Coherence and Flexibility in Dialogue Games for Argumentation

Henry Prakken
Department of Information and Computing Sciences, Utrecht University
Faculty of Law, University of Groningen
The Netherlands
henry@cs.uu.nl

October 18, 2005

Abstract

This article carries out a formal study of dialogue games for argumentation. A formal framework for such games is proposed which imposes an explicit reply structure on dialogues, where each dialogue move either attacks or surrenders to some earlier move of the other participant. The framework is flexible in several respects. It allows for different underlying logics, alternative sets of locutions and more or less strict rules for when they are allowed. In particular, it allows for varying degrees of coherence and flexibility when it comes to maintaining the focus of a dialogue. Its formal nature supports the study of formal properties of specific dialogue protocols, especially on how they respect the underlying logic.
Keywords: argumentation, persuasion, dialouge games, defeasible reasoning.

1 Introduction

This article[1] studies the formal modelling of dialogue games for argumentation. Such games regulate dialogues where two parties argue about the tenability of one or more claims or arguments, each trying to persuade the other participant to adopt their point of view. Hence such dialogues are often called *persuasion dialogues* ([34]). Systems for argumentation dialogues were already studied in medieval times ([3]). The modern study of formal dialogue systems for argumentation probably started with two publications by Charles Hamblin [15, 16]. Initially, the topic was studied only within philosophical logic and argumentation theory (e.g. [21, 22, 34]). From the early nineties the study of argumentation dialogues branched out into two directions: artificial intelligence & law (e.g. [14, 12, 5, 18, 29]) and multi-agent systems (e.g. [2, 23, 25, 24]). There is also some work in general AI, notably [20] and [8].

Dialogue systems define the principles of coherent dialogue. [9] defines coherence in terms of the goal of a dialogue. According to him, whereas logic defines the conditions under which a proposition is true, dialogue systems define the conditions under which an utterance is appropriate, and this is

[1]This article is an extended and substantially revised version of [28]. I thank Martin Caminada and Joris Hulstijn for their comments on an earlier draft of this article, and Gerhard Brewka for suggesting some simplifications in the definitions of Section 2.

the case if the utterance furthers the goal of the dialogue in which it is made. Thus according to Carlson the principles governing the meaning and use of utterances should not be defined at the level of individual speech acts but at the level of the dialogue in which the utterance is made. This justifies why most work on argumentation dialogues, like Carlson, takes a game-theoretic approach to dialogues, where speech acts are viewed as moves in a game and rules for their appropriateness are formulated as rules of the game.

The formalisation of argumentation dialogues for persuasion draws some inspiration from dialogue logic ([19]) but it differs from it in one crucial respect. Dialogue logic aims to define the semantics of logical connectives in terms of rules of attack and defence, and accordingly the goal of a dialogue is to determine whether a proposition is implied by a given set of propositions. Persuasion dialogues, by contrast, are substantive, where the participants ask for and provide substantive reasons for their claims. In consequence, the information available to and agreed by the participants changes during a dialogue and the goal of a persuasion dialogue is to resolve a conflict of opinion. (In addition the players can have their own private goals, such as to win the dialogue.)

A typical persuasion dialogue is the following.

`Paul:` My car is very safe. (*making a claim*)
`Olga:` Why is your car safe? (*asking grounds for a claim*)
`Paul:` Since it has an airbag, (*offering grounds for a claim*)
`Olga:` That is true, (*conceding a claim*) but I disagree that this makes your car safe: the newspapers recently reported on airbags expanding without cause. (*stating a counterargument*)
`Paul:` Yes, that is what the newspapers say (*conceding a claim*) but that does not prove anything, since newspaper reports are very unreliable sources of technological information. (*undercutting a counterargument*)
`Olga:` Still your car is not safe, since its maximum speed is very high. (*alternative counterargument*)

This dialogue (see Figure 1) illustrates several features of argumentation dialogues relevant to the present study. Firstly, it illustrates that players may return to earlier choices and move alternative replies: in her last move Olga states an alternative counterargument after she sees that Paul had a strong counterattack on her first counterargument. Note that she could also have moved the alternative counterargument immediately after the first, to leave Paul with two attacks to counter. The dialogue also illustrates that players may postpone their replies, sometimes even indefinitely: by providing her second argument why Paul's car is not safe, Olga postpones her reply to Paul's counterattack on her first argument for this claim; if Paul fails to successfully attack her second argument, such a reply might become superfluous.

The motivation for the present study is threefold.

Firstly, although the formal theory of dialogue games with argumentation is developing, the present state-of-the art is that there exist a considerable number of systems which are carefully defined but of which the underlying design principles are largely implicit. This makes it hard to compare the various systems and investigate their formal properties. A first aim of this article is to present a formal framework for a class of argumentation dialogues, viz. of dialogues with a clear 'reply structure', where each dialogue move either attacks or surrenders to a preceding move of the other participant. For instance, the above dialogue has the following reply structure, where each *claim*, *why* and *since* move is an attacking and each *concede* move is a surrendering reply. The framework should allow, among other things, for different underlying logics and different sets of speech acts.

A second motivation is to allow for varying degrees of coherence and flexibility when it comes

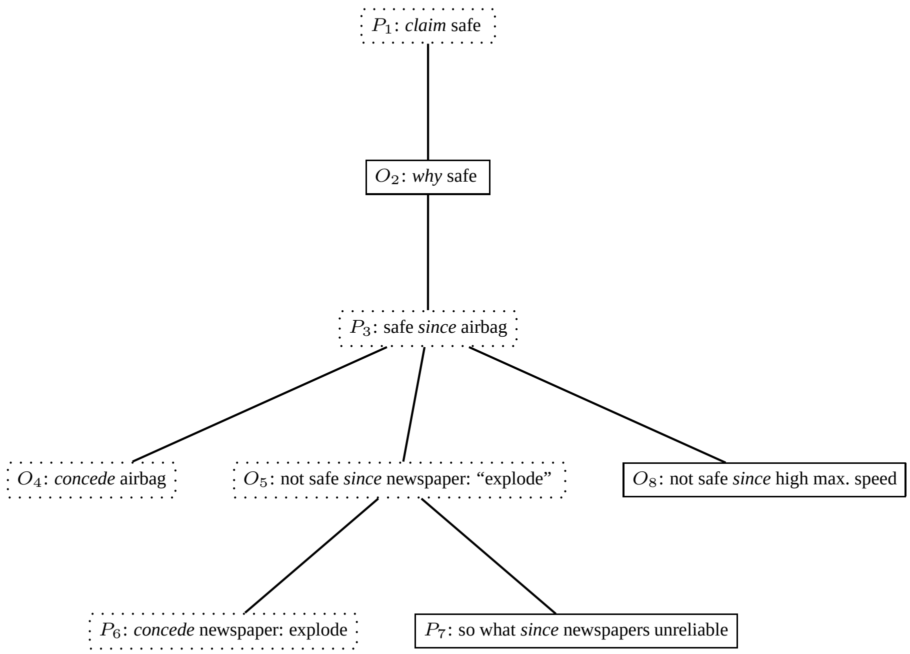

Figure 1:

to maintaining the focus of a dialogue. Dialogue systems can vary in their structural properties in several ways (cf. [20]): whether players can reply just once to the other player's moves or may try alternative replies (*unique- vs. multi-reply protocols*); whether players can make just one or may make several moves before the turn shifts (*unique- vs. multi-move protocols*); and whether the turn shifts as soon as the player-to-move has made himself the winning side or may shift later (*immediate- vs. non-immediate-reply protocols*). Many current systems impose a rather rigid 'control structure' on dialogues. In the most rigid, unique-move and unique- and immediate-reply protocols, the turn switches after each move and each player must respond to the previous move of the other player; see, for instance, [21] and [25] (although in [25] this is relaxed to allow replies to each premise of an argument in turn). This does not allow for alternative replies in one turn (for instance, alternative arguments in support or attack of a proposition), for coming back to choices made at earlier turns, or for postponing replies (perhaps even indefinitely when a reply becomes irrelevant). Other systems, such as [34] allow for multiple moves and alternative replies in one turn, but still do not allow for returning to earlier choices or postponement. An important assumption of the present study is that the degree of structural 'strictness' of a dialogue system depends on the context of a dialogue (likewise [20]). In contexts with little time and resources a unique-move, unique- and immediate reply protocol may be best, to force the participants not to waste resources, while in other contexts with more time and resources it is better to allow the participants more freedom to explore alternatives and return to earlier choices. Therefore, a second aim of this study is to investigate how, within the proposed framework, several more or less structurally flexible protocols can be formulated. The main idea is to exploit the explicit attack-or-surrender-structure of dialogues in the proposed framework for defining the 'dialogical status' of each move as either 'in' (a solid box in the above figure) or 'out' (a dotted box): this will allow for various turntaking and termination rules and for defining various degrees of relevance of moves.

