points towards concepts of type $ResearchFacility$. According to agent's B knowledge, u is also a good university, but the definition is more relaxed: something is a good university if it has at least one good research facility or all the teaching facilities are good. According to agent A perspective on the knowledge of the agent B, u belongs to the concept of good universities, but the definition is perceived as being more restrictive: a good university should have at least one good research facility but also at least one good teaching facility. From the opposite side, agent B imagines that A asserts u as an research institute, where a research institute should have good research facility.

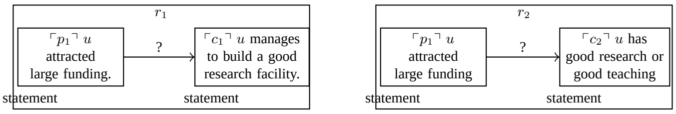
Suppose the agent A conveys different reasons supporting the statement c_1: "u has good research facility" and c_2: "u has either good research or good teaching". For instance:

r_1: "Because u attracted large funding from research projects, it manages to build a good research facility."

r_2: "Because u attracted large funding from research projects, it should have either good research or good teaching."

The above reasons are graphically represented in figure 6.

[2] One can imagine a situation in which an expert explains something to you, you do not understand, but given the reputation or trust relation that you have with the expert, you accept the explanation.



$\ulcorner p_1 \urcorner : Statement$	$\ulcorner p_2 \urcorner : Statement$
$\ulcorner c_1 \urcorner : Statement$	$\ulcorner c_2 \urcorner : Statement$
$r_1 : Reason$	$r_2 : Reason$
($r_1, \ulcorner p_1 \urcorner$): hasPremise	*($r_2, \ulcorner p_1 \urcorner$): hasPremise*
($r_1, \ulcorner c_1 \urcorner$): hasConclusion	*($r_2, \ulcorner c_2 \urcorner$): hasConclusion*

Figure 6. Possible reasons conveyed by the agent A. Are they arguments or explanations?

The question regards how does the agent A treat the reason, as an explanation or as an argument, when conveying it to the agent B. Given the models in figure 4, how the receiving agent B perceives the reason: an explanatory or an argumentative one? The following technical details are introduced to approach these questions.

Agents ontologies/consequent	c_1	c_2
$\mathcal{O}_A$	$\oplus$	$\oplus$
$\mathcal{O}_{AB}$	$\oplus$	$\ominus$
$\mathcal{O}_B$	$\ominus$	$\oplus$
$\mathcal{O}_{BA}$	$\oplus$	$\oplus$

Table 1. The acceptance of the consequents c_1 and c_2 based on agents ontologies.

To distinguish between explanation and argument, the most important issue regards the acceptance of the consequent. In table 1, $\oplus$ denotes that the ontology $\mathcal{O}_i$ entails the consequent c_j. The statement c_1 can be derived from the ontology $\mathcal{O}_A$ (figure 5.3). It cannot be inferred (noted with $\ominus$) by the agent B based on his ontology $\mathcal{O}_B$, because in his interpretation a university with all teaching facilities good but without good research facilities is also a good university.

Instead, the statement c_2 fits the definition of good ontology in $\mathcal{O}_B$. Because agent A accepts its first part "u has good research", he should consider c_2:u has good research or good teaching" as valid. Similarly, agent A considers that agent B cannot infer c_2 ($\ominus$ in table 1), even if the $\mathcal{O}_B$ ontology entails c_2. The agent A has a wrong representation $\mathcal{O}_{AB}$ regarding how the agent B views the statement c_2. Even if the agent B has a wrong model $\mathcal{O}_{BA}$, based on which he believes that the agent A interprets u as a research institute instead of a university, the consequent c_2 is still derived based on axiom $ResearchInstitute \sqsubseteq \exists hasGood.ResearchFacility$.

Firstly, the knowledge of agent A and its model about the knowledge of B represents the world of agent A, noted with w_A. Similarly, the subjective world w_B of agent B consists of the knowledge of B and his view on the knowledge of the agent A. The knowledge of A combined with the knowledge of B, represent the objective world w_O (table 2). A statement is considered $accepted$ if it is entailed by both ontologies. If at least one ontology does not support the statement it is considered $doubted$. The following algebra encapsulates this:

$$\oplus + \oplus = Accepted \qquad \oplus + \ominus = Doubted$$
$$\ominus + \oplus = Doubted \qquad \ominus + \ominus = Doubted$$

In table 2, the agent A treats c_1 as accepted, meaning that from his point of view the reason r_1 represent an explanation. Agent B perceives the sentence c_1 as doubted, therefore it considers that he is hearing an argument. Note that in the objective world w_O, the reason r_1 is actually an argument. Which means that agent A is wrong about the model of his partner B. Consider that the reason r_1 is uttered by the agent B. He believes that he is conveying an argument, which is true in the objective world w_O. Agent A considers that he is receiving an explanation.

World	Ontologies	c_1	c_2
w_O	$\mathcal{O}_A + \mathcal{O}_B$	Accepted	Doubted
w_A	$\mathcal{O}_A + \mathcal{O}_{AB}$	Doubted	Accepted
w_B	$\mathcal{O}_B + \mathcal{O}_{BA}$	Accepted	Doubted

Table 2. The acceptance of the consequents c_1 and c_2 based on agents ontologies.

The statement c_2 being perceived as doubted in w_A, the agent A consider that he is conveying an argument. In the world w_B, the conclusion is accepted, thus agent B hearing an explanation, which is true in the objective world w_O. In this situation, agent B should signal to his partner: "There is no need to persuade me. I agree with the consequent."

The rightness or adequacy of conveying either argument or explanation should be computed relative to the objective world w_O. Given the difference between expecting explanations or arguments (subjective worlds w_A and w_B) and legitimate ones (objective world w_O), the agents may wrongly expect explanations instead of arguments and vice-versa. For the rightness or adequacy of conveying/expecting argument or explanation, the algebra in figure 7 is used.

$Accepted_O + Accepted_X = \oplus_X^w$	agreement rightness
$Accepted_O + Doubted_X = \oplus_X^{\neg w}$	agreement not aware
$Doubted_O + Accepted_X = \ominus_X^{\neg w}$	conflict not aware
$Doubted_O + Doubted_X = \ominus_X^w$	conflict rightness

Figure 7. Rightness/inadvertence regarding expecting/conveying argument or explanation. The first operator represents the actual world w_O, while the second one the subjective perspective of agent X.

The situation resulting by applying the algebra in 7 on the given scenario is presented in table 3. Agent B, even if his model about A is not accurate, manages to figure out the status of both consequents c_1 and c_2. Quite differently, agent A is ignorant with respect to both conclusions.

Agent	Awareness and Ignorance	c_1	c_2
A	$w_O + w_A$	$\ominus_A^{\neg w}$	$\oplus_A^{\neg w}$
B	$w_O + w_B$	$\ominus_B^w$	$\oplus_A^w$

Table 3. Agreement and conflict awareness for agents A and B regarding the consequents c_1 and c_2.

The question is if it is possible for the hearing agent to indicate to the conveyor agent that a wrong assumption has been made. The problem is that no agents are aware of the objective world w_O. Under certain conditions, the inadvertence could be identified and solved. If a mediator would be introduced, aware of w_O, it would be able to identify conflict and to provide guidance for increasing the dialog efficiency. The second option would be by analyzing the communicative acts. If the agent A announces that r_1 is an explanation, agent B can disclose his doubts about c_1. By updating his model $\mathcal{O}_{AB}$, the agent A will re-interpret r_1 as an argument. By specifying pre-conditions and post-conditions of such communicative acts, the participants in the dialog can infer the status of a reason: argument or explanation.

5 COMMUNICATIVE ACTS

The following speech acts are analyzed only from the perspective of distinguishing between argument and explanation. After modeling the communicative acts in DL, their preconditions and postcodntions are formally specified. The dynamics of the dialog is illustrated by a scenario.

5.1 Speech acts in description logic

The definition presented here are in line with the speech acts proposed by Reed [14] for modeling dialogs in the AIF ontology. Our refinement focuses on the distinction between argument and explanation.

Firstly we need to distinguish between explicative and argumentative questions, where a question is linked to the AIF ontology based on the subsumption relation $Question \sqsubseteq LocutionDescriptor$. An argumentative question should have a doubted conclusion as topic, given by axiom 11.

$$ArgumentativeQ \sqsubseteq Question \sqcap \exists\exists hasTopic.DoubtedStatement \tag{11}$$

When conveying an argumentative question a doubt regarding the topic is indicated to receiving agent. In a general model allowing more topics for a single question, one doubted topic is enough to interpret the question as an argumentative one, given by the existential quantification of the role $hasTopic$. Questions of type "How do you know" are a particular case of argumentative ones, given by:

$$HowDoYouKnow? \sqsubseteq ArgumentativeQ \tag{12}$$

For an explicative question all the topics should not be doubted:

$$ExplicativeQ \sqsubseteq Question \sqcap \exists\forall hasTopic.\neg DoubtedStatement \tag{13}$$

Questions of type "Why?" are particularly considered to request for explanation: $Why \sqsubseteq ExplicativeQuestion$.

Similar to [14], a response is a compound concept triggered by a specific question and ended by something which can be a statement, a reason, or another rule, but which remain unspecified at the top level of the ontology:

$$Response \sqsubseteq \exists hasStart.Question \sqcap hasEnd.\top \tag{14}$$

In the AIF ontology the statements can be challenged and when this happens this is a good indicator for us that the particular statement is doubted. In our model, $challenge$ is seen as a particular role, refined by:

$$reject \sqsubseteq challenge, contest \sqsubseteq challenge \tag{15}$$

Beside rules application nodes, AIF includes also conflict nodes. Here, all the roles of type $hasStatement$ of a conflict application rule necessarily point to doubted statements.

$$Conflict \sqsubseteq Rule \sqcap (=2) hasStatement.DoubtedStatement \tag{16}$$

5.2 Pre- and post-conditions

claim argument r

prec	$w_x \models \{c\} \in Doubted^{\mathcal{I}_x}$
	$w_x \models \{p\} \in \neg Doubted^{\mathcal{I}_x} \cap Evidence^{\mathcal{I}_x}$
post	$\mathcal{O}_{yx} \models \{c\} \in Doubted^{\mathcal{I}_x}$
	$\mathcal{O}_{yx} \models \{p\} \in \neg Doubted^{\mathcal{I}_x} \cap Evidence^{\mathcal{I}_x}$
	if $\mathcal{O}_y \models \{c\} \in \neg Doubted^{\mathcal{I}_y} \Rightarrow \mathcal{O}_y \models \{c\} \in \neg Doubted^{\mathcal{I}_y}$

claim explanation r

prec	$w_x \models \{c\} \in \neg Doubted^{\mathcal{I}_x}$
	$w_x \models \{p\} \in \neg Doubted^{\mathcal{I}_x} \cap Cause^{\mathcal{I}_x}$
post	$\mathcal{O}_{yx} \models \{c\} \in \neg Doubted^{\mathcal{I}_x}$
	$\mathcal{O}_{yx} \models \{p\} \in \neg Doubted^{\mathcal{I}_x} \cap Cause^{\mathcal{I}_x}$

Figure 8. Claiming arguments and explanations.

Claim argument The main precondition to utter an argument is that the agent should believe from his knowledge base that a divergence of opinion exists with his partner. Assume that an agent x conveys agent y an argument r having the support p and consequent c (figure 8). The first precondition states that based on the axioms in the world of x the consequent should be interpreted as doubted. The second precondition for the agent x to convey an argument is to consider the precondition p as an evidence which is not doubted at the moment. The first tow post-conditions regards how the world of agent y is updated in the light of new information. Especially, the model $\mathcal{O}_{yx}$ about his partner is updated.

Claim explanation. The precondition to convey an explanation, is that the agent x should interpret the consequent c in his world w_x as not doubted. From the pragmatics of natural dialogs perspective, an explanation occurs only if a request for such an explanation has been conveyed [8, 16]. Such an explanation request signals the possibility that a transfer of understanding may occur. Rather then rejecting an explanation, the explainee would consider it as $irrelevant$.

Argumentative question. An agent x conveys an argumentative question only when the consequent c of the reason is not interpreted as factive in his knowledge base: $O_x \models \{c\} \in \neg Doubted^{\mathcal{I}_x}$. The hearing agent y realizes that the consequent is doubted in his world: $w_y \models \{c\} \in \neg Doubted^{\mathcal{I}_y}$.

Explicative question. An agent x can utter an explicative question if the consequent c of the reason is interpreted as factive in his knowledge base: $O_x \models \{c\} \in \neg Doubted^{\mathcal{I}_x}$. The hearing agent y realizes that the consequent is accepted by his partner: $\mathcal{O}_{YX} \models \{c\} \in \neg Doubted^{\mathcal{I}_y}$.

Challenge. In the common ontology the range of the challenge role is the top level concept $\top$: It means that one can attack a statement, either evidence or cause, but also a reason, either argumentative or explicative. For accepting a reason there two flavors: $agree$ speech act for arguments and $understand$-like acts for explanations.