Finally, the relation of dialogue systems of argumentation with nonmonotonic logic needs more investigation. Argumentation often involves defeasible reasoning, where participants provide plausible but fallible grounds for their claims, so that their arguments can be attacked with counterarguments. Clearly, this is related to nonmonotonic inference, but the relation between dialogue protocols and proof theories for nonmonotonic logic has so far received little attention. An important issue here is how the inherently dynamic nature of dialogues (where the available information changes during a dialogue) can be reconciled with the inherently static nature of nonmonotonic logics (in that they determine the consequences of a given set of propositions). In this respect, the main idea of the present study is that in the course of a dialogue the participants implicitly build a logical structure of arguments and counterarguments relevant to the dialogue topic. For instance, in the above example dialogue the following structure is created (see Figure 2, in which the boxes contain arguments and the links stand for attack relations between arguments): This idea allows a study of

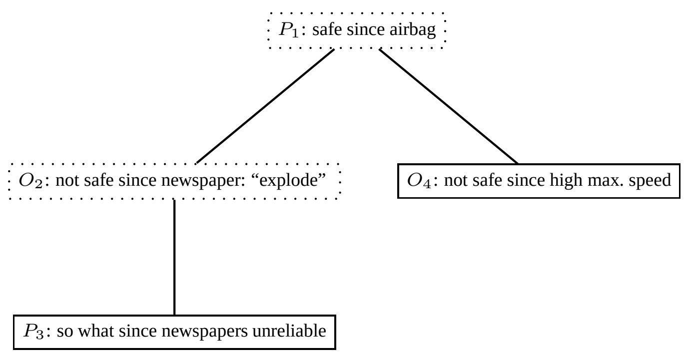

Figure 2:

correspondences between the 'dialogical status' of the initial move of a dialogue and the logical structure created during the dialogue.

The rest of this article is organised as follows. In Section 2 the formal framework is presented, which in Section 3 is instantiated with a 'liberal' system for dialogues with argumentation, allowing the participants much structural freedom at the cost of limited coherence. Section 4 presents an 'any time' winning criterion, defined in terms of the dialogical status of moves. Section 5 then introduces the idea that during a dialogue the participants implicitly build a logical structure of arguments and counterarguments concerning the dialogue's topic. Then formal correspondences are investigated between the 'any time' winning definition and properties of this logical structure. Section 6 then explains why liberal protocols enforce only a limited degree of coherence of dialogues, and studies two more rigid protocols in which each move is (strongly or weakly) relevant to the dialogue topic in a structural sense. Section 7 then studies some further possible protocol rules, after which I and with a discussion of related research and some concluding remarks.

2 A framework for dialogue games for argumentation

In this section a general framework for dialogue systems for argumentation is formulated. First the essentials of logical systems for defeasible argumentation are sketched, which provide the logical basis for dialogue systems for argumentation. Then the framework is briefly outlined and its fixed and variable elements are indicated, after which it is formally defined.

2.1 Argument-based logics for nonmonotonic reasoning

The discussion in this article will, whenever possible, abstract from the logical structure of the parties' individual reasoning. Nevertheless, some choices have to be made. Since argumentation typically involves defeasible reasoning, we need a nonmonotonic logic. Since we are dealing with dialogues for argumentation, one particular form of nonmonotonic logic is very appropriate, viz. logics for defeasible argumentation, or argumentation systems for short (for an overview see [31]). The restriction to argumentation systems is less substantial than it would seem at first sight, since many other nonmonotonic logics can be reformulated in argument-based style [10, 6].

Argumentation systems formalise nonmonotonic reasoning as the construction and comparison of arguments for and against certain conclusions. Nonmonotonicity arises from the fact that new information may give rise to new counterarguments that defeat the original argument. In general, three ways of defeat are distinguished: arguing for a contradictory conclusion (rebutting), arguing that an inference is incorrect (undercutting), or denying a premise (premise-attack); in all three cases also considerations of strength of preference can be involved. In fact, most existing systems allow for only one or two of the kinds of defeat.

Inference in argumentation systems is defined relative to a set of arguments and a binary defeat relation between them. In the literature, several definitions have been proposed. Typically, they classify arguments in three classes: the 'winning' or *justified* arguments, the 'losing' or *overruled* arguments, and the 'ties', i.e., the *defensible* arguments, which are involved in an irresolvable conflict. Corresponding notions of propositional inference can be defined in terms of the status of arguments of which they are conclusions. For modelling persuasion dialogues the question of interest is whether a justified argument for the dialogue topic can be constructed.

One way to define argumentation logics is in the dialectical form of argument games. Such games model defeasible reasoning as a dispute between a proponent and opponent of a claim, who exchange arguments and counterarguments according to certain rules in order to win according to a certain winning condition. A proposition is provably justified on the basis of a set of arguments if its proponent has a winning strategy for a supporting argument. In this article an argument game for Dung's grounded semantics [10] will be used, which at present is the only available argument game for the problem whether an argument is justified. In this game, proponent starts and then both proponent and opponent must defeat the previous argument of the other player. A player wins if the other player cannot move (alternative rules are possible but these will do for present purposes). A strategy for proponent can be represented as a tree with the argument in dispute as its root and branching only after proponent moves, with a child for each argument defeating this move. A strategy is winning if all its branches end with proponent nodes.

2.2 The framework: general ideas

The framework to be developed should allow for variations on a number of issues. It should allow for different underlying argument-based logics (but all with grounded semantics), for various sets of locutions, for different turntaking rules and different rules on whether multiple replies, postponing of replies and coming back to earlier choices is allowed. On the other hand, the framework should impose some basic common structure on all dialogues. The basic structure proposed in the present study is an explicit reply structure on moves, where each move either attacks or surrenders to one earlier (but not necessarily the last) move of the other player. Such a structure is implicit in the protocols of many existing systems but usually not made explicit. It seems especially suited for "verbal struggles" (a term coined by [4] in their classification of speech act verbs). I do not claim, however, that all dialogues should or do conform to this structure. It may, for instance, be less suited for dialogues where the focus is more on investigation or deliberation than on settling a conflict of opinion. Another assumption of the framework is that during a dialogue the players implicitly build a structure of arguments and counterarguments related to the dialogue topic.

Thus according to the present approach a dialogue can be regarded in three ways. One can look at the order in which the moves are made, in which case a dialogue is regarded as a linear structure. One can also look at the reply relations between the moves, in which case the dialogue is conceived of as a tree. Finally, one can look at the arguments that are exchanged in reply to each other, in which case the dialogue is regarded as a dialectical structure of arguments and counterarguments. These three ways to look at argumentation dialogues were illustrated in the introduction by the three different presentations of the dialogue between Paul and Olga.

2.3 The framework formally defined

Now the framework will be formally defined. All dialogues are assumed to be for two parties arguing about a single dialogue topic $t \in L_t$, the *proponent* (P) who defends t and the *opponent* (O) who challenges t. As for notation, for any player p, we define $\overline{p} = O$ iff $p = P$ and $\overline{p} = P$ iff $p = O$.

The top level definition of the framework is as follows.

Definition 1 A *dialogue system for argumentation* (dialogue system for short) is a pair $(\mathcal{L}, \mathcal{D})$, where $\mathcal{L}$ is a logic for defeasible argumentation and $\mathcal{D}$ is a dialogue system proper.

The elements of the top level definition are in turn defined as follows. Logics for defeasible argumentation are defined as an instance of Dung's [10] abstract framework with a specific, tree-based form of arguments (cf. e.g. [26, 27, 33]) and conforming to grounded semantics.[2]

Definition 2 A *logic for defeasible argumentation* $\mathcal{L}$ is a tuple $(L_t, R, Args, \rightarrow)$, where L_t (the *topic language*) is a logical language, R is a set of *inference rules* over L_t, *Args* (the *arguments*) is a set of AND-trees of which the nodes are in L_t and the AND-links are inferences instantiating rules in R, and $\rightarrow$ a binary relation of *defeat* defined on $Args$. For any argument A, $prem(A)$ is the set of leaves of A (its premises) and $conc(A)$ is the root of A (its conclusion).

An *argumentation theory* T_F within $\mathcal{L}$ (where $F \subseteq L_t$) is a pair $(A, \rightarrow_{/A})$ where A consists of all arguments in *Args* with only nodes from F and $\rightarrow_{/A}$ is $\rightarrow$ restricted to $A \times A$. T_F is called *finitary* if none of its arguments has an infinite number of defeaters.

For any set $A \subseteq Args$ the *information base* $I(A)$ is the set of all formulas that are a premise of an argument in A. The *closure* $Cl(A)$ of a set of arguments $A \in Args$ is the argumentation theory $T_{I(A)}$.

An argument B *extends* an argument A if *conc*$(B) = \varphi$ and $\varphi \in prem(A)$ (for example, r *since* s extends p *since* q, r). The concatenation of A and B (where B extends A) is denoted by $B \otimes A$. Defeasible inference in $\mathcal{L}$ is assumed to be defined according to grounded semantics. The defeat relation of $\mathcal{L}$ is assumed to satisfy the following property: if A defeats B, then for all C extending A and D extending B it holds that $C \otimes A$ defeats $D \otimes B$.

[2]For present purposes a very detailed formal definition is not needed; for the full details the reader is referred to the references.

The idea of an argumentation theory is that it contains all arguments that are constructible on the basis of a certain theory or knowledge base. Note that each link of an argument corresponds to a (deductive or defeasible) inference rule in R. The present framework fully abstracts from the nature of these rules; for a detailed account see the above references. Note also that the assumption on the defeat relation is not completely innocent: it is not satisfied in systems where arguments are compared on their 'weakest links', as in, for instance, Pollock's work (e.g. [26, 27]).

Definition 3 A *dialogue system proper* is a triple $\mathcal{D} = (L_c, P, C)$ where L_c (the communication language) is a set of locutions, Pr is a *protocol* for L_c, and C is a set of effect rules of locutions in L_c, specifying the effects of the locutions on the participants' *commitments*.

A communication language is a set of locutions and two relations of attacking and surrendering reply are defined on this set.

Definition 4 A *communication language* is a set L_c of *locutions*. Each $s \in L_c$ is of the form $p(c)$ where p is an element of a given set P of performatives and c either is a member or subset of L_t, or is a member of $Args$ (of some given logic $\mathcal{L}$). On Lc two binary relations R_a and R_s of *attacking* and *surrendering reply* are defined. Both relations are irreflexive and in addition satisfy the following conditions:

1. $\forall a, b, c : (a, b) \in R_a \Rightarrow (a, c) \notin R_s$
2. $\forall a, b, c : (a, b) \in R_s \Rightarrow (c, a) \notin R_a$

The function $att : R_s \longrightarrow \mathcal{P}(R_a)$ assigns to each pair $(a, b) \in R_s$ one or more *attacking counterparts* $(c, b) \in R_a$.

Condition (1) says that a locution cannot be an attack and a surrender at the same time, and condition (2) says that surrenders cannot be attacked (this is since they effectively end a line of dispute).

The protocol for L_c is defined in terms of the notion of a dialogue, which in turn is defined with the notion of a move:

Definition 5 (Moves and dialogues)

- The set M of *moves* is defined as $\mathbb{N} \times \{P, O\} \times L_c^p \times \mathbb{N}$, where the four elements of a move m are denoted by, respectively:
  - $id(m)$, the *identifier* of the move,
  - $pl(m)$, the *player* of the move,
  - $s(m)$, the *speech act* performed in the move,
  - $t(m)$, the *target* of the move.
- The set of *dialogues*, denoted by $M^{\leq\infty}$, is the set of all sequences $m_1, \ldots, m_i, \ldots$ from M such that
  - each i^{th} element in the sequence has identifier i,
  - $t(m_1) = 0$;
  - for all $i > 1$ it holds that $t(m_i) = j$ for some m_j preceding m_i in the sequence.

The set of *finite dialogues*, denoted by $M^{<\infty}$, is the set of all finite sequences that satisfy these conditions. For any dialogue $d = m_1, \ldots, m_n, \ldots$, the sequence $m_1, \ldots, m_i$ is denoted by d_i, where d_0 denotes the empty dialogue. When d is a dialogue and m a move then d, m denotes the continuation of d with m.

Note that the definition of dialogues implies that several speakers cannot speak at the same time.

When $t(m) = id(m')$ we say that m *replies to* m' in d and also that m' is the *target of* m in d. We will sometimes slightly abuse notation and let $t(m)$ denote a move instead of just its identifier. When $s(m)$ is an attacking (surrendering) reply to $s(m')$ we will also say that m is an attacking (surrendering) reply to m'.

A protocol also assumes a turntaking rule.

Definition 6 A *turntaking function* T is a function

- $T : M^{<\infty} \longrightarrow \mathcal{P}(\{P, O\})$

such that $T(\emptyset) = \{P\}$. A *turn* of a dialogue is a maximal sequence of stages in the dialogue where the same player moves.

When $T(d)$ is a singleton, the brackets will be omitted. Note that this definition allows that more than one speaker has the right to speak next.

We are now in the position to define the central element of a dialogue game, the 'rules of the game', in other words, the protocol.

Definition 7 (Protocols) A *protocol* on M is a set $P \subseteq M^{<\infty}$ satisfying the condition that whenever d is in P, so are all initial sequences that d starts with.

A partial function $Pr : M^{<\infty} \longrightarrow \mathcal{P}(M)$ is derived from P as follows:

- $Pr(d) =$ undefined whenever $d \notin P$;
- $Pr(d) = \{m \mid d, m \in P\}$ otherwise.

The elements of *dom*(Pr) (the domain of Pr) are called the *legal finite dialogues*. The elements of $Pr(d)$ are called the moves allowed after d. If d is a legal dialogue and $Pr(d) = \emptyset$, then d is said to be a *terminated* dialogue.

All protocols are further assumed to satisfy the following basic conditions for all moves m_i and all legal finite dialogues d.

If $m \in Pr(d)$, then:

- R_1: $pl(m) \in T(d)$;
- R_2: If $d \neq d_0$ and $m \neq m_1$, then $s(m)$ is a reply to $s(t(m))$ according to L_c;
- R_3: If m replies to m', then $pl(m) \neq pl(m')$;
- R_4: If there is an m' in d such that $t(m) = t(m')$ then $s(m) \neq s(m')$.
- R_5: For any $m' \in d$ that surrenders to $t(m)$, m is not an attacking counterpart of m'.

Together these conditions capture a lower bound on coherence of dialogues. Note that they state only necessary conditions for move legality. Rule R_1 says that a move is legal only if moved by the player-to-move. R_2 says that a replying move must be a reply to its target according to L_c, and R_3 says that one cannot reply to one's own moves. Rule R_4 states that if a player backtracks, the new move must be different from the first one. ('backtracking' in this article is taken to mean any alternative reply to the same target in a later turn). Finally, R_5 says that surrenders may not be 'revoked'. At first sight, it would seem that R_5 could be formulated as "$t(m)$ does not have a surrendering reply in d". However, later we will see that it makes sense to attack one premise of an argument even if another of its premises has been surrendered.

Finally, a commitment function is a function that assigns to each player at each stage of a dialogue a set of propositions to which the player is committed at that stage.

Definition 8 A *commitment function* is a function

- C: $M^{\leq \infty} \times \{P, O\} \longrightarrow \mathcal{P}(L_t)$.

such that $C_\emptyset(p) = \emptyset$. $C_d(p)$ denotes player p's commitments in dialogue d.

Informally, commitments are the players' public standpoints, which they are expected to defend upon challenge. As is well known, commitments should not be confused with beliefs. Beliefs are internal to the player and need not be known by the other player, while commitments are incurred publicly. It is perfectly possible that a participant believes φ but is committed to $\neg\varphi$ (think of a suspect in a trial who knows he is guilty but pleas his innocence.) Commitments are typically incurred by making or conceding claims and stating, and they are typically given up by withdrawing claims or premises. Commitments can be used to define a dialogue's termination (for instance, when opponent becomes committed to proponent's initial claim) and outcome (in terms of the joint commitments of the players at termination). Commitments can also be used to regulate legality of moves, for instance, by requiring them not to contradict or challenge their own commitments.

3 Liberal dialogue systems

So far any 'verbal struggle' could fit the framework. It will now be specialised for argumentation with a particular communication language and some basic protocol rules motivated by this language. In fact, a class of liberal[3] dialogue systems will be defined (parametrised by a logic $\mathcal{L}$), in which the participants have much freedom, and which is intended to be the core of all other dialogue systems of this study.

The communication language allows for making a claim, for challenging, conceding and retracting a claim, for supporting a claim with an argument, and for attacking arguments with counterarguments or by challenging their premises.

Commitment rules are defined, but the commitments are only used in defining termination and outcome of dialogues; they do not constrain move legality (see for that Section 7.1 below). There are only weak relevance requirements of moves, viz. those given by the reply structure of L_c, and there are no restrictions at all on length of turns. Basically, a speaker may continue speaking as long as he is not interrupted by the listener, and he may make any move as long as according to L_c it is a well-formed reply to some earlier move of the listener. So liberal dialogues greatly rely for their coherence on the cooperativeness of the dialogue participants.

[3]In [30] the term 'liberal disputes' was used for what in this article will be called 'relevant dialogues'; see Section 6 below.

Table 1: Speech acts for liberal dialogues

Acts	Attacks	Surrenders
claim φ	*why* φ	*concede* φ
why φ	*argue* A (*conc*$(A) = \varphi$)	*retract* φ
argue A	*why* φ ($\varphi \in prem(A)$) *argue* B (B defeats A)	*concede* φ ($\varphi \in prem(A)$ or $\varphi = conc(A)$)
concede φ		
retract φ		

3.1 The communication language

The communication language is listed in Table 1. In examples below, when an argument contains a single inference, it will usually be listed as *conclusion since premises*. Note that counterarguments must defeat their target according to $\mathcal{L}$. Attacking counterparts of a surrender are at the same line of the surrender except for the second line of the *argue* A row: *argue* B is an attacking counterpart of *concede* φ only if the conclusion of B negates or is negated by φ. (So the attacking counterpart of conceding a premise is an premise-attack and the attacking counterpart of conceding a conclusion is a rebuttal.)

3.2 The commitment rules

The following commitment rules seem to be uncontroversial and can be found throughout the literature. (Below s denotes the speaker of the move; effects on the other parties' commitments are only specified when a change is effected.)

- If $s(m) = claim(\varphi)$ then $C_s(d, m) = C_s(d) \cup \{\varphi\}$
- If $s(m) = why(\varphi)$ then $C_s(d, m) = C_s(d)$
- If $s(m) = concede(\varphi)$ then $C_s(d, m) = C_s(d) \cup \{\varphi\}$
- If $s(m) = retract(\varphi)$ then $C_s(d, m) = C_s(d) \ - \{\varphi\}$
- If $s(m) = argue(A)$ then $C_s(d, m) = C_s(d) \cup prem(A) \cup \{conc(A)\}$

3.3 Turntaking

As for the turntaking function, proponent starts with a unique move (which introduces the topic of the dialogue), opponent then replies and after that it is simply assumed that it is always the speaker's turn; in other words, the turn shifts as soon a new speaker succeeds in saying something.

- T_L: $T(d_0) = P, T(d_1) = O$, else $T(d) = \{P, O\}$.

Thus protocol rule R_1 is always satisfied for any dialogue with at least two moves.

3.4 The protocol

The protocol for liberal dialogues adds two protocol rules to those of the general framework.

If $m \in Pr(d)$, then:

- R_6: If $d = \emptyset$, then $s(m)$ is of the form $claim(\varphi)$ or *argue* A.
- R_7: If m concedes the conclusion of an argument moved in m', then m' does not reply to a *why* move.