5.3 Dialog dynamics

Consider the dialog in education domain from figure 9, taking place between a scholar S and administrator A: Assume that after the move m_1 both parties correctly identified the reason r_1, interpreting the statement ⌜1⌝ as the premise and the statement ⌜2⌝ as the conclusion. (figure 10). Moreover, the conveyor agent S interprets ⌜1⌝ as a cause which makes possible to assign more founds for investments. Given no support for rejecting the statements ⌜2⌝ and ⌜1⌝, based on axiom (7) they are interpreted as facts by the agent S: $Fact^{I_S} = \{\ulcorner 2 \urcorner, \ulcorner 1 \urcorner\}$. With the causal premise ⌜1⌝ and a factual consequent ⌜2⌝, both axioms (3) and (4) being satisfied by the reason r_1. Thus, it represents an explanation for the agent S, given by $Explanation^{I_S} = \{r_1\}$.

m_1	S:	Because the global income of our department has increased, it brings the possibility to assign more funds for teaching and research facilities.
m_2	A:	Are you sure that the global income has increased?
m_3	S:	Because the number of students has increased, the partial income has increased.
m_4	A:	Partial income has been affected by the wage being increased.
m_5	S:	Is it so? My wage did not increase.
m_6	A:	The wage expenses has risen due to the recruitment of new staff in the last semester.
m_7	S:	Maybe that's why my wage did not increase.
m_7	A:	Anyhow, knowing that .... it is a good idea to increase the research facilities.

Figure 9. Dialog in education domain.

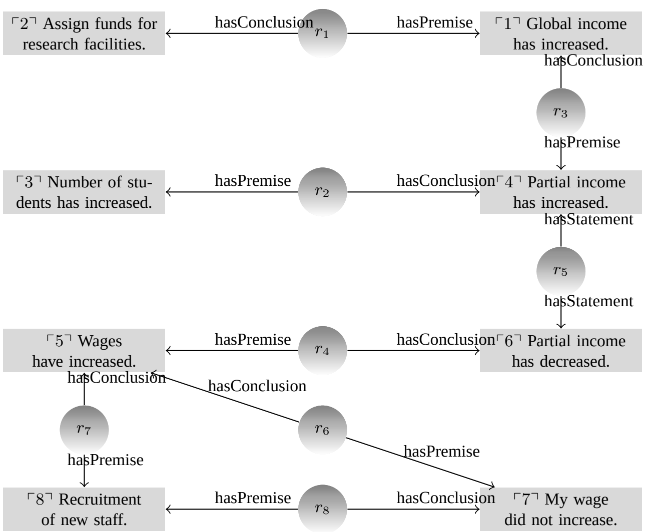


Figure 10. Supporting reasons in dialog.

Assume that the agent A contests all the statements that are not proved, given by:

$$\forall isProved.\perp \sqsubseteq \exists contest.Statement \qquad (17)$$

No proof existing at this moment, the statement ⌜1⌝ is labeled as contested by the agent A: $contest^{I_A} = \{(A, \ulcorner 1 \urcorner)\}$. $Contest$ being a subrole of $challenge$ (according to common vocabulary in figure 10), the statement ⌜1⌝ is interpreted as doubted statement based on definition (5). It means that the preconditions to utter an argumentative question are satisfied.

The move m_2 clearly introduces some doubts regarding the statement ⌜1⌝, meaning that the agent A has no difficulties to interpret the question *Are you sure that...*, notes with q_1, as an argumentative question (line 3 in figure 11), with the topic represented by the statement ⌜1⌝, given by: $(q_1, \ulcorner 1 \urcorner) : hasTopic)$. Consequently, based on the common axiom (14), both agents become aware that the topic ⌜1⌝) is doubted in the current dialog: $DoubtedStatement^{I_S} = DoubtedStatement^{I_A} = \{\ulcorner 1 \urcorner\}$.

At this moment, agent S solves the inconsistency risen by the axioms $Fact \sqsubseteq \neg DoubtedStatement$, $Fact^{I_S} = \{\ulcorner 1 \urcorner\}$ after the move m_1, and $DoubtedStatement^{I_S} = \{\ulcorner 1 \urcorner\}$ after the move m_2 by removing his initially wrong interpretation of ⌜1⌝ as a fact.

Both agents identify the move m_3 as a reason $Reason^{I_S} = Reason^{I_A} = \{r_1, r_2\}$, with the corresponding premise ⌜3⌝ and consequent ⌜4⌝. Given the interpretation of the premise ⌜3⌝ as a $Cause$ by the agent S, and no rejection of the consequent, the reason r_2 is also interpreted at this moment as an explanation: $Explanation^{I_S} = \{r_2\}$.

The move m_3 represents also the response of agent S triggered by the question q_1. The formalization says that starting by the question q_1 $(r_2, q_1) : hasStart$, agent S answers with r_2, where r_2 is interpreted as an response by the agent S uttering it, based on axiom (16). Being interpreted as a response by the conveyor, one of the statements in r_2 should have be related with the topic risen by q_1. Thus, according to the cognitive map of S, the cognitive consistency is assured by the reason r_3. Because it has the conclusion ⌜1⌝ which doubted and the premise ⌜1⌝ representing a fact, the reason r_3 represents an argument from the agent S viewpoint.

Recall that the topic of the q_1 question is the statement ⌜1⌝, but the topic itself does not explicitly appear in the declaration r_2. It means that the hearing agent A can correctly interpret it as the response for q_1, but also as an independent declaration in the dialog flow, with the issue risen by q_1 still open. One option would be to ask for clarifications regarding the membership of the individual r_2 to the $Response$ class, or the second one, simply to react to the just uttered sentence r_2. The clarification may come on the form of the r_2 reason, which will synchronize the cognitive maps of the two agents.

In the current dialog, A chooses to focus on one of the statements risen by r_2 because it is aware of a conflict regarding the statement ⌜4⌝. Based on definition 11, the statement is categorized by the agent A as doubted, thus interpreting the reason r_2 as an argument: $Argument^{I_A} = \{r_2\}$.

In move m_4, the premises and the conclusion of reason r_4 are correctly identified by both agents. The conflict between the statements "partial income has increased" and "partial income has decreased" is also clear. Based on common axiom 11 regarding conflict rules, both agents become aware the the consequents ⌜4⌝ and ⌜6⌝ are doubted. At this moment r_4 and r_2 should be interpreted as arguments by both parties: $Argument^{I_S} = \{r_1, r_2, r_3, r_4\}$, respectively $Argument^{I_A} = \{r_1, r_2, r_4\}$. Being the agent who proposed the argument, the agent A is not aware of any attack relation on the premise ⌜5⌝ supporting it. Therefore, according to agent's A knowledge base, the statement is a fact: $Fact^{I_A} = \{\ulcorner 5 \urcorner\}$.

The move m_5 indicates that agent S has a different opinion. Firstly, it rises the argumentative question q_2: "Is it so?". Based on it and on the common knowledge in axiom 14, agent S realizes that the statement $\ulcorner 5 \urcorner$ is doubted. Agent S also provides evidence $\ulcorner 5 \urcorner$ in support of his argument r_6.

At move m_6, knowing that the statement $\ulcorner 5 \urcorner$ is doubted, the agent A can rise only arguments supporting it. The argument r_7 is valid because its premise $\ulcorner 8 \urcorner$ is not attacked at this moment of dialog, according to the knowledge base of the conveyor agent A. According to the current interpretation function of A, the statement $\ulcorner 8 \urcorner$ is both evidence for r_7 and also a fact.

In the move m_7, agent S interprets the statement $\ulcorner 8 \urcorner$ as an explanation why his salary did not increase, given that the global income of the department has increased: $Explanation^{I_S} = \{r_8\}$. Depending on the next moves and possible challenge relations on $\ulcorner 8 \urcorner$ from the administrator A, the reason r_8 may shift to an argument. Note that at this moment a transfer of understanding takes place.

Move	I_S	I_A
m_1	$r_1 : Reason$ $(r_1, \ulcorner 1 \urcorner) : hasPremise$ $(r_1, \ulcorner 2 \urcorner) : hasConclusion$ $\ulcorner 1 \urcorner) : Cause$	$r_1 : Reason$ $(r_1, \ulcorner 1 \urcorner) : hasPremise$ $(r_1, \ulcorner 2 \urcorner) : hasConclusion$ $(A, \ulcorner 1 \urcorner) : contest$
m_2	$q_1 : ArgumentativeQ$ $(q_1, \ulcorner 1 \urcorner) : hasTopic$	$q_1 : ArgumentativeQ$ $(q_1, \ulcorner 1 \urcorner) : hasTopic$
m_3	$(r_2, q_1) : hasStart$ $(r_2, r_2) : hasEnd$ $r_2 : Response$ $r_3 : Reason$ $(r_3, \ulcorner 4 \urcorner) : hasPremise$ $(r_2, \ulcorner 1 \urcorner) : hasConclusion$ $r_2 : Reason$ $(r_2, \ulcorner 3 \urcorner) : hasPremise$ $(r_2, \ulcorner 4 \urcorner) : hasConclusion$ $\ulcorner 3 \urcorner : Cause$	$r_2 : Reason$ $(r_2, \ulcorner 3 \urcorner) : hasPremise$ $(r_1, \ulcorner 4 \urcorner) : hasConclusion$
m_4	$r_4 : Reason$ $(r_4, \ulcorner 5 \urcorner) : hasPremise$ $(r_4, \ulcorner 6 \urcorner) : hasConclusion$ $r_5 : ConflictRule$ $\ulcorner 6 \urcorner : DoubtedStatement$ $\ulcorner 4 \urcorner : DoubtedStatement$	$r_4 : Reason$ $(r_4, \ulcorner 5 \urcorner) : hasPremise$ $(r_4, \ulcorner 6 \urcorner) : hasConclusion$ $r_5 : ConflictRule$ $\ulcorner 6 \urcorner : DoubtedStatement$ $\ulcorner 4 \urcorner : DoubtedStatement$ $\ulcorner 5 \urcorner : Fact$
m_5	$q_2 : ArgumentativeQ$ $(q_2, \ulcorner 5 \urcorner) : hasTopic$ $\ulcorner 7 \urcorner : Evidence$ $r_6 : Reason$ $(r_6, \ulcorner 7 \urcorner) : hasPremise$ $(r_6, \ulcorner 5 \urcorner) : hasConclusion$	$q_2 : ArgumentativeQ$ $(q_2, \ulcorner 5 \urcorner) : hasTopic$ $\ulcorner 7 \urcorner : Statement$ $r_6 : Reason$ $(r_6, \ulcorner 7 \urcorner) : hasPremise$ $(r_6, \ulcorner 5 \urcorner) : hasConclusion$
m_6	$r_7 : Reason$ $(r_7, \ulcorner 8 \urcorner) : hasPremise$ $(r_7, \ulcorner 5 \urcorner) : hasConclusion$	$r_7 : Reason$ $(r_6, \ulcorner 8 \urcorner) : hasPremise$ $(r_6, \ulcorner 5 \urcorner) : hasConclusion$ $\ulcorner 8 \urcorner : Evidence$
m_7	$r_8 : Reason$ $(r_8, \ulcorner 8 \urcorner) : hasPremise$ $(r_8, \ulcorner 7 \urcorner) : hasConclusion$ $\ulcorner 8 \urcorner : Cause$	$r_8 : Reason$ $(r_8, \ulcorner 8 \urcorner) : hasPremise$ $(r_8, \ulcorner 7 \urcorner) : hasConclusion$

Figure 11. Dialog interpretation for each agent. I_A and I_S are the interpretation functions for the agent A, respectively S.

The following observations sum up the analysis of the dialog.

- Some reasons are explicit, some are implicit. For instance, the implicit conflicting rule r_5 is identified by both agents, whilst the implicit reason r_3 is known only by the agent S.
- An agent may consider that he conveys an explanation, but actually it represents in argument. (i.e r_2 after the move m_3).
- An agent may consider that he conveys arguments, but the reason represent an explanation.
- In the light of new information, the wrong interpretation may be updated (i.e. with the uttering of an argumentative question in move m_2, the reason r_1 is interpreted by the the agent S as an argument and not as an explanation based on initial assumptions in move m_1).
- Understanding can arise from conveying arguments: the explanation r_8 is constructed based on statements from two arguments r_7 and r_6.

Move	$Expl^{I_S}$	Arg^{I_S}	$Expl^{I_A}$	Arg^{I_A}
m_1	r_1			r_1
m_2		r_1		r_1
m_3	r_2	r_1,r_3		r_2
m_4		r_1,r_3, r_2, r_4		r_1, r_2, r_4
m_5		r_1,r_3, r_2, r_4, r_6		r_1, r_2, r_4, r_6
m_6		r_1,r_3, r_2, r_4, r_6, r_7		r_1, r_2, r_4, r_6, r_7
m_7	r_8	r_1,r_3, r_2, r_4, r_6, r_7		r_1, r_2, r_4, r_6, r_7

Table 4. Dynamics of argument and explanation in dialog.

6 DISCUSSION AND RELATED WORK

Explanation and argumentation capabilities [12] for more persuasive agents have already considered some aspects of user modeling. We have improved on this integration by also including the difference of the DL knowledge bases of agents. The informal approach [18] has been developed in this paper into a computational model of both argument and explanation.

Bex exploits in [3] argument-explanation complementarity for legal reasoning, while [12] for building more persuasive agents. Interleaving argument and explanation in natural dialogs has been investigated in [2] and [10]. Except for McBurney and Parsons', these models do not contain multiple perspectives.

Given different types of explanation patterns in social sciences, we have limited the approach to causal explanations. A broader investigation would include *constructive explanations*, explaining events by accounting knowledge structures such as scripts and plans or *contrastive explanations*, explaining surprising events by showing the deviation from expectation based on the available knowledge structures. One can also distinguish between conversational explanations and scientific explanations. The second category includes domain specific explanations: computing explanations, historical explanations, legal explanations, evolutionary biology explanations, which means that the top level ontology of explanations needs to be extended for each specific scientific field. Restricting explanations to causality, supports the idea that explanations are asymmetrical: if j explains F, then F does not explain j. Instead, arguments are not necessarily asymmetrical.

The problem is more complex when, besides knowledge, one considers different reasoning capabilities, but also different goals, preferences, or values of the agents. An argument may be more valuable from the individual perspective or from the collective viewpoint. In individualistic cultures values like egalitarianism, competitiveness, and self-reliance are higher ranked compared to hierarchies or cooperativeness in collective cultures. Consider the argument "Higher trained persons have good communication competencies" would be easily accepted by societies promoting activeness and implication of citizen, but most probable will be rejected by the Asian societies which rank lower eloquence skills.

Explanation aims to transfer understanding. For human agents, understanding occurs in different degrees, relative to their knowledge bases, beliefs, and goals. Cognitive understanding requires similar ontologies, but assumes agents have different goals and beliefs. The explainer should be able to explain how it comes to the conclusion and what hypotheses he had considered and rejected. The smallest degree of understanding, making sense, demands a coherent explanation, which usually is also an incomplete one. It means that, when the explainee conveys "I understand" speech act, the explainer can shift to an examination dialog in order to figure out the level of understanding, rather than a crisp value understand/not understand as suggested by Walton [16]. Acceptability standards of explanation can be defined similarly to the standard of proof in argumentative theory [6].

In their explanatory argumentation framework [15], the authors are showing how to apply abstract argumentation in scientific debates. We have been concerned here in mixing argument and explanation using DL knowledge so that human agents would be able to easily follow such a process. Therefore, our explanation was directed towards explaining on the knowledge level of the explainee, and not on explaining the workings of the abstract argumentation mechanism.

7 CONCLUSIONS AND FURTHER WORK

Our contributions are: (i) evidencing the instrumental role of knowledge structures through argumentation and explanation; (ii) providing guidelines to determine whether something in a dialog is an argument or an explanation [16]; (iii) modeling explanations similar to arguments in the AIF ontology. By using description logic to define the differences, its reasoning services are exploited aiming at automatic classification of arguments and explanations.

Ongoing work regards the exploitation of a more expressive DL-language for representing the model an agent has on his partner, as for instance a multi-agent version extension of $\mathcal{ALC}$ with multimodal operators, as introduced by [1]. Here, the belief, knowledge and temporal operators are encapsulated within the language itself.

Acknowledgements

We are grateful to the anonymous reviewers for their useful comments. Adrian Groza is supported by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557.

REFERENCES

[1] Franz Baader and Armin Laux, 'Terminological logics with modal operators', in *IJCAI*, pp. 808–815. Morgan Kaufmann, (1995).
[2] Floris Bex and Henry Prakken, 'Investigating stories in a formal dialogue game', in *Conference on Computational Models of Argument*, pp. 73–84, Amsterdam, The Netherlands, The Netherlands, (2008). IOS Press.
[3] Floris J. Bex, Peter J. Van Koppen, Henry Prakken, and Bart Verheij, 'A hybrid formal theory of arguments, stories and criminal evidence', *Artificial Intelligence and Law*, **18**(2), 123–152, (June 2010).
[4] Sarah Brem and Lance Rips, 'Explanation and evidence in informal argument', *Cognitive Science*, **24**, 573–604, (2000).
[5] Katarzyna Budzynska and Chris Reed, 'Speech acts of argumentation: Inference anchors and peripheral cues in dialogue', in *Computational Models of Natural Argument*, (2011).
[6] Thomas F. Gordon, Henry Prakken, and Douglas Walton, 'The carneades model of argument and burden of proof', *Artificial Intelligence*, **171**(10-15), 875–896, (2007).
[7] Cristoph Lumer, 'The epistemological theory of argument how and why?', *Informal Logic*, **25**, 213–242, (2005).
[8] Gregory Randolph Mayes, 'Resisting explanation', *Argumentation*, **14**, 361–380, (2000).
[9] Gregory Randolph Mayes, 'Argument explanation complementarity and the structure of informal reasoning', *Informal Logic*, **30**(1), 92–111, (2010).
[10] Peter McBurney and Simon Parsons, 'Representing epistemic uncertainty by means of dialectical argumentation', *Annals of Mathematics and Artificial Intelligence*, **32**(1-4), 125–169, (August 2001).
[11] Hugo Mercier and Dan Sperber, 'Why do humans reason? arguments for an argumentative theory', *Behavioral and Brain Sciences*, (34), 57–111, (2011).
[12] B. Moulin, H. Irandoust, M. Blanger, and G. Desbordes, 'Explanation and argumentation capabilities: Towards the creation of more persuasive agents', *Artificial Intelligence Review*, **17**, 169–222, (2002). 10.1023/A:1015023512975.
[13] Fabio Paglieri and Cristiano Castelfranchi, 'Why argue? Towards a cost-benefit analysis of argumentation', *Argument and Computation*, **1**, 71–91, (2010).
[14] Chris Reed, Simon Wells, Joseph Devereux, and Glenn Rowe, 'Aif+: Dialogue in the argument interchange format', in *COMMA*, pp. 311–323, (2008).
[15] Dunja Seselja and Christian Strasser, 'Abstract argumentation and explanation applied to scientific debates', *Synthese*, (2012). in press.
[16] Douglas Walton, 'A dialogue system specification for explanation', *Synthese*, **182**(3), 349–374, (2011).
[17] Douglas Walton and David M. Godden, 'Redefining knowledge in a way suitable for argumentation theory', in *Dissensus and the Search for Common Ground: Proceedings of OSSA June 2007*, pp. 1–13, (2007).
[18] Larry Wright, 'Reasoning and explaining', *Argumentation*, **16**, 33–46, (2002).

Combining Explanation and Argumentation in Dialogue

Floris Bex[1] and **Douglas Walton**[2]

Abstract. Explanation and argumentation can be used together in such a way that evidence, in the form of arguments, is used to support explanations. In a hybrid system, the interlocking of argument and explanation compounds the problem of how to differentiate between them. The distinction is imperative if we want to avoid the mistake of treating something as fallacious while it is not. Furthermore, the two forms of reasoning may influence dialogue protocol and strategy. In this paper a basis for solving the problem is proposed using a dialogue model where the context of the dialogue is used to distinguish argument from explanation.

1 INTRODUCTION

The hybrid model of [1][4] combines arguments and explanations in such a way that an argument can support an explanation. The idea of argumentation and explanation being combined is also familiar in the notion of *inference to the best explanation*. But in general, there is a difference between argument and explanation, and as we will show in this paper, it would be a fundamental error to criticize an argument as falling short of standards for a rational argument, when what was put forward was actually an explanation.

A problem is that in many cases of natural language discourse, the same piece of discourse can reasonably be interpreted as either an explanation or an argument. Similarly, a question 'Why?' can be interpreted as either asking for a reason that supports some claim of the speaker or as asking for an explanation for some observed anomaly. So here we have a pervasive problem, which can only be solved if we can find some clear and useful method of distinguishing between explanations and arguments. It is not only a problem for logic and discourse analysis, but also for explanation systems in computing [7], and particularly for hybrid models that combine argument with explanation [4][18].

Our solution to the problem of distinguishing argument and explanation lies in dialogue, more specifically, in speech act theory [26]. According to this view, it is the illocutionary force of the speech act in a dialogue that determines whether reasoning is argumentation or explanation [5]. Illocutionary force can be seen as the intention of uttering some locution: one can say p with an intention of explaining p, arguing for p, challenging p, promising p and so on. We thus argue that the distinction between argument and explanation is not a logical one but rather that the only correct way of making this distinction is to look at the dialogical context.

The question is then how to determine the purpose or intention of uttering a locution. In other words, how do we know whether some assertion is meant to explain a proposition or argue for it? The solution lies in the different purposes of explanation and argumentation. Argumentation is meant to convince someone else, explanation is aimed at helping them understand. Hence, the rules for argumentation and explanation are different.

There are various reasons for wanting to properly distinguish between argumentation and explanation. For example, we might want to be able to handle situations in which argumentation is fallacious whilst explanation is not. Furthermore, confusion of argumentation and explanation may lead to undesirable misunderstandings and unwanted behaviour in multi-agent dialogue, as the use of either argumentative or explanatory techniques may influence dialogue protocol and strategy. Finally, the distinction is important in the analysis of natural language texts.

In this paper, we discuss argumentation and explanation and how to distinguish between them. We also discuss an example of the fallacy of begging the question, which in case of an argument is a fallacy but for explanation is not. In section 3 we then show how argument and explanation can be combined in a dialogical setting and how the rules for arguing differ from the rules for explaining.

2 ARGUMENTATION AND EXPLANATION

How can one determine, in a given text of discourse where it is said that one event occurred because of another event, the text should be taken as representing an argument or an explanation? The problem is that cases where a given text of discourse could be interpreted as expressing either and argument or an explanation are fairly common, as an instructor of an informal logic course can tell you. Another factor is that in artificial intelligence, something called a justification explanation been recognized [7], suggesting that argument and explanation are often combined and work together. Suffice it to say that abductive reasoning, also commonly called inference to the best explanation, is just such a species of argument. There is also a tendency among students who are learning to use argumentation techniques in introductory logic courses, once they have learned some tools to analyze and evaluate arguments, to see any text of discourse they are given as expressing an argument. This can be a problem. The student who treats an explanation as an erroneous argument committing a fallacy, for example the fallacy of arguing in a circle, when the argument is really an explanation, has committed an error by misapplying logic.

Logic textbooks attempt to solve this problem by offering a pragmatic test to determine, in a given case, whether a passage expresses an argument or an explanation, namely by looking at how the discourse is being used in the given case. If it is being used to prove something that is in doubt, it is an argument. If it is being used to convey understanding of something that does not make sense or is incomprehensible, it is an explanation. The focus of this way of drawing the distinction is on the proposition or event that is

[1] Argumentation Research Group, University of Dundee, United Kingdom.
[2] Centre for Research in Reasoning, Argumentation and Rhetoric (CRRAR), University of Windsor, Canada.

to be explained or proved. If it is not subject to doubt (e.g. it is generally accepted as true, or can be taken for granted as true), the bit of text in question should be taken as an explanation. If it is subject to doubt, that is, if it is unsettled whether it is true or not, then the bit of text in question should be taken as an argument.

Let's look at two examples of explanations cited in the most widely used logic textbook [14, p. 19]. Here is the first one: the Challenger spacecraft exploded after liftoff because an O-ring failed in one of the booster rockets. Classifying this assertion as an argument or an explanation depends on whether the statement that the Challenger spacecraft exploded after liftoff should be taken as a statement that is accepted as factual or whether it should be taken to be a statement that is subject to doubt and that requires proof, or at least some supporting evidence, before it is accepted. The statement that the O-ring failed is not being used to prove the statement that the spacecraft exploded. That the spacecraft exploded is not in doubt. Most of us graphically remember seeing the exploding spacecraft on TV. The passage quoted above is not trying to prove that statement by providing evidence or reasons that support or imply it. The passage assumes that it is an accepted matter of fact that the spacecraft exploded, and is trying to show why it exploded. So the passage contains an explanation, as opposed to an argument. Because it is generally taken as common knowledge that the Challenger spacecraft exploded after liftoff, the whole causal statement is taken as an explanation.

The same principle applies to the second example: cows can digest grass, while humans cannot, because their digestive systems contain enzymes not found in humans. Should we take it as an accepted fact that cows can digest grass while humans cannot, or should we take this statement is subject to doubt and something that needs to be proved before it can be accepted? Again, it seems fairly plausible that the statement that cows can digest grass while humans cannot is generally accepted as part of common knowledge. If so it doesn't need to be proved, and the compound statement joined by the causal 'because' connective should be taken as an explanation.

We need to be aware, however, that this distinction based on common knowledge is not the only criterion required to distinguish arguments from explanations in a natural language text of discourse. Another part of the evidence or the so-called indicator words, like 'therefore', 'since', 'accordingly', and so forth. The problem is that the same indicator words are often used with respect to both arguments and explanations. Hence in any individual case one has to look carefully at the details of the actual text of discourse in the given case.