R_6 says that each dialogue begins with either a claim or an argument. The initial claim or, if a dialogue starts with an argument, its conclusion is the *topic* of the dialogue. R_7 restricts concessions of an argument's conclusion to conclusions of counterarguments. This ensures that propositions are conceded at the place in which they were introduced. Consider the following dialogue:

P_1:	p *since* q
O_2:	*why* q
P_3:	q *since* r
$O_4[P_3]$:	*concede* q

R_7 invalidates O_4 as a reply to P_3; it should instead be targeted at P_1, which is when the proponent introduced q.

Definition 9 A *dialogue system for liberal dialogues* is now defined as any dialogue system with L_c as specified in Table 1, with turntaking rule T_L and such that a move is legal if and only if it satisfies protocol rules R_1-R_7.

Note that systems for subsets of L_c can be defined as slight variations of systems for liberal dialogues by simply declaring the use of certain moves illegal at all times.

4 Termination and outcome of dialogues

Next termination and outcome of dialogues must be defined. In practice, termination of dialogues is often conventional so that an 'any time' definition of a dialogue outcome is called for.

4.1 Termination of dialogues

In the philosophical literature on two-party-persuasion, the most usual termination criterion is that a dialogue terminates if and only if the opponent concedes proponent's main claim or the proponent retracts his main claim. Above in Definition 7 instead the usual 'mathematical' approach was followed, in which a dialogue is defined as terminated just in case no legal continuation is possible. So to capture the 'philosophical' definition, a dialogue system should ideally be defined such that the players run out of legal moves just in case the main claim is conceded or retracted.

However, more can be said about termination of dialogues. In general the individual knowledge bases of the players will evolve during a dialogue: the players may learn from each other, they may ask advice of third parties, or they may perform other knowledge-gathering actions, such as consulting databases or making observations. For this reason, a player will rarely run out of attacking moves, since it is (theoretically) always possible to find an argument for a claim or a counterargument to an argument. So it will rarely be possible to *force* the other player to concede or retract the main claim. In addition, a ‘filibustering’ player can always challenge the premises of any new argument. For these reasons realistic dialogues will often not terminate by retraction or concession of the main claim, but by external agreement or decision to terminate it, so formal termination results are of limited practical value.

4.2 Outcome of dialogues

When the traditional philosophical termination rule is adopted, the obvious outcome rule is to declare proponent the winner if opponent has conceded his main claim and to declare opponent the winner if proponent has retracted his main claim. The winner can then be defined such that the proponent wins if the opponent has conceded his main claim and the opponent wins if the proponent has retracted his main claim.

However, for dialogues that can terminate by convention this may in certain contexts be too restrictive; a player may avoid losing simply by never giving in and continue debating till the other player becomes tired and agrees to terminate. To deal with contexts where this is undesirable, ‘any time’ outcome definitions need to be studied, which allocate ‘burdens to attack’ to the players, so that if at a certain dialogue stage a participant has not yet fulfilled his burden to attack, he may be the ‘current’ loser even if he has not conceded (opponent) or retracted (proponent) the main claim. Besides for identifying the current winner, such a notion can also be used to regulate turntaking and to define relevance of moves, as will be explained in detail in Section 6.

Consider the following simple liberal dialogue:

P_1: *claim* p
O_2: *why* p

At this stage it seems reasonable that P’s main claim is not successfully defended, since there is an unanswered challenge. So P has the burden to attack this challenge on the penalty of being the current loser. Suppose P fulfills this burden with

P_3: p *since* q

Then it seems reasonable to say that P’s claim is successfully defended, since its only challenge has been met, so the burden to attack has shifted back to the opponent.

One way to define an ‘any time’ outcome notion is simply to apply a ‘black-box’ logical proof theory for $\mathcal{L}$ to the premises of all arguments moved at a certain dialogue stage that are not challenged or retracted. If the main claim is justified in $\mathcal{L}$ on the basis of these premises, proponent is the current winner, otherwise opponent is the current winner. This is the approach taken by, for example, Gordon [13], Loui [20] and Brewka [8]. Earlier [30] I argued that a more natural approach is to incorporate the proof theory of $\mathcal{L}$ into the dialogue protocol as much as possible, and then to prove that the dialogue outcome corresponds to what logically follows. Thus the protocol is arguably more realistic as a model of human dialogues, which may be beneficial in several contexts.

In the following section this approach is formally defined.

4.3 Dialogical status of moves

Next an any-time outcome notion that does not appeal to a black-box logical consequence notion will be defined. The definition is in terms of a pragmatic notion of the *dialogical status* of a move, which formalises the informal ideas explained in the previous subsection. In particular, a move that is (or is not) successfully defended against an attacking reply is said to be *in* (or *out*). The definition of dialogical status assumes a notion of a surrendered move, which needs to be defined separately for each instantiation of the framework of Section 2.

Definition 10 (Dialogical status of moves) All attacking moves in a finite dialogue d are either *in* or *out* in d. Such a move m is *in* iff

1. m is surrendered in d; or else
2. all attacking replies to m are *out*

Otherwise m is *out*.

Definition 11 (The current winner of a dialogue) The status of the initial move m_1 of a dialogue d is *in favour of* $P(O)$ and *against* $O(P)$ iff m_1 is *in* (*out*) in d. We also say that m_1 favours, or is against p. Player p *currently wins* dialogue d if m_1 of d favours p.

For liberal dialogue systems and all further systems to be discussed in this article the notion of a surrendered move is defined as follows.

Definition 12 A move m in a dialogue d is *surrendered* in d iff

- it is an *argue* A move and it has a reply in d that concedes A's conclusion; or else
- m has a surrendering reply in d.

Proposition 13 For each finite dialogue d there is a unique dialogical status assignment.

A counterexample for infinite dialogues is an infinite sequence of attacking moves $m_1, m_2, \ldots, m_i, \ldots$ each replying to the immediately proceding move. This dialogue has two dialogical status assignments: one in which all even moves are *in* and all odd moves are *out*, and one with the converse assignments.

Now the 'current' winner of a dialogue can be defined as follows:

Definition 14 For any dialogue d the proponent wins d if m_1 is *in*, otherwise the opponent wins d.

5 Soundness and fairness results for liberal dialogues

The question now arises to what extent the any-time winning definition corresponds with the underlying logic. For instance, it may be asked whether if proponent wins a dialogue about topic t, the information currently agreed upon in the dialogue gives rise to a justified argument for t. This is an aspect of the *soundness* of a protocol. The reverse question can also be asked: if the currently agreed information defeasibly implies the dialogue topic, does proponent currently win? This is an aspect of the protocol's *fairness*. In this section these questions and some related questions will be investigated for liberal dialogues.

The basic idea is to prove a relation between the dialogue and the dialectical graph of arguments and counterarguments that is implicitly built during a dialogue. At a first stab, we want to prove that the initial move of a dialogue is *in* just in case the dialectical graph contains a justified argument for the dialogue topic. However, this must be qualified. Some arguments in the graph may have premises that were retracted, or they may have premises that were challenged but for which no further argument was given. Such premises are not defended and arguments using them should not be taken into account. Therefore, we are interested in proving that the initial move of a dialogue is *in* just in case the *defended part* of its dialectical graph contains a justified argument for the dialogue topic. In fact, as will be discussed below, such a result can only be proven for dialogues in which the players play logically 'perfectly'.

5.1 The dialectical graph of a dialogue

Next the idea will be formalised that during a dialogue the players implicitly build a dialectical graph of arguments and counterarguments related to the dialogue topic. The soundness and fairness results will be proven in terms of this graph. The basic idea is that the dialectical graph sometimes contains as a subgraph a winning strategy in the argument game for grounded semantics (recall that strategies for this semantics can be represented by trees of arguments). In this respect it is important to note that the dialectical graph will not simply represent the arguments moved in a dialogue but *occurrences* of such arguments: (parts of) arguments may occur more than once in the graph, namely if they are moved more than once in the dialogue.

When dialogue moves are regarded as operations on a dialectical graph, two such operations must be considered. The first is adding a new argument, either as an argument for an initial claim, or as a counterargument; in the latter case defeat relations are also added. The second operation is 'backward' extending an argument by providing an argument for one of its premises. This second kind of move can operate on more than one argument in the dialectical graph at the same time, as the dialogue in Figure 3 illustrates. (A solid box means that a move is *in* and a dotted box that it is *out*.) In this dialogue proponent gives two alternative arguments for his first 'top level' premise q,

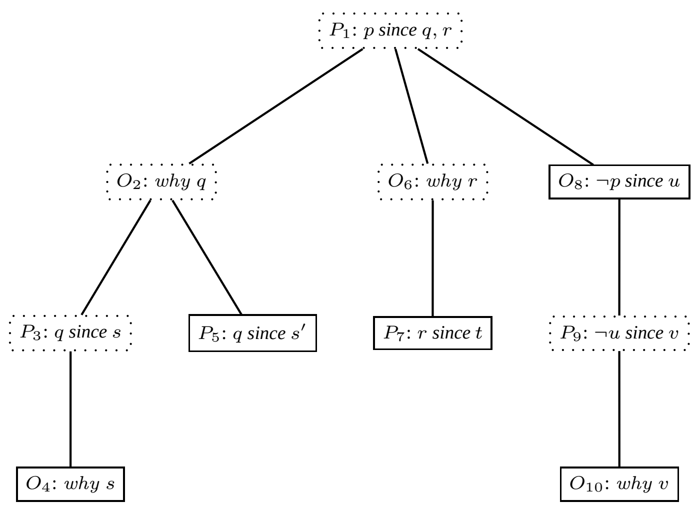

Figure 3: A dialogue tree.

and then gives an argument for his second 'top level' premise r: clearly, this argument for r must be combined independently with both alternative arguments for q. Hence, P_7 extends two arguments at the same time, viz.

$$\frac{\dfrac{s}{q} \quad r}{p} \qquad \text{and} \qquad \frac{\dfrac{s'}{q} \quad r}{p}$$

This example also shows that sometimes a counterargument moved in a dialogue replies to more than one argument in its dialectical graph: O_8 replies to both of proponent's alternative arguments for p. Accordingly, this dialogue induces the dialectical graph of Figure 4.