In the context of argumentation, premises are offered as proof of a conclusion or a claim, often in order to persuade someone or settle an issue that is subject to doubt or disputation. A number of computational models of argumentation have emerged and matured in the past twenty-or-so years [20] and the computational aspects of the dialectics of argument and of the structure of argument are well understood (cf. [19]).

In the context of explanation, the *explananda* (facts to be explained) are explained by a coherent set of *explanans* (facts that explain). The usual purpose of explanation is not necessarily to convince someone but rather to help someone understand why the explananda are the case. Computational models for explanation are mainly based on the technique of abductive (model-based) reasoning, which has been studied in the context of medical and system diagnosis [9]; other examples of computational explanation are [8], which models explanatory dialogues, and [24], which uses explanations for natural language understanding.

Despite the interest in dialogue treatments of explanation, the formal dialectical systems deriving from the early work of Hamblin treat only arguments. In Hamblin's '*Why-Because System with Questions'* [12, pp. 265-276], there are two participants who take turns making moves following syntactical rules (protocols). For example, when one party asks the question 'Why A?', the other party must reply with one of three speech acts: Assertion A; No commitment A; Statements B, B → A (where → represents the material conditional of propositional calculus). The language is that of propositional calculus, but it could be any other logical system with a finite set of atomic statements [12. p. 265]. As each party moves, statements are either inserted into or retracted from its commitment set of the party who made the move. A record of each party's commitments is kept and updated at each next move. On Hamblin's account, "a speaker is committed to a statement when he makes it himself, or agrees to it as made by someone else, or if he makes or agrees to other statements from which it clearly follows" [13, p. 136]. Interestingly, a why-question can only be a request for the other to present an argument, never an explanation.

Despite the important role explanations can play in argumentative dialogue, there have not been many attempts to combine argumentation and explanation into one formal model. Perhaps the most thorough work thus far is [1][4], in which arguments in the framework of [19] are combined with abductive-causal reasoning based on standard models of explanation [9] in one *hybrid theory*. The basic idea of this hybrid approach is as follows. A logical model of abductive-causal reasoning takes as input a causal theory (a set of causal rules) and a set of observations that has to be explained, the explananda, and produces as output a set of hypotheses that explain the explananda in terms of the causal theory. Arguments can be used to support and attack stories, and these arguments can themselves be attacked and defeated. Thus, it is possible to reason about, for example, the extent to which an explanation conforms to the evidence. This is important when comparing explanations: the explanation that is best supported and least falsified by arguments is, ceteris paribus, the best explanation.

2.1 Argumentation and explanation in dialogue

Dialogues consist of a series of locutions or utterances made by the participants. As a simple example of a dialogue, take the following exchange between Allen and Beth.

(1) Allen: The Evanston City Council should make it illegal to tear down the city's old warehouses.
(2) Beth: What's the justification for preserving them?
(3) Allen: The warehouses are valuable architecturally.
(4) Beth: Why are they so valuable?
(5) Allen: The older buildings lend the town its distinctive character.

During a dialogue, the participants construct and navigate an underlying reasoning structure [23], a static rendition of the claims, arguments and explanations proposed. For example, in the above dialogue one of the arguments made is ‘The warehouses are architecturally valuable *therefore* the Evanston city council should make it illegal to tear them down’. The link between a dialogue and this underlying structure can be explained by combining speech act theory [26] with Hamblin-style dialogue theory. A speech act can be analyzed as a locutionary act (the actual utterance, e.g. ‘What’s the justification for preserving them?’), but also as an illocutionary act which consists of the illocutionary force, meaning that it functions a kind of move in a dialogue. For example, one may include p in different kinds of moves like asserting p, asking p, challenging p, promising p and so on. In our example, speech acts (1) and (2) have the same propositional content, namely ‘The Evanston City Council should make it illegal to tear down the city’s old warehouses’. The illocutionary force, however, differs between (1) and (2): where (1) is uttered with the intention of *asserting* ‘The Evanston City Council should make it illegal to tear down the city’s old warehouses’, (2) can be seen as an instance of *requesting an argument* for this sentence. Figure 1 shows the example dialogue at the top, which is connected to the underlying reasoning structure via illocutionary relations.

There are different types of dialogue [30], each with a different goal. In persuasion dialogues, for example, one of the players makes a claim which he had to defend, while the other player’s goal is to dispute this claim. Another example of a dialogue type is *inquiry dialogue* [30], the aim of which is to increase knowledge. The participants in such a dialogue collectively gather, organize and assess hypothetical explanations and evidence for and against these explanations. Hence, Walton [28] identifies both explanation and argumentation as functions of an inquiry dialogue. Aleven [1] has defined an inquiry dialogue based on the hybrid theory in which the participants build explanations and then support and critically analyze these explanations using arguments. In this type of dialogue, the participants collectively build a hybrid theory of explanations and arguments.

2.2 The problem of distinguishing argumentation and explanation

The very first problem in attempting to analyze the concept of an explanation is to attempt to provide criteria to determine when some piece of discourse that looks like it could be either an explanation or an argument should be taken to fit into one category or the other. One possible way of distinguishing between argumentation and explanation might be to look at the product of our reasoning, that is, the underlying reasoning structure. At first sight, it often seems an explanation is abductive and causal whilst an argument is modus-ponens style, non-causal reasoning. The basic idea of causal abductive inference is that if we have a general rule $p \rightarrow_c q$, meaning p causes q, and we observe q, we are allowed to infer p as a possible explanation of q. In contrast, argumentation is often seen as reasoning from a premise p to a conclusion q through an inference rule $p \rightarrow_c q$, where this rule need not necessarily be causal. However, as it turns out it is also possible to give abductive or causal *arguments* (cf. [31]; causal argument). Similarly, one may perform explanatory reasoning by taking a rule $q \rightarrow_c p$, meaning *q is evidence for p* (see [4] for a discussion on evidential and causal reasoning).

As was previously argued in [5], argument and explanation can only be properly distinguished by looking at the dialogical context of reasoning. In order to determine this context, we need not just look at the original intention of the speaker (i.e. the illocutionary force of a speech act) but also at the broader dialogical context, such as the utterance that was replied to by the speaker and the intentions of the other participants. Consider the example in Figure 1. Allen makes his first move by asserting that the old warehouses should be preserved, and then Beth asks for a justification for this claim. Here it is clear that Beth is requesting an argument to justify Allen’s claim. Allen then provides this, but then Beth asks him the why-question: why are they so valuable? The speech act could be interpreted as requesting either an argument (challenging) or an explanation (Figure 1). Allen’s first reply to a challenge constitutes an argument but Allen’s second reply is ambiguous.

2.2.1 Circular Arguments and Explanations

Circular reasoning has long been a concern in logic. The fallacy of arguing in a circle has been included under the heading of informal fallacies in logic textbooks since the time of Aristotle [12]. But circularity is not been concerned exclusively with respect to arguments. Circular explanations are often condemned by the logic textbooks as unhelpful and confusing. But the reasons for condemning circular explanations are different from those for condemning circular argumentation [27].

The fallacy of arguing in a circle, or begging the question, is committed by an instance of circular reasoning that fails to work as an argument supposed to prove the conclusion that is in doubt. A

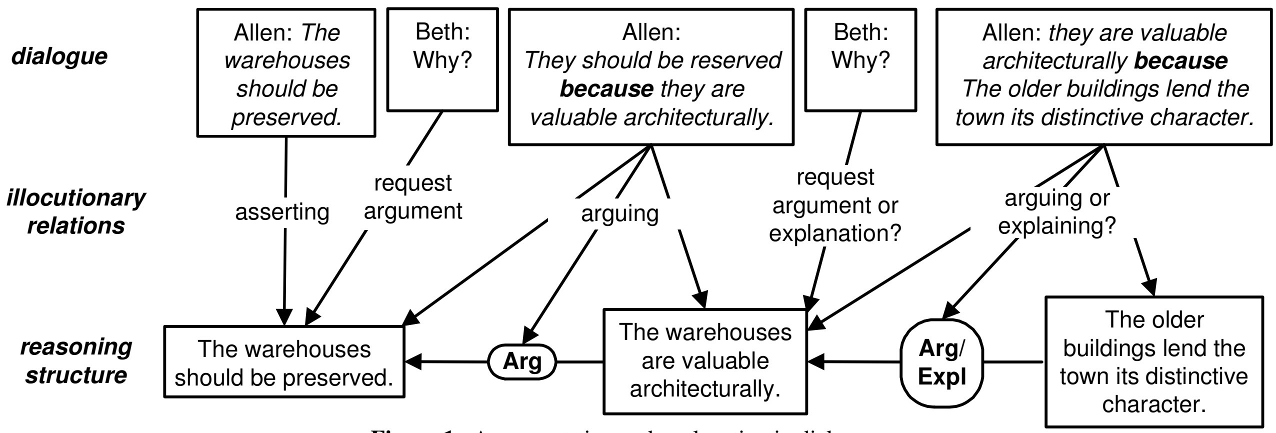


Figure 1. Argumentation and explanation in dialogue.

standard textbook example is provided by the following short dialogue between a man, Smith, and his bank manager.

Manager: Can you give me a credit reference?
Smith: My friend Jones will vouch for me.
Manager: How do we know *he* can be trusted?
Smith: Oh, I assure you he can.

Here we can detect a sequence of circular reasoning. The trustworthiness of Smith is supposed to depend on the testimony of his friend Jones, but the trustworthiness of Jones depends on the testimony of his friend Smith. This obviously won't work because of the circularity in the procedure of providing evidence to support a claim in an argument. If Jones's trustworthiness can be vouched for by some source independent of Smith, then the argument would work, and would no longer commit the fallacy of begging the question. In this kind of case, we cannot prove claim q by relying on premise p and then try prove p by backing it up by using q as a premise. It does not follow, however, that all circular arguments are fallacious as we now indicate.

To extend the example a bit further, suppose that a third-party could vouch for Jones, and that the trustworthiness of this third party is not dependent on the trustworthiness of either Smith or Jones. Then there would still be a circle in the argumentation structure, as shown in Figure 2, but the two text boxes on the right function as premises in a linked argument supporting the trustworthiness of Jones. This new argument gives us a way of breaking out of the circle that we were locked into in the previous argument represented by the dialogue above. The argumentation as a whole shown in Figure 2 has a circle in it, but when evaluated a whole it does not commit the fallacy of begging the question.

The problem with real cases where the fallacy of begging the question is a serious danger is that the circle is embedded in a text where it may be mixed in with much other discourse. This danger becomes even more serious when the discourse combines argumentation with explanation. But if you can find such a circle in an argument, it represents quite a serious criticism of that argument. A rational argument used to persuade a respondent to accept its conclusion must not be based on premises that can only be accepted if part of the evidence for one of these premises depends on the prior acceptance of the conclusion itself. If, so the argument is useless to prove the conclusion. The argument lacks what has been called a probative function [27].

The situation is different for explanations. They need to be evaluated in a different way. When a circular explanation is fallacious it is because it is uninformative or useless in transferring understanding. As with arguments, however, an explanation can be circular, but still be useful as an explanation. One reason is that there are feedback processes in nature, and to explain what is happening, the account given needs to go in a circle. For example, the more overweight a diabetic gets, the more insulin is produced in his blood, but the more insulin there is in his blood, the more he eats, and the more he becomes overweight. In this vicious circle, the problem becomes worse and worse by a continual process of feedback that escalates it. To understand that the process is circular helps to explain the whole picture of what is going on.

Let us return to our warehouse dialogue from section 2.1. First, let us assume that Allen's reply (5) is a speech act of arguing that creates an argument '*the older buildings lend the town its distinctive character* ***so*** *the warehouses are valuable architecturally'* (Figure 3). Now extend the dialogue as follows:

(6) Beth: OK agreed. But why do the older buildings lend the town its distinctive character?
(7) Allen: The warehouses are valuable architecturally.

When examining this dialogue we might be suspicious about the possibility that it contains the fallacy of begging the question. After all, when Allen is asked by Beth about the justification for preserving the old warehouses (4), Allen replies that the warehouses are valuable architecturally (5). But then later, at his last move in the dialogue (7), he reverts back to making the same statement again. It definitely appears that the dialogue is circular. The question then is whether the circularity is benign or vicious.

Let's interpret Beth's question (6) as a request for explanation. Now the reasoning in the dialogue is no longer just a sequence of argumentation, but a mixture of argumentation and explanation (Figure 3). In order to prove his claim that the warehouses are valuable architecturally, Allen has used the premise that the older buildings lend the town its distinctive character. But then he has used the former as an explanation to help Beth understand the latter. The sequence of replies is then circular but not fallacious. Allen is merely explaining why the older buildings lend the town its distinctive character. Since Beth has agreed to this proposition, Allen does not need to prove it, and so there is no interdependency in the sequence of argumentation of the kind required for the committing of the fallacy of begging the question. There is no failure to fulfill the probative function of the kind that signals circular reasoning of a kind associated with committing the fallacy of begging the question. Allen is not using premise p to prove conclusion q and then using q as a premise required to prove p.

This is an unusually subtle case to disentangle. There is a circularity there, but it is benign one where the explanation fits into the argumentation in a way that is not an obstruction to the dialogue. The circularity could help Beth to understand the situation. So it does have a legitimate function. There is circular reasoning, but no circular argumentation.

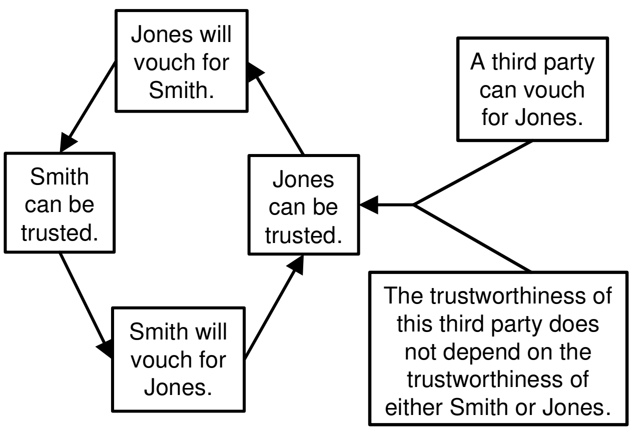


Figure 2. Circular Reasoning in the Credit Reference Example.

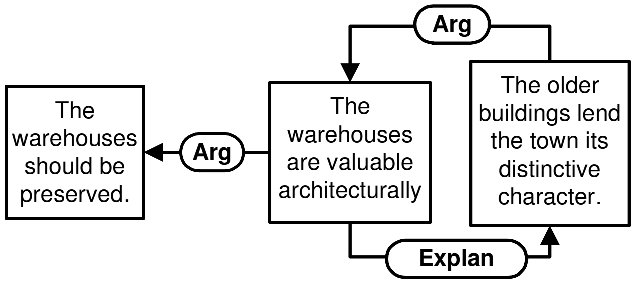


Figure 3. Mixed Version of the Warehouse Example

3 DEFINING EXPLANATION IN DIALOGUE

How then, given the text of discourse, are we to determine whether the text is better taken to represent an argument or an explanation? The test widely adopted in logic textbooks uses the distinction between an accepted fact and a disputed claim was discussed in section 2. But we need to go even beyond that and look more broadly at how arguments and explanations function as different kinds of moves in a dialogue. An argument is a speech act used to *convince* the hearer of some unsettled claim and an explanation is a speech act used to help the hearer to *understand* something. This distinction can be drawn as one of a difference of purpose of discourse. Since the distinction is drawn this way, it can be seen to be based on a dialogue model of communication in which two parties take turns in putting forward speech acts. As argued above, in order to then determine whether something is an argument or an explanation, we need not just look at the original intention of the speaker (i.e. the illocutionary force of a speech act) but also at the broader dialogical context.

Defining explanation as a speech act put forward with the aim of transferring understanding from an explainer to an explainee raises further questions. What is understanding, and how can it be transferred from one party to another? Research in AI and cognitive science shows that communicative agents understand the actions of other agents because they share "common knowledge" of the way things can normally be expected to proceed in familiar situations in everyday life. This common knowledge can be modeled as explanation schemes or *scripts* [24]. An explanation scheme is a generic scenario, an abstract rendering of a sequence of actions or events of a kind. For example, the restaurant-script contains information about the standard sequence(s) of events that take place when somebody goes to dine in a restaurant.

Explanation schemes can be instantiated by particular explanations and thus the scheme provides the conditions for the explanation's coherence [1]. Take, for example, a man who enters a restaurant, orders a hamburger and then removes his pants and offers the waiter his pants. This particular story is incoherent, because it does not adhere to the typical restaurant scheme. But if this story fits another explanation scheme it can still be coherent. Suppose information is added that the waiter spilled hot soup on the man's legs. This new information would fill out the story in such a way that it hangs together as a coherent script about what happens when someone spills hot liquid on one's clothes. Thus, an explanation may be causal, motivational, teleological, and so on.

A dialogue model of explanation can then be constructed by building it around the notion of the mutual comprehensibility of a story, or connected sequence of events or actions that both parties can at least partially grasp in virtue of their common knowledge about the ways things can be generally expected to happen in situations they are both familiar with. This is the route taken by Schank and his colleagues in cognitive science (cf. [24]). According to them, explanation is a transfer of understanding from one party to another in a dialogue, where understanding is clarified scripts, "frozen inference chains stored in memory". On Schank's theory, failures of understanding of kinds that trigger a need for an explanation occur because of an anomaly, a gap in a story that contains a part where it fails to make sense, or even where the whole story fails to make sense because it does not "add up". An explanation, on this approach, is a repair process used to help someone account for the anomaly by using scripts that could be taken from script libraries.

3.1 A Dialogue System for Argument and Explanation

We now propose an example of a dialogue system for argumentation and explanation, based on the protocols presented by [6][29]. Our dialogue system consists of a *communication language* that defines the possible speech acts in a dialogue, a *protocol* that specifies the allowed moves at any point in the dialogue, *commitment rules*, which specify the effects of a speech act on the propositional commitments of the dialogue participants. Furthermore, we assume that both players have their own separate knowledge bases containing argumentation schemes and explanation schemes, which form the basis of arguments and explanations proposed in the dialogue [22].

In a game for argumentation and explanation, essentially two types of dialogue are combined: explanation dialogue [29][17][8] and examination dialogue [10]. In a pure explanation dialogue the explainer is trying to transfer understanding to the explainee; an examination dialogue can be used to test (evaluate) an explanation. Examination dialogues are more adversarial. For example, the answerer's inconsistency in previous replies can be attacked using probing counter-arguments to test his trustworthiness (for example, as a witness). Figure 4 shows the combination of explanation and examination dialogues as a process.

The speech acts of a game for explanation and argumentation are presented in the typical format $F\ p$, where F is the illocutionary

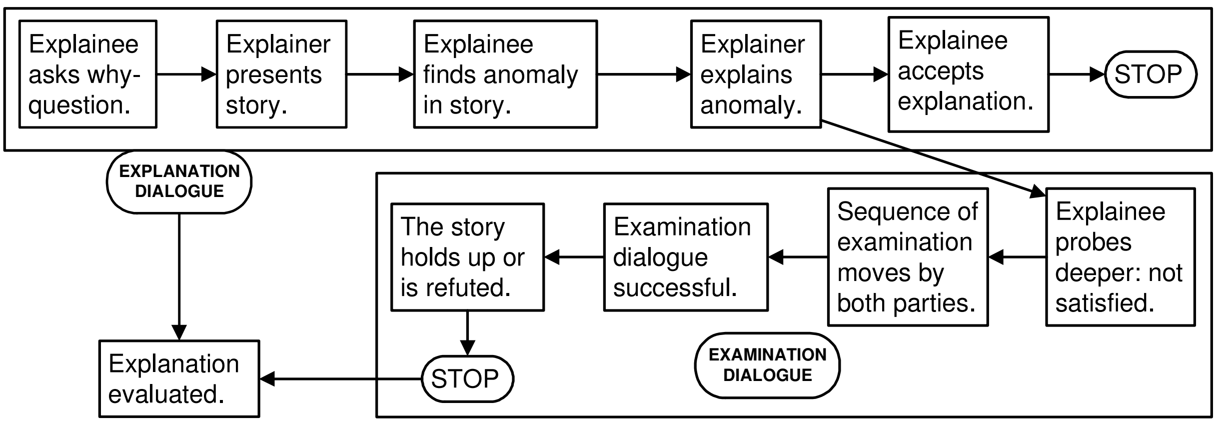


Figure 4. Explanation and examination dialogues combined.

force and p is the propositional content.

1. *claim* φ. The player claims a proposition φ.
2. *argue* ψ *because* φ. The player states an argument ψ *because* φ based on an argumentation scheme S_A from the player's knowledge base.
3. *challenge* φ. The player asks for an argument for φ.
4. *concede* φ. The player admits that proposition φ is the case.
5. *retract* φ. The player declares that he is not committed (any more) to φ.

These speech acts are standard in systems for argumentative dialogue (cf. [16]). Now, for explanation we need other speech acts, as defined by [6][29].

6. *explain* ψ *because* φ. The player provides an explanation ψ *because* φ based on an explanation scheme S_E from the player's knowledge base.
7. *explanation request* φ. The player asks for an explanation of φ.
8. *inability to explain* φ. The player indicates that he cannot explain φ.
9. *positive response*: The player indicates that he understands an explanation.
10. *negative response*: The player indicates that he does not understand an explanation.

Note that with explanation, the issue is not whether a player is convinced (i.e. wants to be committed to a proposition) but rather whether he understands a proposition.

Commitment rules specify the effect of moving one of the speech acts. A player becomes committed to any claim, argument or explanation he puts forward, and also to any claim he concedes to. Commitments can be retracted by the *retract* speech act.

The following standard protocol rules are part of the dialogue system (cf. [28]).

1. The players each take their turn.
2. The players cannot move the exact same speech act twice.
3. Players cannot commit to propositions which would make their commitments inconsistent.
4. Players are only allowed to *argue* for propositions to which they are committed but the other player is not.
5. Players are only allowed to *argue* against propositions to which the other player is committed and they are not.
6. A *challenge* φ move may only follow either a *claim* φ move or an *argue* ψ *because* φ move.
7. A *challenge* φ move can only be responded to by either an *argue* φ *because* ψ move or a *retract* φ move.
8. Players are only allowed to *challenge* propositions to which the other player is committed and they are not.
9. Players can only *concede* to propositions to which the other player is committed.
10. Players can only *retract* propositions to which they are committed.

The above rules capture the basics of argumentative dialogue. The rules encapsulate the idea that argumentation is an activity aimed at proving (or disproving) some claim: once both parties are committed to a claim, there is no point in arguing any further.

For explanation the rules are different, as explanation is aimed at improving understanding. Both parties can be committed to a claim, but one of the two may not fully understand it.

11. Players are only allowed to request explanations of propositions to which both players are committed.
12. Players are only allowed to request explanations of propositions for which they themselves do not have an explanation scheme in their knowledge base.
13. A *request explanation* φ move can only be responded to by an *explain* φ *because* ψ move or an *inability to explain* φ move.
14. Players are only allowed to explain propositions to which both players are committed.
15. Players are only allowed to explain propositions for which they have an explanation scheme in their knowledge base and the other party does not.
16. An explain move is always followed by either a *positive response* or a *negative response*.

Note how explaining is in a sense analogous to arguing but with a different aim, namely making someone understand a proposition instead of committing them to it.

The system can be applied to the two examples taken from the logic textbook [14], the Challenger spacecraft example and the example about the digestive system of a cow. These are classified as explanations because of the rules stating that players are only allowed to argue for or against propositions to which the other player is not committed. In the one example is taken as common knowledge that the Challenger spacecraft exploded after liftoff. In the other example, it is taken to be common knowledge that cows can digest grass while humans cannot. Therefore both parties can be taken to be committed to both these propositions. Hence in both examples, it would be inappropriate for either party to argue either for or against these propositions. However it would be appropriate for either party to offer an explanation.

Briefly, it can be shown how a script is involved in the spacecraft example as follows. To make the explanation successful the party to whom it was directed must have enough general knowledge about how rockets work, how a rocket can explode, and to connect an O-ring failure to a leakage of fuel. There must also be knowledge about what might normally be expected to happen when a fuel leak occurs during the operation of the rocket motor. The receiver of the explanation must also know that the booster rockets are attached to the spacecraft in such a way that if the booster rocket explodes, the whole spacecraft that is attached to it will also explode. To connect all these events into a coherent script that explains how the spacecraft exploded after liftoff the receiver of the explanation must already have the common knowledge required to understand how this series of events and objects is connected up into a coherent story.

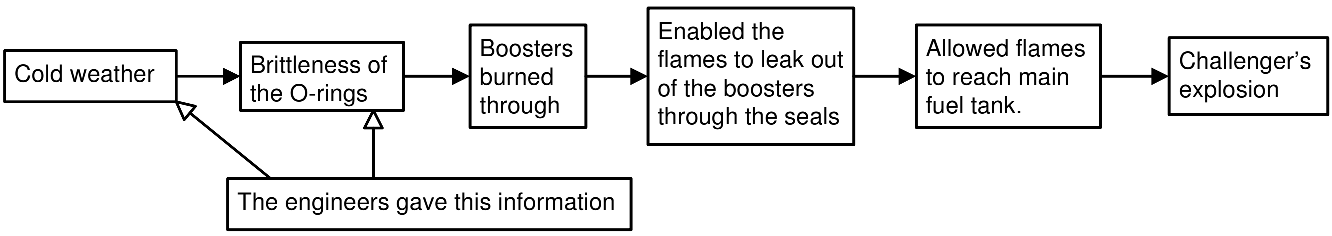


Figure 5. Explanation supported by evidence

How the system applies to the example dialogue about the warehouses is indicated in Figure 1 in the account given of the illocutionary relations in that figure. The evidence for classifying moves as arguments or explanations is indecisive in the instance where Beth asks Allen the question 'Why are the warehouses so valuable?' As noted, the speech act could be interpreted as requesting either an argument or an explanation. There was another ambiguous speech act when Beth asks Allen why the warehouses are so valuable. This speech act could be interpreted as requesting either an argument with explanation, as noted in the discussion of the case in section 2.2. The system manages these cases by analyzing them as instances where the evidence given in the dialogue exchange is insufficient to classify the speech act as either an argument on explanation. The system needs to then follow up by shifting to an examination dialogue where the dialogue participant who asked the question needs to be examined and must indicate whether he or she is putting forward the speech act as an argument on explanation. In many instances, especially the short one like those found in the logic textbooks, the text of the case is merely given, and there is no possibility of examining the questioner. In such cases we need to make a determination based on the given textual and contextual evidence. It is our contention that this determination needs to be made in the framework provided by our hybrid system of dialogue for argument and explanation.