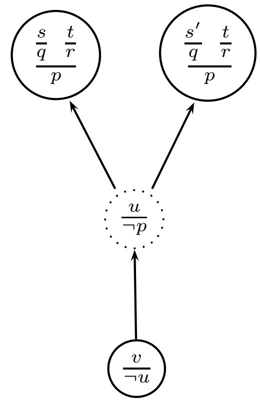

Figure 4: The dialectical graph of Figure 3

Moves that extend an argument are even more involved. Consider an *argue* B reply to a *why* attack on an *argue* A move, such that $conc(B) = \varphi$. If A was already extended before, then the effect of the *argue* B move on the dialectical graph depends on whether:

1. A was extended but not on φ;
2. A was extended on φ.

In the first case, the *argue* B move simply further extends any current extension of A by replacing its leaf node φ with the tree B. (Note that more than one extension of A may exist since A may have been extended in alternative ways on another premise.) In the second case, *argue* B is in fact an alternative way to extend A on φ, so it first copies each current version of A and then extends each such copy on φ with B.

Let us now consider the formal definition of the dialectical graph of a dialogue. First, since separate occurrences of arguments in a dialogue must be individuated, it must be carefully defined when an argument extends another.

Definition 15 An argument B *extends* an argument A in a dialogue d if B was moved as an *argue* B reply to a *why* attack on an *argue* A move. We also say that B extends A on *conc*(B).

Next the set of 'current versions' of arguments moved in a dialogue is defined. For convenience the propositions in an argument will be labelled with the move in which they were moved; when there is no danger of confusion, the definitions below will ignore the difference between labelled and nonlabeled versions of arguments.

Definition 16 The set $Args_d$ of *arguments of dialogue* d contains all trees T satisfying the following conditions:

1. The root of T and its children are of the form m_i:φ, m_i:ψ_1,. . ., m_i:ψ_n such that
   (a) $s(m_i) =$ *argue* A; and
   (b) A does not extend another argument in d; and
   (c) φ and $\psi_1, \ldots, \psi_n$ are the conclusion and premises of A.

2. For any other node $m_i{:}\varphi \in T$ its children are all nodes $m_j{:}\psi$ for an *argue* A move m_j in d such that

    (a) A extends an argument moved in m_i on φ; and

    (b) ψ is a premise of A.

We say that $A \in Args_d$ is a *current version* of A' as moved in an *argue* A' move m if A' is a subtree of A and the nodes of A' are labelled with m.

Now adding the reply relations to the set of arguments of a dialogue yields the dialectical graph of the dialogue.

Definition 17 The *dialectical graph* g_d *of* a dialogue d is the directed graph $(Args_d, R)$ where R is a binary relation on $Args_d$ such that $(x, y) \in R$ iff there exist node labels m in x and m' in y such that m replies to m' in d. If $(x, y) \in R$ we say that x replies to y in g_d.

Proposition 18 For any dialectical graph $g_d = (Args_d, R)$ the relation R is acyclic.

Here it is crucial that if an argument is repeated in a dialogue, its second occurrence is a new node in the dialectical graph. Note also that a dialectical graph may consist of various unconnected parts (for instance, if alternative arguments for the initial claim are moved).

5.2 Other notions concerning dialogues

To formulate the soundness and fairness results, some further notions are needed.

5.2.1 Defended arguments

Recall that we want that the proponent wins just in case the defended part of the theory constructed during the dialogue implies the dialogue topic. This must be made more precise. We are interested in only those arguments of which the premises are not challenged and of which no node is retracted.

Definition 19 An argument $A \in g_d$ is *defended* in d iff:

- for no premise φ of A labelled with m it holds that m has a $why\ \varphi$ reply in d that was not replied-to in d; and
- for no node φ of A labelled with m it holds that m has a $why\ \varphi$ reply in d replied to with a *retract* φ move.

The set *DefArgs*$_d$ of *defended arguments* of d is the set of all arguments that are defended in d. Its closure $Cl(DefArgs_d)$ is abbreviated as C^d.

The defended part of a dialectical graph is now defined as follows.

Definition 20 The *defended part* d_d of a dialectical graph g_d is the dialectical graph $(DefArgs_d, R)$.

By way of illustration the defended part of the dialectical graph of Figure 4 is displayed in Figure 5, where the dialogical status is computed on the basis of the defended arguments only (leaving the move labels implicit). This figure reveals that proponent lost not because he failed to move the right arguments, but because he could not defend the premises of all his arguments.

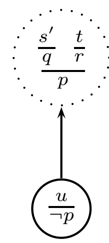

Figure 5: The defended part of Figure 4.

5.2.2 Logical completion of dialogues

An additional subtlety must be explained. As I discussed earlier in [30], actually terminated dialogues might not yet be 'logically' terminated: an argument moved at some earlier stage might also be a legal counterargument against some later argument, or several premises stated during the dialogue give, when combined, rise to an additional legal argument. A simple example of this kind is displayed in Figure 6. Consider first dialogue 1. Suppose that the premises of arguments A and C,

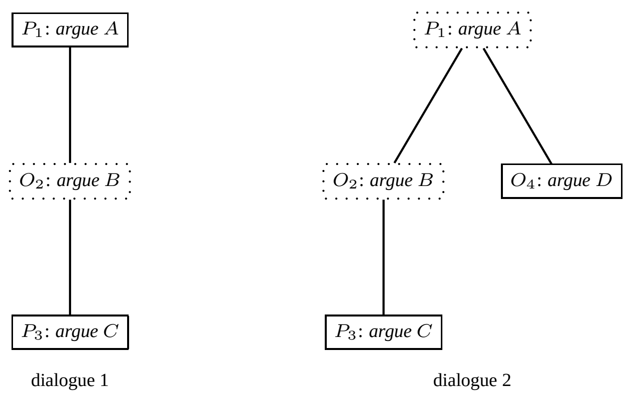

Figure 6: Two dialogues

when combined, enable an argument D, which is a legal reply to P_1. Then the argument graph of dialogue 1 (which is equal to dialogue 1) is not a winning strategy for P since it does not contain all replies to P_1 so that it is not a strategy. So even though P_1 is *in*, the currently agreed information does not defeasibly imply the dialogue topic. Note that Dialogue 2, although a strategy for P, is not a winning strategy. In [30] it is shown that with a suitable underlying logic the example can be constructed such that C and D are identical, which shows that the problem arises even if the outcome definition is restricted to the arguments actually moved in a dialogue.

This example illustrates that the soundness and fairness results must be made conditional on the assumption that the players have played logically 'perfectly', i.e., that they have moved all such 'implied' arguments. To make this precise, the following notions are needed:

Definition 21 A dialogue d is *logically completed* iff for all minimal arguments $A \in Cl(Args_d)$ such that A defeats an argument B in $Args_d$, it holds that A is a child of all occurrences of B in g_d.

The minimality condition is needed since otherwise no realistic dialogue will be logically closed, since arguments can always be extended in irrelevant ways to make a new argument. Even so logical closure is quite a strong property of dialogues. Yet it is a realistic notion since nothing prevents a player from losing a dialogue by mistake, for example, by unnecessarily conceding or withdrawing the dialogue topic, or by failing to state an available counterargument.

5.2.3 Winning parts of a dialogue

Next it is useful to identify the part of a dialogue that 'makes' the initial move have its dialogical status.

Definition 22 Let d be a dialogue currently won by player p. A *winning part* d^p of d is recursively defined as follows.

1. First include m_1;
2. for each move m of p that is included, if m is surrendered, include all its surrendering replies, otherwise include all its attacking replies;
3. for each attacking move m of $\overline{p}$ that is included, include one attacking reply m' that is *in* in d.

Informally, the idea is that, by omitting all moves of p that are surrenders or from which p has backtracked, d^p contains that part of d that makes p win. In general, d^p is not unique, since p might have moved alternative attacking replies to a move, neither of which were successfully challenged by $\overline{p}$.

The following facts hold about winning parts:

Proposition 23 A dialogue d contains a winning part for p just in case p currently wins d.

Proposition 24 Of any winning part d^p the leaves are either surrenders by $\overline{p}$ or attackers by p.

Proposition 25 Of any winning part d^p all moves of p are *in* and all moves of $\overline{p}$ are *out* in both d and d^p.

5.3 A soundness and fairness result

The first soundness and fairness result can now be stated as follows. Observe that it is proven only for finite dialogues and only for dialogues in which the players do not move surrenders (because surrenders can give up a 'winning' game without having to).

Proposition 26 Let d be a logically completed finite liberal dialogue without surrenders. Then P currently wins d iff g_d contains an argument for the dialogue topic that is justified on the basis of C^d.

Since the proof of the only-if part of this proposition in fact only depends on logical completeness of a winning part for P, it can be strengthened as follows.

Corollary 27 Let d be a finite liberal dialogue without surrenders currently won by P such that at least one winning part for P is logically closed. Then g_d contains an argument for the dialogue topic that is justified on the basis of C^d.

Basically, the $\Rightarrow$ part of the theorem tells us that for determining the winner of a dialogue in which the players have played logically perfectly, no black-box nonmonotonic theorem provers need to be invoked, since the dialogical status of the initial move always 'respects' the defeasible consequences of the dialogue's defended information base. The $\Leftarrow$ part tells us that the protocol is 'fair' in that when the information base created during a logically completed dialogue implies that proponent should win, proponent in fact wins.

The result does not hold without the condition that each dialogue is logically completed. Figure 6 above provides a simple counterexample.

5.4 Other aspects of soundness and fairness

Logical completion is quite a strong condition for dialogues. The question therefore arises under which conditions it is possible to logically complete a dialogue. More precisely, suppose that the closure C^d of the set of defended arguments of a dialogue d makes proponent's argument justified but the dialogue is not logically completed; suppose also that from now on the players are only allowed to move arguments from C^d: can proponent 'ultimately win' the dialogue, that is, can he make opponent run out of legal moves in any such continuation of d? This is another aspect of the fairness of a protocol. The converse question can also be asked: suppose proponent can ultimately win a dialogue if the players may only move defended arguments: is the dialogue topic implied by C^d? This can be regarded as a further aspect of a protocol's *soundness*.

In [30] these questions were studied for argument games, i.e., dialogue games in which only arguments can be moved. To answer these questions for protocols with additional locutions, some notions must be made more precise.

Definition 28 A *logical completion* of a dialogue d of type t is a continuation of d according to the protocol for t-dialogues in which at any dialogue stage d' only *argue* A moves are moved and where all such A are in $C^{d'}$. A logical completion is *terminated* if no legal move of this kind can be made. A player p *wins* a logical completion d' of d if d' is terminated and the initial move favours p.

Note that the set of defended arguments can grow during a logical completion, viz. if a defended argument is moved to answer an as yet unanswered *why* question. Note also that a terminated logical completion usually is not a terminated dialogue according to Definition 7, since the players will usually be able to move other kinds of moves, such as *why*, *concede* and *retract* moves, or *argue* moves with arguments composed of new information.

Proposition 29 Let d be a finite liberal dialogue without surrenders. If P has a winning strategy at the start of a logical completion d' of d, then C^d contains a justified argument for the dialogue topic.

Proposition 30 Let d be a finite liberal dialogue without surrenders such that C^d is finitary. If C^d contains a justified argument for the dialogue topic and during any logical completion of d the set C^d remains constant, then P has a winning strategy at the start of a logical completion of d.

Proposition 30 does not hold without the condition that the set of defended arguments remains constant during a logical completion. To sketch a counterexample, suppose g_d contains two arguments A and B such that B replies to A but only A is defended. Then the proponent trivially has a winning strategy for A in an argument game on the basis of C^d. Suppose also that C^d contains an argument C which can be moved in reply to a challenge of B's premise: then if opponent moves C to extend B, it may be that both A and the combination of B and C are defended and then proponent does not have a winning strategy for A any more on the basis of the new C^d.

6 Protocols for relevant dialogues

In this section it will be shown that liberal protocols only weakly enforce structural coherence of dialogues and then two notions of strong and weak (structural) relevance of moves will be defined that remedy this.

6.1 Motivation

Liberal protocols promote relevance through the protocol rules R_2, R_4 and R_5. According to these rules, Figure 7 displays two legal liberal dialogues, sharing the first three moves. In dialogue 1 move

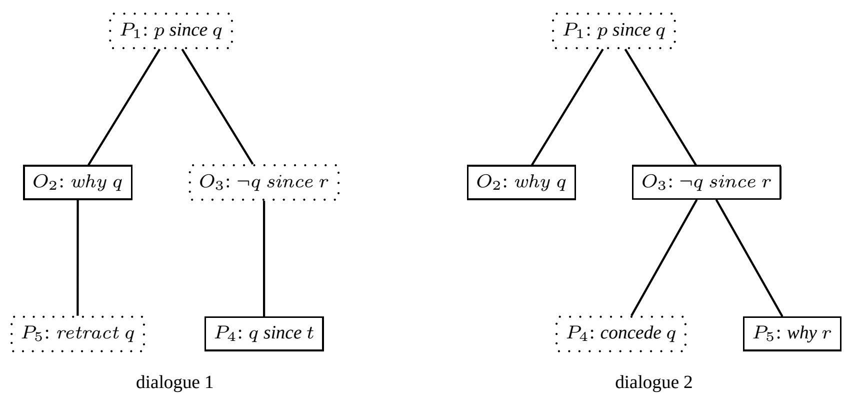

Figure 7: Two somewhat incoherent dialogues.

O_3 is in a certain sense superfluous since with O_2 the opponent already launched another attack on P_1. Also, P in his second turn first attacks O's argument for $\neg q$ with P_4 and then retracts q with P_5. Arguably, this behaviour of the proponent is not very coherent: first he counterattacks an attack on his initial argument and then he surrenders to another attack on that argument. Dialogue 2 displays a variant of such rather incoherent behaviour: proponent first concedes the conclusion of O_3's argument with P_4 and then attacks its premise with P_5.

Figure 8 displays another legal liberal dialogue that is not entirely coherent. Here the opponent in his third turn attacks with O_7 in a line of the dialogue which P has meanwhile implicitly retreated with P_5: his current reason for q is v. Arguably O's argument for $\neg u$ is at this point irrelevant for the dialogue topic.

These dialogues illustrate that there is a need for stricter protocols, where each move is relevant to the dialogue topic. A rigorous way to enforce relevance of moves is to have a unique-move and unique-reply protocol. Then in the dialogues of Figure 7 O's first turn ends after his challenge of q, after which P has to choose between retracting q or defending it. And in Figure 8 the dialogue ends after P_5's retraction and P is penalised for making the 'wrong' choice of argument at P_3. However, this comes with a price. Firstly, as argued in the introduction, it may not in all contexts be fair to disallow the players to repair mistakes or to move alternative arguments in the same turn. Secondly, there are more subtle reasons to allow backtracking. Consider the implied arguments of a dialectical graph. Fairness demands that when such arguments are relevant for the outcome of a dialogue, their moving should be legal at least at one stage during the dialogue. However, as shown in [30], the more a protocol restricts the possibility to move alternatives to earlier moves, the more it runs the

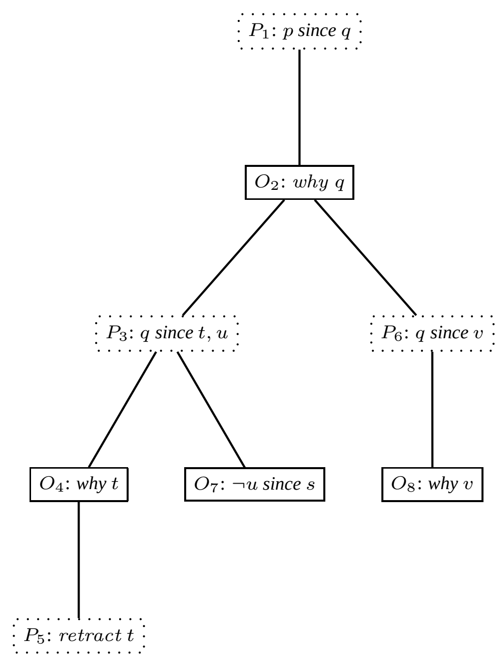

Figure 8: Another somewhat incoherent dialogue.

risk of making such arguments illegal at any stage. Consider again Figure 6 and assume that the protocol is unique-reply, that is, it disallows backtracking. In dialogue 1 after P_2 it is too late for opponent to move D, so although D is implied by the information base of the dialogue, it can at no stage be moved. This example shows that unique-reply protocols can be unfair.

In general, any 'any time' definition of the dialogue outcome can be used to constrain turntaking and promote relevance: for instance, protocols could be made immediate-reply and all moves can be required to have an effect on the outcome of the dialogue. These ideas will now be made more precise in terms of the dialogical status of moves.

Intuitively, a replying move is structurally relevant if it is capable of changing the dialogical status of the initial move, given the various ways the players have backtracked and surrendered. Two typical grounds for irrelevance of a move are that it is made in a dialogue branch from which the other adversary has retreated (cf. move O_7 in Figure 8), or in a dialogue branch containing a surrendered move, of which the status therefore cannot be changed (cf. move P_5 in dialogue 2 of Figure 7).

6.2 Relevance defined

The requirement that each move be relevant allows the players maximal freedom on issues such as backtracking and postponing replies while yet ensuring a strong focus of a dialogue. The present notion of relevance extends the one of [30], which only applied to argument games.

As for the formal definition of relevance, as just explained, what is crucial is a move's effect on the status of the initial move. In order to determine relevance of surrendering moves, their effect is checked as if they were their attacking counterpart. Thus a move is relevant iff any attacking counterpart with the same target would change the status of the initial move of the dialogue. This is formally defined as follows.

Definition 31 (Relevance) An attacking move in a dialogue d is *relevant* iff it changes the dialogical status of d's initial move. A surrendering move is relevant iff its attacking counterparts are relevant.

Together the above definitions imply that a reply to a surrendered move is never relevant. Note also that, if not surrendered, an irrelevant target can become relevant again later in a dialogue, viz. if a player returns to a dialogue branch from which s/he has earlier retreated.

To illustrate these definitions, consider Figure 9 (where + means *in* and - means *out*.). The dialogue tree on the left is the situation after P_7. The tree in the middle shows the dialogical status of the moves when O has continued after P_7 with O_8, replying to P_5: this move does not affect the status of P_1, so O_8 is irrelevant. Finally, the tree on the right shows the situation where O has instead continued after P_7 with O'_8, replying to P_7: then the status of P_1 has changed, so O'_8 is relevant.