4 RELATED RESEARCH

We have presented only relatively simple examples, or at any rate short ones, that can fit the space confines of this paper. However, we would suggest as a project for further research applying the dialogue system comprising both arguments and explanations to longer examples of dialogues of the kind that can already be found in the literature. This literature is about explanation systems, but it could be helpful to re-examine the examples used in them, as well as other longer texts containing explanations, using this new system. In some instances applying our system to problematic cases where there are ambiguous instances of questions that could be requests for either explanations or arguments, participants will need to extend the dialogue by having a clarification dialogue used to deal with ambiguity.

In addition to the dialogue systems that combine argumentation and explanation as proposed in [6][29], there are numerous explanations systems that incorporate the ideas about transferring understanding through explanations. For example, ACCEPTER [15] is a computational system for story understanding, anomaly detection and explanation evaluation. In this system, explanations are directed towards filling knowledge gaps revealed by anomalies. Examples of explanations processed by ACCEPTER along the lines of the dialogue sequence above, include the death of a race horse, the explosion of the space shuttle Challenger, the recall of Audi 5000 cars for transmission problems, and an airliner that leaves from the wrong departure gate [15][38].

The schemas in ACCEPTER's memory are represented as MOPS (memory organization packages) representing stereotyped sequences of events. MOPS help an agent understand by providing expectations on how things can normally be expected to go in a familiar situation. MOPS are comparable to the stories used in the hybrid theory. A simplified version of the explanation of the explosion of the space shuttle Challenger modeled by Leake [15, 39-53] can be used to show how this example fits nicely into the way of treating explanations in the hybrid theory.

This version of the explanation [15, 39] can be summed up as follows. The boosters burned through, allowing flames to reach the main fuel tank, causing an explosion. According to the engineers, the explosion was caused by the booster seals being brittle and the cold weather. The explanation given is that the Challenger's explosion was caused by the flame in the booster rockets, and prior to that by the cold weather which was the cause of the brittleness of the O-rings which enabled the flames to leak out through the seals. This causal sequence can be displayed in the hybrid theory as shown in figure 5. The arrows with filled heads represent causal relations, while the arrows with white heads represent arguments.

The explanation given in the example in section 2 explained the Challenger explosion by presenting the story that the spacecraft exploded because the O-ring failed in one of the booster rockets. This story leaves out intervening causal steps made explicit in the fuller story represented in figure 5. Also, we see at the bottom left of figure 5, there was additional information given by testimony of the engineers. This testimony can be seen as an argument supporting the two initial items in the causal story sequence along the top and right. This supplemented explanation expands the story of what happened, yielding better understanding of why the Challenger explosion happened. It does this by filling further information in the causal sequence in the story and by adding in evidence supporting part of the story.

Cawsey's work [8] on computational generation of explanatory dialogue and Moore's dialogue-based analysis of explanation for advice-giving in expert systems [17] also took a dialogue approach. Moore defines explanation as an inherently incremental and interactive process that requires a dialogue between an explanation presenter who is trying to explain something and a questioner who has asked for an explanation.

An interesting piece of related research is [3], which uses scripts or story schemes to model cases about the facts. These cases can then be argued with using the argumentative moves of CATO [1], which were originally developed for reasoning with legal cases. What this means is that [3] have a skeleton dialogue system that uses scripts to perform argumentation instead of explanation. This conforms with our findings: it is not the logical structure of the reasoning or the schemes used in reasoning that determines whether something is explanation or argumentation but the context of the dialogue in which the reasoning is performed and the schemes are used.

5 CONCLUSION

In this paper, we have discussed the problem of distinguishing between argumentation and explanation. In many cases, the same piece of discourse can reasonably be interpreted as either an explanation or an argument, and the logical structure of the reasoning proposed also does not conclusively distinguish between the two. The distinction is important for several reasons. First, there are situations in which argumentation may be fallacious whilst explanation is not, as illustrated by our examples of circular reasoning in section 2.2.1. Second, explanation and argumentation serve different aims and it is important that there is no confusion in multi-agent dialogue; if a request for explanation is interpreted as a request for argumentation, this may lead to undesirable misunderstandings and unwanted behaviour by agents. We have shown that such confusions can easily lead to the committing of logical fallacies. The illustration we have used to make this point is the specific fallacy of begging the question, also known as arguing in a circle. Finally, the distinction is important for the connection between argumentation, story-based explanation and discourse analysis, as argumentation schemes and explanation schemes can play important roles in the analysis of natural language texts [21][11].

Our solution involves looking at the context of dialogue to determine whether reasoning is argumentation or explanation. Whether something is argumentation or explanation is determined by the intention of uttering a locution, and this intention can be inferred from the context of the dialogue, such as the speech act that was replied to and the knowledge and intentions of the other players. This context of dialogue can be modeled as a dialogue system (section 3). In this sense, our dialogue system for argumentation and explanation does not only provide normative rules for coherent dialogue (as is usual), but it also helps us describe the difference between argumentation and explanation in dialogue.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of EPSRC under grant EP/G060347/1 for Floris Bex and the support of the Social Sciences and Humanities Research Council of Canada for Insight Grant 435-2012-0101 for Douglas Walton.

REFERENCES

[1] V. Aleven. *Teaching Case Based Argumentation Through an Example and Models*. PhD Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 1997.

[2] F. Bex, *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Dordrecht: Springer, 2011.

[3] F. Bex, T. Bench-Capon, and B.Verheij, 'What makes a story plausible? The need for precedents', in *Legal Knowledge and Information Systems. JURIX 2011: The Twenty-Fourth Annual Conference*, ed., K.D. Atkinson, pp. 23–32, (2011).

[4] F.J. Bex, P. J. van Koppen, H. Prakken, and B. Verheij. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law*, 18:2 (2010).

[5] F.J. Bex and K. Budzynska. Argument and explanation as contexts of reasoning. 10th Workshop on Computational Models of Natural Argument (CMNA 2010).

[6] F.J. Bex and H. Prakken. Investigating Stories in a Formal Dialogue Game. *Computational Models of Argument. Proceedings of COMMA 2008. Frontiers in Artificial Intelligence and Applications* 172:73-84 (2008). IOS Press, Amsterdam.

[7] J. Cassens & A. Kofod-Petersen, 'Designing Explanation-Aware Systems: The Quest for Explanation Patterns', *Explanation-Aware Computing: Papers from the 2007 AAAI Workshop.* In T.R. Roth-Berghofer, S. Schulz, and D.B. Leake, editors, Technical Report WS-07-06, Menlo Park, California, AAAI Press, 2007, 20-27.

[8] A. Cawsey, *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*, Cambridge, Mass., MIT Press, 1992.

[9] L. Console and P. Torasso. A spectrum of logical definitions of model–based diagnosis. *Computational Intelligence* (2010) 7:3, 133 – 141.

[10] Dunne, P. E., Doutre, S. and Bench-Capon, T. J. M. (2005). Discovering Inconsistency through Examination Dialogues. *Proceedings IJCAI-05*, Edinburgh, 1560-1561.

[11] M.A. Finlayson. Deriving Narrative Morphologies via Analogical Story Merging, in *New Frontiers in Analogy Research* (2009) New Bulgarian University Press, Sofia, pp. 127-136.

[12] C. L. Hamblin. *Fallacies*, London, Methuen, 1970.

[13] Hamblin, C. L. (1971). Mathematical Models of Dialogue. *Theoria* 37, 130-155.

[14] P. Hurley, *Logic: A Concise Introduction*, Belmont, California, Wadsworth, 2003.

[15] D.B. Leake, *Evaluating Explanations*, Hillsdale, New Jersey, Erlbaum, 1992.

[16] P. McBurney and S. Parsons. Dialogue games for agent argumentation. In I. Rahwan and G. Simari (eds.): *Argumentation in Artificial Intelligence* (2009) Berlin, Germany: Springer, pp. 261-280.

[17] J. D. Moore, *Participating in Explanatory Dialogues*, Cambridge, Mass., MIT Press, 1995.

[18] D. Poole, 'Learning, bayesian probability, graphical models, and abduction', in *Abduction and Induction: essays on their relation and integration*, eds., P. Flach and A. Kakas, Kluwer, (1998).

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

[20] 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).

[21] C. Reed, R. Mochales Palau, G. Rowe, and M.-F. Moens, Language Resources for Studying Argument. *Proceedings of the 6th Language Resources and Evaluation Conference (LREC-2008)*, 2008.

[22] C. Reed and D. Walton. Argumentation schemes in dialogue. In H.V Hansen, C.W. Tindale, R.H. Johnson, and J.A. Blair, editors, Dissensus and the Search for Common Ground (Proceedings of OSSA), 2007.

[23] C. Reed, S. Wells, K. Budzynska and J. Devereux. Building arguments with argumentation: the role of illocutionary force in computational models of argument. *Computational Models of Argument: Proceedings of COMMA 2010*, IOS Press, Amsterdam (2010), 415 – 426.

[24] L.J. Rips. Circular Reasoning. *Cognitive Science*, 26, 2002, 767-795.

[25] R.C. Schank, *Explanation Patterns: Understanding Mechanically and Creatively*, Hillsdale, New Jersey, Erlbaum, 1986.

[26] J.R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

[27] D. Walton. Are Circular Arguments Necessarily Vicious? *American Philosophical Quarterly*, 22, 1985, 263-274

[28] D.N. Walton. The New Dialectic: Conversational Contexts of Argument, University of Toronto Press, Toronto, 1998.

[29] D. Walton. A Dialogue System Specification for Explanation, *Synthese* (2011), 182(3), 349-374.

[30] D. Walton and E.C.W. Krabbe. *Commitment in Dialogue*. Albany: SUNY Press, 1995.

[31] D. Walton, C.A. Reed and F. Macagno. *Argumentation Schemes.* Cambridge: Cambridge University Press, 2008.

Defining the structure of arguments with AI models of argumentation

Bin Wei[1] and **Henry Prakken**[2]

Abstract. The structure of arguments is an important issue in the field of informal logic and argumentation theory. In this paper we discuss how the ‘standard approach’ of Walton, Freeman and others can be analysed from a formal perspective. We use the *ASPIC*$^+$ framework for making the standard model of argument structure complete and for introducing a distinction between types of individual arguments and types of argument structures. We then show that Vorobej’s extension of the standard model with a new type of hybrid arguments is not needed if our formal approach is adopted.

1 Introduction

The structure of arguments is an important issue in the field of informal logic and argumentation theory. The main issue is to define the different ways in which premises and conclusions can be combined to generate different structural argument types. The ‘standard approach’ was introduced by Stephen N. Thomas in [1] and was further developed by, among others, Walton [2] and Freeman [3]. Vorobej [4] extended their approach with an additional argument type called “hybrid arguments”. This paper aims to show how formal AI models of argumentation can be used to further extend and clarify these informal models of the structure of arguments. In particular, we argue that they have some limitations, since their classifications are incomplete and since they do not distinguish between types of individual arguments and structures consisting of several arguments. Moreover, we argue that Vorobej’s proposal can be clarified by making a distinction between deductive and defeasible arguments.

We aim to achieve our aims by applying the *ASPIC*$^+$ framework of [6]. We use it to make three specific contributions: (1) to make the standard classifications complete; to (2) indicate and explain why convergent and divergent arguments are not arguments but argument structures; and (3) to indicate and explain why Vorobej’s class of hybrid arguments is not needed if an explicit distinction is made between deductive and defeasible arguments.

This paper is organized as follows. In section 2, we introduce the standard informal model of argument structure and Vorobej’s [4] extension with so-called hybrid arguments. In Section 3, we present a simplified version of the *ASPIC*$^+$ framework. We then use this framework in section 4 to complete the standard model and to distinguish between types and structures of arguments. In section 5, we discuss Vorobej’s notion of hybrid arguments and how it can be captured in *ASPIC*$^+$. Section 6 concludes the paper.

[1] The Institute of Logic and Cognition, Sun Yat-Sen University, Guangzhou, China. email: srsysj@gmail.com
[2] Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands

2 Standard approaches to argument structure

We first introduce the main approaches to argument structures, notably the approach by e.g. Walton [2] and Freeman [3], which we will call the ‘standard’ approach and Vorobej’s [4] extension with so-called hybrid arguments.