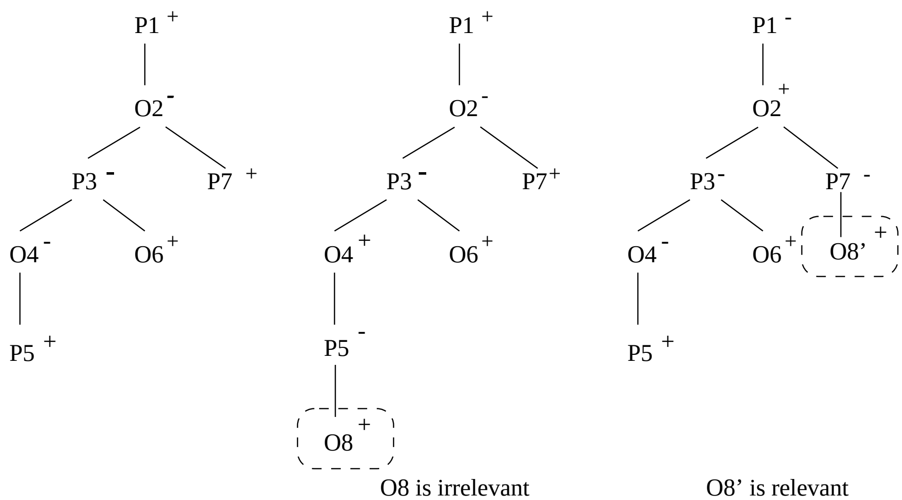

Figure 9: Dialogical status of moves.

To be a protocol for relevant dialogue, a protocol must also satisfy some additional conditions on the notions of move legality, turntaking and winning. The following protocol rule is added to those of liberal dialogue systems.

If $m \in Pr(d)$, then:

- R_8: if m is a replying move, then m is relevant in d.

To prevent premature termination of a dialogue this rule must be combined with an immediate-reply turntaking rule (cf. [20]):

- T_i: $T(d_0) = P$ and if $d \neq d_0$ then $T(d) = p$ iff $\overline{p}$ currently wins d.

Together, R_8 and T_i enforce that when a player is to move, s/he keeps moving until s/he has changed the status of the initial move his or her way (since after such a change no further move of the same player can be relevant). In other words, each player first moves zero or more relevant surrenders, and then moves zero or one relevant attacker: if no attackers are moved, this is because the player has no legal moves.

Definition 32 A *dialogue system for relevant dialogues* is any dialogue system with L_c as specified in Table 1, with turntaking rule T_i and such that a move is legal if and only if it satisfies protocol rules R_1-R_8.

6.3 Properties of relevant dialogues

Proposition 23 can be strengthened for relevant dialogues.

Proposition 33 Every relevant dialogue d contains a unique winning part.

About soundness and fairness, it is easy to see that Proposition 26 and Corollary 27 still hold for relevant dialogues. The key is the condition of logical completion, which excludes dialogues from consideration where irrelevance prevents the moving of an implied argument.

Proposition 34 Let d be a logically completed finite relevant dialogue without surrenders. Then P currently wins d iff g_d contains an argument for the dialogue topic that is justified on the basis of C^d.

However, the condition of logical completion is even stronger for relevant than for liberal dialogues, since it is easy to imagine dialogues where moves in a logical completion are irrelevant. Therefore, it becomes even more important to investigate the other two properties for relevant dialogues.

Proposition 35 Let d be a finite relevant dialogue without surrenders. If P has a winning strategy at the start of a logical completion d' of d, then C^d contains a justified argument for the dialogue topic.

Proposition 36 Let d be a finite relevant dialogue without surrenders such that C^d is finitary. If C^d contains a justified argument for the dialogue topic and during any logical completion of d the set C^d remains constant, then P has a winning strategy at the start of a logical completion of d.

6.4 Protocols for weakly relevant dialogues

Comparing systems for liberal and for relevant dialogues, the main advantage of the relevance requirement is that it keeps a dialogue focussed by ensuring that no resources are wasted on 'superfluous' moves, i.e., moves that have no bearing on the status of the initial move. However, there are reasons to study a weaker sense of relevance. Perhaps the main drawback of the relevance condition is that it must be combined with an immediate-reply turntaking rule, which prevents the moving in one turn of alternative ways to change the status of the initial move. This may be a drawback, for instance, in discussions where the parties cannot immediately reply to each other, and therefore reply to all moves of the preceding turn (as in parliamentary debate).

This disadvantage of the relevance rule can be met with a weakening of the notion of relevance, to require only that each attacking move creates a new winning part of the speaker or removes a winning part of the hearer.

Definition 37 (Weak relevance) An attacking move in a dialogue d is *weakly relevant* iff it creates a new winning part of d for the speaker or removes a winning part of the hearer. A surrendering move is weakly relevant iff its attacking counterparts are weakly relevant.

Relevance according to Definition 31 will now be called *strong relevance*. Clearly, each strongly relevant move is also weakly relevant. The relevance rule is now weakened as follows:

- R'_8: if m is a replying move, then m is weakly relevant in d.

Finally, the turntaking rule is relaxed as follows.

- T_w: $T(d_0) = P$. If $d_i \neq d_0$ then $T(d_i) = pl(m_i)$ if $\overline{pl(m_i)}$ currently wins d and $T(d_i) = \{P, O\}$ if $pl(m_i)$ currently wins d.

This says that interrupting the speaker is allowed but not obligatory as soon as the speaker has made himself the current winner.

Definition 38 A *dialogue system for weakly relevant dialogues* is any dialogue system with L_c as specified in Table 1, with turntaking rule T_w and such that a move is legal if and only if it satisfies protocol rules R_1-R_7, R'_8.

The structure of weakly relevant dialogues differs in two main respects from that of strongly relevant dialogues. Firstly, a player has some freedom to make additional moves after he has made himself the current winner, possibly creating additional winning parts. Secondly, each player must counterattack all attacks of the other player in order to make himself the current winner.

It is straightforward to prove that the soundness and fairness results for liberal and relevant dialogues still hold for weakly relevant dialogues.

To illustrate the weak notion of relevance, consider again the dialogue between Paul and Olga from the introduction. A weakly relevant version of this dialogue is when Olga moves her second argument for 'not safe' directly after her first in the same turn, after which Paul must attack both of them in his next turn to change the status of his main claim.

7 Other possible protocol rules

So far liberal protocols formulate basic requirements for coherence of dialogues and weak and strongly relevant protocols add to this two ways to strengthen these requirements. Many more protocol rules are possible and in this section a few of them will be discussed. It is important to note, however, that whether these rules make sense depends on the context and nature of the application. Some relevant factors are the available resources and the importance of the interests that are at stake. Also, when the players are humans they can often be assumed to be both rational and cooperative so that much can be left to their individual strategies and conventions. On the other hand, when the players are pieces of software with programmed self-interest, it may be necessary to explicitly model and implement such strategies and conventions, since such software agents will explore the entire space of possibilities to avoid losing.

Thus four categories of dialogue rules can be distinguished. *Basic* protocol rules should be respected in all discussions. *Context-dependent* protocol rules hold whenever this is appropriate in a certain context of application. *Conventions* formulate behaviour that participants should ideally have to promote coherence of the dialogue, for example, 'do not challenge too much or retract too rapidly'. Finally, *player strategies and heuristics* are meant to promote the player's individual goal, viz. to win the dialogue: for example, 'no irrational surrenders', that is, surrenders when one is not losing. The assertion and concession attitudes of [25] (for example "concede a proposition only if you cannot construct a counterargument") can also be regarded as player heuristics.

7.1 Respecting commitments

A further means to promote coherence of dialogues is by using the players' commitments in regulating move legality. Players can, for instance, be required to keep their own commitments consistent or restore consistency upon demand, or not to challenge their own commitments. In this section the addition will be studied of protocol rules referring to commitments to the protocols discussed so far. However, since commitments are a topic of their own, the discussion will be restricted to some simple rules and a more advanced treatment of commitments will be left for future research.

If $m \in Pr(d)$ and $pl(m) = p$, then:

- R_9: $C_p(d, m)$ is consistent.
- R_{10}: If $s(m) = why\ \varphi$, then $C_p(d) \not\vdash \varphi$.

R_9 ensures logical consistency of the players' commitments and R_{10} prevents players from challenging a proposition to which they are themselves committed.

As is easy to verify, the proofs of Propositions 26 and Propositions 34 are not affected by the addition of these two rules. Again the key is the condition that a dialogue is logically completed, which excludes dialogues from consideration where a player's commitments prevent the moving of an implied argument. On the other hand, this makes the condition of logical closure an even stronger one than for dialogue systems without protocol rules on commitments. As a result, the other fairness and soundness properties do not hold when R_9 and R_{10} are added. The dialogue in Figure 10 about topic p provides a counterexample to Proposition 36 for strongly and weakly relevant protocols (similar examples can be constructed against Propositions 35 and 29). Let $d = P_1, \ldots, O_6$.