2.1 Standard approach

The standard approach to the structure of arguments was introduced by Stephen N. Thomas in [1]. He divided the arguments into (1) *linked arguments*, which means that every premise is dependent on the others to support the conclusion, (2) *convergent arguments*, which means that premises support the conclusion individually, (3) *divergent arguments*, which means that one premise supports two or more conclusions, and (4) *serial arguments*, which means that one premise supports a conclusion which supports another conclusion.

Walton then further discussed the structure of arguments in [2]. We present the informal definitions of the concepts of structures of arguments according to his latest description in [5]. The corresponding diagrams are shown in Figure 1.

Definition 1. *The types of arguments are informally defined as follows:*

(1) An argument is a single argument *iff it has only one premise to give a reason to support the conclusion.*
(2) An argument is a convergent argument *iff there is more than one premise and where each premise functions separately as a reason to support the conclusion.*
(3) An argument is a linked argument *iff the premises function together to give a reason to support the conclusion.*
(4) An argument is a serial argument *iff there is a sequence* $\{A_1, \ldots, A_n\}$ *such that one proposition* A_i *acts as the conclusion drawn from other proposition* A_{i-1} *as premise and it also functions as a premise from which a new proposition* A_{i+1} *as conclusion is drawn.*
(5) An argument is a divergent argument *iff there are two or more propositions inferred as separate conclusions from the same premise.*
(6) An argument is a complex argument *iff it combines at least two arguments of types (2),(3),(4) or (5).*

Example 1. *Walton gives the following examples of, respectively, a convergent, divergent and linked argument:*

(1) (A) Tipping makes the party receiving the tip feel undignified; (B) Tipping leads to an underground, black-market economy; (C) Tipping is a bad practice.

(2) (A) *Smoking has been proved to be very dangerous to health; (B) Commercial advertisements for cigarettes should be banned; (C) Warnings that smoking is dangerous should be printed on all cigarette packages.*
(3) (A) *Birds fly; (B) Tweety is a bird; (C) Tweety flies.*

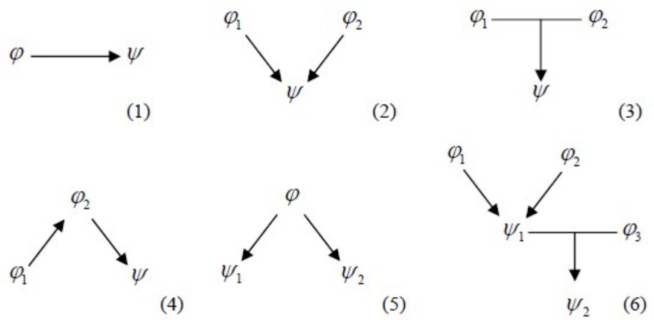


Figure 1. Structures of argument

In Example 1(1), the three statements form a convergent argument, since statements (A) and (B) function separately as a reason to support the conclusion (C). By contrast, in Example 1(2) these three statements form a divergent argument, since statement (B) and (C) are inferred as separate conclusions from the same premise (A). Finally, Example 1(3) is a linked argument, since neither premise alone gives any reason to accept the conclusion.

2.2 Hybrid arguments

In [4], Mark Vorobej argued that the standard approach needs to be extended with a class of hybrid arguments. To discucss this class, we must first present Vorobej's basic definitions of types of arguments.

Definition 2. *An argument A is:*

- simple *iff A has exactly one conclusion. Otherwise, A is* complex.
- convergent *iff A is simple and each premise in A is relevant to C, where relevance is treated as a primitive dyadic relation obtaining in each instance between a set of propositions and a single proposition.*

Definition 3. *A linked set and linked argument are defined as follows:*

- *A set of premises Δ forms a* linked set *iff*
  (1) Δ contains at least two members;
  (2) Δ is relevant to C, and
  (3) no proper subset of Δ is relevant to C.
- *An argument A is* linked *iff A is simple and each premise in A is a member of some linked set.*

Vorobej then motivates this new class of hybrid arguments with examples like the following one.

Example 2. *Consider example (F) as follows:*

- *(F): (1) All the ducks that I' ve seen on the pond are yellow. (2) I've seen all the ducks on the pond. (3) All the ducks on the pond are yellow.*

Vorobej observes that (2) in isolation is not relevant to (3), so this is not a convergent argument. Secondly, (1) is relevant to (3), so (1) is not a member of any linked set, so this is also not a linked argument. Vorobej regards (F) as a hybrid argument, since (1) is relevant to the conclusion (3) and (2) is not relevant to the conclusion (3) but (1) and (2) together provide an additional reason for (3), besides the reason provided by (1) alone.

Vorobej provides the following definition of hybrid arguments in terms of a relation of supplementation between premises.

Definition 4. *The relation of supplementation and hybrid argument are defined as follows:*

- *A set of premises Σ* supplements *a set of premises $\triangle$ iff*
  (1) Σ is not relevant to C;
  (2) Δ is relevant to C;
  (3) $\Sigma \cup \Delta$ offers an additional reason R in support of C, which Δ alone does not provide, and
  (4) Σ and Δ are the minimal sets yielding R which satisfy clauses (1),(2) and (3).
- *An argument A is a hybrid iff A is simple and contains at least one supplemented (or supplementing) set.*

In Example 2 premise (2) supplements premise (1). The argument is therefore a hybrid argument.

3 The *ASPIC*$^+$ framework

The *ASPIC*$^+$ framework of [6] models arguments as inference trees constructed by two types of inference rules, namely, strict and defeasible inference rules. The framework has in [6, 7, 8, 9] been shown to capture a number of other approaches to structured argumentation, such as assumption-based argumentation [10], forms of classical argumentation [11] and Carneades [12]. In this paper we use a simplified version of *ASPIC*$^+$ framework, with negation instead of an arbitrary contrariness function over the language, with just one instead of four types of premises, and without preferences.

Definition 5. *[Argumentation system] An* argumentation system *is a tuple $AS = (\mathcal{L}, \mathcal{R})$ where*

- *$\mathcal{L}$ is a logical language closed under negation ($\neg$). Below we write $\psi = -\varphi$ when either $\psi = \neg\varphi$ or $\varphi = \neg\psi$.*
- *$\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.*

Definition 6. *[Knowledge bases] A* knowledge base *in an argumentation system $(\mathcal{L}, \mathcal{R})$ is a set $\mathcal{K} \subseteq \mathcal{L}$.*

Arguments can be constructed step-by-step by chaining inference rules into trees. In what follows, for a given argument the function $Prem$ returns all its premises, $Conc$ returns its conclusion Sub returns all its sub-arguments, while $TopRule$ returns the last inference rule applied in the argument.

Definition 7. *[Argument] An* argument *A on the basis of a knowledge base $\mathcal{K}$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R})$ is:*

1. *φ if $\varphi \in \mathcal{K}$ with: $Prem(A) = \{\varphi\}$; $Conc(A) = \varphi$; $Sub(A) = \{\varphi\}$; $TopRule(A) =$ undefined.*
2. *$A_1, \ldots A_n \to/\Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $Conc(A_1), \ldots, Conc(A_n) \to/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$.*
   $Prem(A) = Prem(A_1) \cup \ldots \cup Prem(A_n)$; $Conc(A) = \psi$; $Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$; $TopRule(A) = Conc(A_1), \ldots, Conc(A_n) \to/\Rightarrow \psi$.

An argument is *strict* if all its inference rules are strict and *defeasible* otherwise.

Definition 8. *[Maximal proper subargument] Argument A is a* maximal proper subargument *of B iff A is a subargument of B and there does not exist any proper subargument C of B such that A is a proper subarugment of C.*

Example 3. *Consider a knowledge base in an argumentation system with $R_s = \{p, q \rightarrow s; u, v \rightarrow w\}$; $R_d = \{p \Rightarrow t; s, r, t \Rightarrow v\}$; $K = \{p, q, r, u\}$.*

An argument for w is displayed in Figure 2. Strict inferences are displayed with solid lines and defeasible inferences with dotted lines. Formally the argument and its subarguments are written as follows:

$A_1 : p$ $\quad A_6 : A_1, A_2 \rightarrow s$
$A_2 : q$ $\quad A_7 : A_3, A_4, A_6 \Rightarrow v$
$A_3 : r$ $\quad A_8 : A_5 \rightarrow n$
$A_4 : t$ $\quad A_9 : A_8 \Rightarrow u$
$A_5 : m$ $\quad A_{10} : A_7, A_9 \rightarrow w$

We have that
$Prem(A_{10}) = \{p, q, r, t, m\}$
$Conc(A_{10}) = w$
$Sub(A_{10}) = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}\}$
$MaxSub(A_{10}) = \{A_7, A_9\}$
$Toprule(A_{10}) = \{u, v \rightarrow w\}$

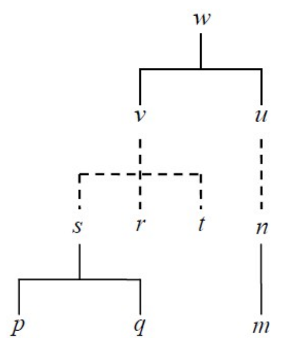


Figure 2. An Argument

In *ASPIC*$^+$ there are three syntactic forms of *attacks*: an *undercutter* attacks the inference rule, a *rebuttal* attacks the conclusion, and an *underminer* attacks a premise. Rebutting and undercutting attacks can only be targeted at (conclusions of) defeasible inference rules. So the argument in Figure 2 can only be rebutted on the (inferences of) the conclusions v and u. Attacks combined with preferences defined by an argument ordering yield three kinds of defeat. For the formal definitions of attack and defeat see [6].

4 Types and structures of argument

We now give a new classification of arguments in terms of the *ASPIC*$^+$ framework and then define so-called argument structures, which are collections of arguments with certain features. We first define two kinds of $unit$ arguments and then define several other argument notions consisting of these two $unit$ types in different ways. We finally define various structures of argument in terms of the various definitions of argument types.

Definition 9. *The types of arguments can be defined as follows:*

(1) An argument A is a $unit\ I$ argument iff A has the form $B \Rightarrow \psi$ and subargument B is an atomic argument $B : \varphi$. We call the inference rule $\varphi \Rightarrow \psi$ a $unit\ I$ inference.
(2) An argument A is a $unit\ II$ argument iff A has the form $B_1, \ldots, B_n \Rightarrow \psi$ and subarguments $A : B_1, \ldots, B_n$ are atomic arguments $B_1 : \varphi_1, \ldots, B_n : \varphi_n$. We call the inference rule $\varphi_1, \ldots, \varphi_n \Rightarrow \psi$ a $unit\ II$ inference.
(3) An argument A is a $multiple\ unit\ I$ argument iff all inferences $r_1, \ldots, r_n$ in the argument A are $unit\ I$ inferences.
(4) An argument A is a $multiple\ unit\ II$ argument iff all inferences $r_1, \ldots, r_n$ in the argument A are $unit\ II$ inferences.
(5) An argument A is a $mixed$ argument iff A has at least one $unit\ I$ subargument and $unit\ II$ subargument.

We display the diagrams of argument types in Figure 3. For simplicity, we assume $n = 2$ in these diagrams and show only one case of a $mixed$ argument.

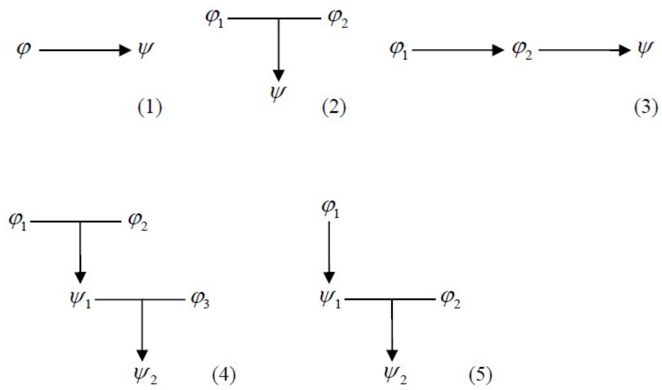


Figure 3. Argument types

Proposition 1. *Every argument is of exactly one argument type.*

Proof. Firstly, we prove the existence of an argument type by induction on the number of unit inferences. For $n = 1$, argument A corresponds to a $unit\ I$ argument. For $n = k > 1$, argument A corresponds to a $multiple\ unit\ I$ argument, a $multiple\ unit\ II$ argument, or a $mixed$ argument. For $n = k + 1$, we represent argument A as $B_1, \ldots, B_n \Rightarrow \psi$, where $m \leq n$. Consider the following possibilities:

(1) If A_i is a $multiple\ unit\ I$ argument and r_{k+1} is an $unit\ I$ inference, then according to Definition 7 and Definition 9(3), A is a $multiple\ unit\ I$ argument.
(2) If A_i is a $multiple\ unit\ I$ argument and r_{k+1} is an $unit\ II$ inference, then according to Definition 7 and Definition 9(5), A is a $mixed$ argument.
(3) If A_i is a $multiple\ unit\ II$ argument and r_{k+1} is an $unit\ I$ inference, then according to Definition 7 and Definition 9(5), A is a $mixed$ argument.
(4) If A_i is a $multiple\ unit\ II$ argument and r_{k+1} is an $unit\ II$ inference, then according to Definition 7 and Definition 9(4), A is a $multiple\ unit\ II$ argument.

(5) If A_i is a $mixed$ argument and r_{k+1} is an $unit\ I$ or $unit\ II$ inference, then according to Definition 7 and Definition 9(5), A is a $mixed$ argument.

Secondly, we prove the property of uniqueness of argument type. Assume there exists an argument A corresponding to two or more argument types: then there must exist two or more top rules in the argument, and then there are two or more conclusions in A, which contradicts the definition of argument. □

Consider again Example 3. We have that A_1, A_2, A_3, A_4, A_5 are atomic arguments, A_8 is a $unit\ I$ argument, A_6 is a $unit\ II$ argument, A_9 is a $multiple\ unit\ I$, A_7 is a $multiple\ unit\ II$, and A_{10} is a $mixed$ argument.

We next define several argument structures, which are sets of arguments with certain properties.

Definition 10. *A set of arguments* $\{A_1, \ldots, A_n\}$ *is* interconnected *iff for any argument* A_i *there exists an argument* A_j *such that* $Conc(A_i) \in Prem(A_j)$ *or* $Con(A_j) \in Prem(A_i)$ *or* $Con(A_i) = Con(A_j)$ *or* $Prem(A_i) = Prem(A_j)$.

Definition 11. *The set of argument structures*[3] *is defined as follows:*

(1) A set of arguments $\{A_1, \ldots, A_n\}$ *is in a* $serial\ convergent\ structure$ *SCS iff there are only* $unit\ I$ *arguments in the set of arguments* $\{A_1, \ldots, A_n\}$ *and for any* A_i *and* A_j *we have* $Conc(A_i) = Conc(A_j)$.
(2) A set of arguments $\{A_1, \ldots, A_n\}$ *is in a* $serial\ divergent\ structure$ *SDS iff there are only* $unit\ I$ *arguments in the set of arguments* $\{A_1, \ldots, A_n\}$ *and for any* A_i *and* A_j *we have* $Prem(A_i) = Prem(A_j)$.
(3) A set of arguments $\{A_1, \ldots, A_n\}$ *is in a* $linked\ convergent\ structure$ *LCS iff it contains only* $unit\ II$ *arguments and for any* A_i *and* A_j *we have* $Conc(A_i) = Conc(A_j)$.
(4) A set of arguments $\{A_1, \ldots, A_n\}$ *is in a* $linked\ divergent\ structure$ *LDS iff it contains only* $unit\ II$ *arguments and for any* A_i *and* A_j *we have* $Prem(A_i) = Prem(A_j)$.
(5) A set of arguments $\{A_1, \ldots, A_n\}$ *is in a* $mixed\ structure$ *MS iff it is interconnected and it is not of the form of either SCS, SDS, LCS and LDS.*

We display the diagrams of argument structures in Figure 4. For simplicity, we assume $n = 2$ in the diagrams and show only one case of $mixed\ structure$.

Corollary 1. *A set of arguments* $S = \{A_1, \ldots, A_n\}$ *is interconnected if for any* $A_i \in S$, *there is an argument A, such that* A_i *is a maximal proper subargument of A.*

4.1 Reconsidering the standard approach

First, we consider the correspondence between the standard approach and our new approach. It is easy to see that single, linked and serial arguments, respectively, correspond to $unit\ I$, $unit\ II$ and $multiple\ unit\ I$ arguments.

However, *convergent* and *divergent* arguments are not arguments any more, since a *convergent* 'argument' now is an argument structure consisting of a number of distinct $unit\ I$ arguments for the same conclusion, while a $divergent$ 'argument' now is an argument structure consisting of a number of distinct $unit\ I$ argument with the

[3] The structure here is different from the structure in informal approaches, where it refers to the structure of an *individual* argument.

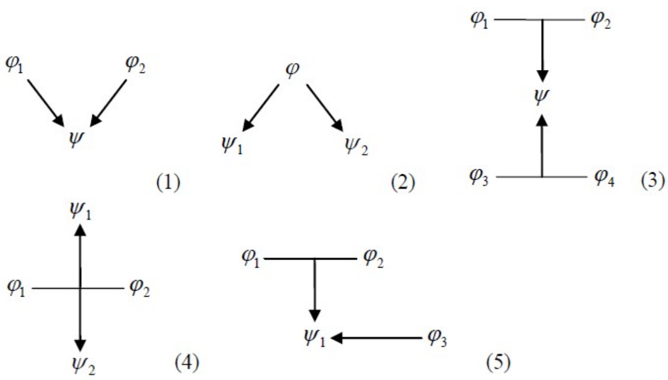


Figure 4. Argument structures

same premise. For instance, in Example 1(1) there are two arguments $A \Rightarrow (C)$ and $B \Rightarrow (C)$ for the same conclusion (C), and in Example 1(2), there are two arguments $A \Rightarrow (B)$ and $A \Rightarrow (C)$ with the same premise (A) where but different conclusions.

Therefore, the classes of convergent, divergent 'arguments' are not arguments but argument *structures*. Actually, they correspond to the $serial\ convergent\ structure$ SCS and the $serial\ divergent\ structure$ SDS. Moreover, the class of complex arguments in the standard approach is not an argument if it contains SCS or SDS, but instead corresponds to the mixed argument structure MS. Otherwise, it corresponds to a mixed argument.

From the above analysis we see that the standard approach is incomplete and, moreover, does not distinguish types of individual argument from types of argument structures. We can conclude that the new classification in terms of the *ASPIC*$^+$ framework is helpful in clarifying and complementing the standard approach.

5 The problem of hybrid arguments

In this section we analyze why Vorobej's class of hybrid arguments is not needed if our approach is adopted. In our new approach, Vorobej's hybrid 'arguments' are not arguments but argument structures consisting a number of arguments. More specifically, they are of type $mixed\ structure$ MS or $linked\ convergent\ structure$ LCS.

We first make a notion explicit and redefine a definition. In [4] the notion of relevance is implicit and Vorobej treated it as a primitive dyadic relation. We note that there are two kinds of relevance: *defeasible relevance* indicates the support from a set of arguments to the conclusion via a defeasible inference, while *strict relevance* indicates the support form a set of arguments to the conclusion via a strict inference.

In the *ASPIC*$^+$ framework, we write $S \vdash \varphi$ if there exists a strict argument for φ with all premises taken from S, and $S \mid\!\sim \varphi$ if there exists a defeasible argument for φ with all premises taken from S. Then Definition 4 can be rewritten as follows:

Definition 12. *A set of premises* Σ *supplements a set of premises* Δ *iff (1)* $\Sigma \not\mid\!\sim C$ *and* $\Sigma \neq \emptyset$*; (2)* $\Delta \mid\!\sim C$*; (3)* $\Sigma \cup \Delta \vdash C$ *or* $\Sigma \cup \Delta \mid\!\sim C$*, and (4)* $\Sigma \cup \Delta$ *is the minimal set satisfying clauses (1),(2) and (3) when* $\Sigma \cup \Delta \vdash C$.

If a set of premises $\Sigma = \{P_1, \ldots, P_m\}$ supplements a set of premises $\Delta = \{Q_1, \ldots, Q_n\}$, then we have two arguments A

and B, where argument A is of the form $Q_1, \ldots, Q_n \Rightarrow C$ and argument B is of the form $P_1, \ldots, P_m, Q_1, \ldots, Q_n \Rightarrow C$ or $P_1, \ldots, P_m, Q_1, \ldots, Q_n \rightarrow C$.

Thus, the hybrid argument here is a (1) $mixed\ structure$ MS consisting of a $unit\ I$ argument and a $unit\ II$ argument, if $m = 1$, or (2) a $linked\ convergent\ structure$ LCS consisting of two $linked$ arguments, if $m > 1$.

We now first reconsider Example 2.

- (F): (1) All the ducks that I' ve seen on the pond are yellow. (2) I've seen all the ducks on the pond. (3) All the ducks on the pond are yellow.

Arguably, (1) supports (3) because of the defeasible inference rule of enumerative induction:

- All observed F's are G's $\Rightarrow$ all F's are G's.

Moreover, (1) and (2) together arguably support (3) because of a deductive version of enumerative induction:

- All observed F's are G's, all observed F's are all F's $\rightarrow$ all F's are G's.

We then see that the apparently hybrid argument is in fact a convergent structure consisting of two separate arguments for the same conclusion, sharing one premise:

$A = 1 \Rightarrow$ All the ducks on the pond are yellow.
$B = 1, 2 \rightarrow$ All the ducks on the pond are yellow.

Actually, all examples in [4] can be reconstructed in terms of these two kinds of structures:

Example 4. *Consider examples (G) and (J) as follows:*

- *(G): (1) My duck is yellow. (2) Almost without exception, yellow ducks are migratory. (3) My duck is no exception to any rule. (4) My duck migrates.*
- *(J): (1) Data quacks. (2) Data has webbed feet. (3) 95% of those creatures who both quack and have webbed feet are ducks. (4) Data is a duck.*

In example (G), we have that $\{(1), (2)\} \mid\!\sim (4)$ and $\{(1), (2), (3)\} \vdash (4)$, so we have two arguments A and B for the same conclusion:

- $A = 1, 2 \Rightarrow (4)$ with a defeasible inference rule: almost without exception X's are Y's, a is an $X \Rightarrow a$ is a Y;
- $B = 1, 2, 3 \rightarrow (4)$ with a strict inference rule: almost without exception X's are Y's, a is a X, a is no exception to any rule $\rightarrow a$ is a Y.

In example (J), there are four arguments A, B, C and D based on $\{(1)\} \mid\!\sim (4)$, $\{(2)\} \mid\!\sim (4)$, $\{(1), (2)\} \mid\!\sim (4)$ and $\{(1), (2), (3)\} \mid\!\sim (4)$:

- $A = 1 \Rightarrow (4)$ with a defeasible inference rule: x quacks $\Rightarrow x$ is a duck;
- $B = 2 \Rightarrow (4)$ with a defeasible inference rule: x has webbed feet $\Rightarrow x$ is a duck;
- $C = 1, 2 \Rightarrow (4)$ with a defeasible inference rule that aggregates the two previous inference rules;
- $D = 1, 2, 3 \Rightarrow (4)$ with a defeasible inference rule: a is a Y, a is a Z, 95% of x's who are both Y and Z are $T \Rightarrow a$ is a T.

On our account argument (G) is a $linked\ convergent\ structure$ and argument (J) is a $mixed\ structure$.

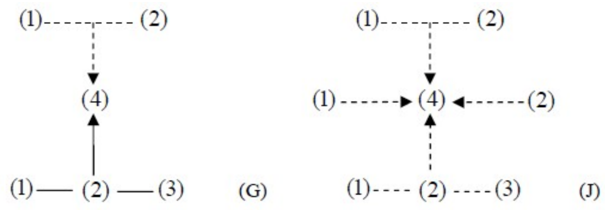


Figure 5. Hybrid Arguments

6 Conclusion

In this paper we showed how AI models of argumentation can be used to clarify and extend informal-logic approaches to the structure of arguments. We defined a complete classification of types of arguments, we showed that convergent and divergent 'arguments' are not arguments but sets of arguments and we showed that Vorobeij's 'hybrid arguments' can be defined by explicitly distinguishing between deductive and defeasible inferences.

ACKNOWLEDGEMENTS

Bin Wei was supported by the Chinese MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (12JJD720006).

REFERENCES

[1] S. N. Thomas, *Practical Reasoning in Natural Language*, Third edition, Prentice-hall in Englewood Cliffs.N.J, 1986.
[2] D. N. Walton, *Argument Structure: A pragmatic Theory*. Toronto, University of Toronto Press, 1996.
[3] James B. Freeman, *Argument Structure: Representation and Theory*. Springer, 148, 2011.
[4] Mark Vorobej, Hybrid Arguments, *Informal Logic*, Vol.17, No.2, 289-296, 1995.
[5] D. N. Walton, *Fundamentals of Critical Argumentation*, University of Cambridge Press, 139-149, 2006.
[6] H. Prakken, An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2): 93-124, 2010.
[7] S.J. Modgil and H. Prakken, Revisiting preferences and argumentation, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, 1021-1026.
[8] B. van Gijzel and H. Prakken, Relating Carneades with abstract argumentation, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, 1113-1119.
[9] B. van Gijzel and H. Prakken, Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation, *Argument and Computation*, 3(1): 21-47, 2012.
[10] P.M. Dung and P. Mancarella and F. Toni, Computing ideal sceptical argumentation, *Artificial Intelligence*, 171: 642-674, 2007.
[11] N. Gorogiannis and A. Hunter, Instantiating abstract argumentation with classical-logic arguments: postulates and properties, *Artificial Intelligence*, 175: 1479-1497, 2011.
[12] T.F. Gordon and H. Prakken and D.N. Walton, The Carneades model of argument and burden of proof, *Artificial Intelligence*, 171: 875-896, 2007.

Author Index