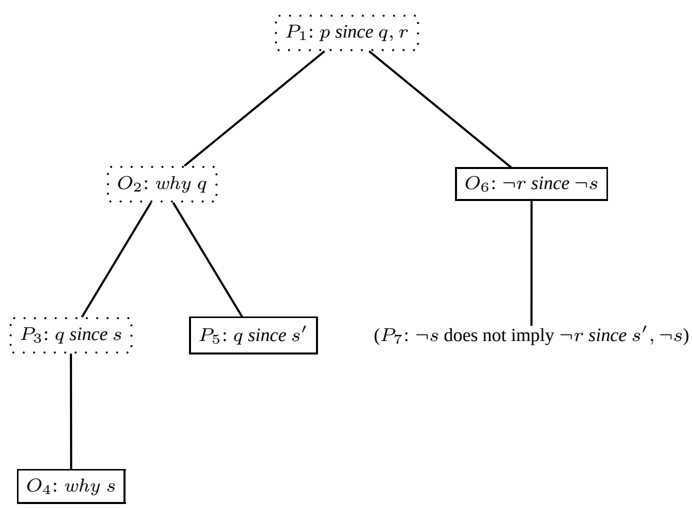

Figure 10: A counterexample to fairness with commitments.

Suppose that C^d contains an undercutting defeater of O_6 (displayed between parentheses as P_7) and supports no *argue* move for the opponent after P_7. Then proponent's initial argument is justified on the basis of C^d. Yet he has no winning strategy in a logical completion of d since P_7 makes his commitments inconsistent. Allowing the players to move retractions in logical completions will not help since retracting s in reply to O_4 is irrelevant. Note that this dialogue is also a counterexample for liberal protocols since Proposition 30 assumes that dialogues contain no surrenders. It remains to be investigated whether allowing players to retract in logical completions can restore fairness of liberal protocols.

It should be noted that in the present setup there is a tension between the effects of moves as replies to targets, which are local, and their effects as operations on commitment sets, which are global. As a reply to a target, a move's direct effect is on the target's dialogical status. As an operation on commitments, a move's direct effect is on the speaker's commitment set, which is global to the dialogue. The dialogue in Figure 11 illustrates this tension. After P_8 the protocols of this

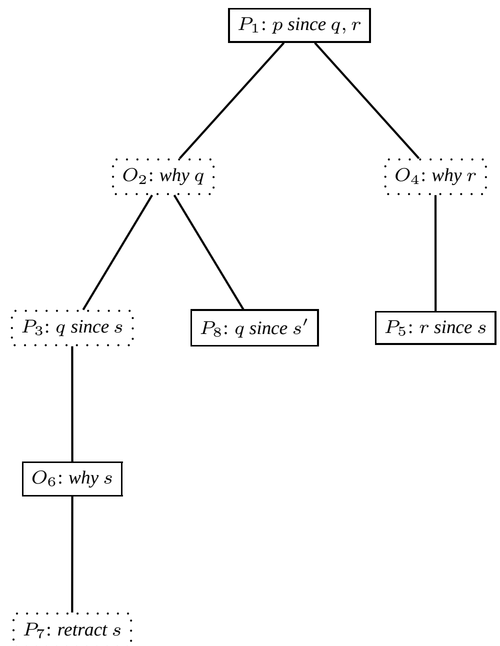

Figure 11: Replies vs. commitments.

study allow opponent to continue with challenging s in P_5 and proponent to reply to this challenge with an argument for s. This dialogue is not very coherent since proponent already retracted s. Additional conventions could be added that reflect the global nature of commitments, for example, that a proposition may only be supported with an argument if the speaker is still committed to the proposition, or that once a proposition is retracted, no new commitment to the same proposition may be incurred by the same player.

7.2 Conventions on retraction

As noted by [34] and [17], the more a protocol allows retractions, the less coherent dialogues may become, since a player can always retreat and try something else when he sees he is in trouble, thus obstructing the dialogue goal of conflict resolution. In the present approach, commitments are 'sticky' in that a claim or premise can only be retracted if it is challenged. Also, in relevant dialogues even challenged claims of premises may sometimes not be retracted, viz. when they are irrelevant. This is not as innocent as it seems, since sometimes an irrelevant retraction would, when made, open new ways of attack for the speaker, as is illustrated by Figure 10 above. So in the present setup claims and backtracking moves come with a price. This will encourage rational agents to be careful in stating their claims and backtracking from certain lines of discussion. In addition, a further protocol rule might be added that disallows retractions before the challenge it replies to is attacked with an argument and before that argument is attacked in some way.

7.3 Circularity of reasoning

Another issue is circularity of reasoning, which has two aspects. An individual argument as constructed by a player may be circular, and a player's behaviour may be 'dialectically circular' in that a proposition defended by an argument is reused as a premise of a counterattack if the initial argument is attacked. [21]'s way of preventing circular reasoning was to disallow the reuse of a proposition that is 'under challenge' i.e., that was challenged by the other player and who has not yet given up the challenge. However, this only addresses the first aspect of circularity, and it does not work for dialogues with alternative replies. Besides, it remains to be seen whether all circular arguments should be prevented. For instance, [25] allow a premise taken from a fixed knowledge base to be supported with itself, as a way to express that this premise is assumed known in the dialogue.

8 Related research

As noted in the introduction, there are three main streams in the formal study of argumentation dialogues, viz. philosophical logic/argumentation theory (AT), artificial intelligence and law (AI & Law), and multi-agent systems (MAS). In addition, there is some work in general AI. As for the use of logic, the underlying logic of dialogue systems in AT is monotonic; the only way to attack an argument is by challenging its premises and the winner of a dialogue is determined by applying classical logic to the players' commitment sets (see Section 4 above). The latter is still the case in MAS-research, although that does use nonmonotonic logics in the internal design of the dialogue participants, to apply heuristics called 'assertion' and 'acceptance attitudes' (for example, "claim or accept a proposition only if you have a justified argument for it in your own knowledge base"). In AI & Law and general AI the logic used in the protocols is nonmonotonic: counterarguments are possible and the 'current' outcome of a dialogue is defined in terms of a nonmonotonic logic.

Next some of the main systems will be discussed in more detail. The discussion of the speech acts will for ease of comparison use the terminology of the present framework instead of that used by the authors, when they deviate.

Two examples of AT research are [21] and [34]. Mackenzie's system has been historically influential, mainly for its set of locutions. Of the locutions discussed in the present paper he has the *claim*, *why*, *concede* and *retract* locutions. Arguments are moved implicitly, by replying to a *why* move with a *claim*. In addition, Mackenzie has a *question* speech act, which asks the hearer to declare a standpoint with respect to a proposition, and a *resolve* speech act for demanding resolution of conflicts in or logical implication by commitments. MacKenzie does not define outcomes or termination of dialogues. In fact, this makes his system underspecified as to the dialogue goal, so that it can be extended to various types of dialogues. The protocol is unique-move and unique-reply but it nevertheless hardly enforces coherence of dialogues. Only the moves required after *why* and *question* and the use of the *resolve* move are constrained; the participants may freely exchange unrelated claims, and may freely challenge, retract or question. For instance, the following dialogue is legal: P: *claim* p, O: *claim* q, P: *question* r, O: *claim* $\neg r$, P: *retract* s.

Walton & Krabbe [34] add an explicit *argue* locution to those of [21]. However, the only way to attack an argument is by challenging its premises so the underlying logic is monotonic. The dialogues allowed by [34] are much more focussed than Mackenzie's, even though this allows that more than one move is made in one turn and alternative arguments for the same challenged proposition are moved. This is achieved by a mixture of rules implicitly based on the reply structure of the locutions and rules based on respecting commitments. However, each move from the last turn must be replied-to (though other moves may be made as well), so that backtracking and postponement of replies are impossible. On the other hand, the absence of an explicit reply structure makes the protocol more liberal than the ones of the present paper in several respects. For instance, it allows retractions of commitments even if they were not challenged.

Gordon's work on the *Pleadings Game* [13] is seminal AI & Law work on the modelling of legal procedures as dialogue games. The Game was intended as a normative model of civil pleading in Anglo-American legal systems, where the participants aim to identify the issues to be decided in court. The logic is conditional entailment [11], which has a model-theoretic semantics and an argument-based proof theory. The logic is used among other things to determine the winner at termination by checking whether the dialogue topic is defeasibly implied by the shared commitments of the participants. The game contains speech acts for conceding and challenging a claim, for stating and conceding arguments, and for challenging challenges of a claim. The latter has the effect of leaving the claim for trial. My distinction between attacking and surrendering replies is implicit in Gordon's distinction between three kinds of moves that have been made during a dialogue: the *open moves*, which have not yet been replied to, the *conceded moves*, which are the arguments and claims that have been conceded, and the *denied moves*, which are the claims and challenges that have been challenged and the arguments that have been attacked with counterarguments. The protocol is multi-move but unique-reply. At each turn a player must respond in some allowed way to every open move of the other player that is still 'relevant' (in a sense similar but not identical to that of Section 6), and may reply to any other open move. If no allowed move can be made, the turn shifts to the other player, except when this situation occurs at the beginning of a turn, in which case the game terminates. Move legality is further defined by specific rules for the various speech acts, which are mostly standard.

MAS-researchers from Toulouse and Liverpool have developed an approach to specify dialogue systems for various types of dialogues. I here focus on their systems for persuasion, taking the most recent papers ([25, 24]) as the basis for discussion. The logic is [1]'s argument-based nonmonotonic logic, which is an instance of [10]'s grounded semantics. The communication language consists of claims, challenges, concessions and questions. Arguments are moved implicitly as *claim* replies to *why* moves (where sets of propositions may be claimed). The protocols largely conform to an implicitly assumed reply structure on L_c and have a rigid, unique-move and unique-reply nature (except that each premise of an argument may be responded to in turn). There is special attention for the internal structure of the participants and their dialogical behaviour. Participants have their own, possibly inconsistent belief base and they are assumed to adopt an assertion and acceptance attitude, which they must respect throughout the dialogue. Also, claims moved in support of other claims must be from the participant's internal belief base, so that arguments 'bottom out' in one step. Unlike in all other work discussed thus far, various properties of the protocols and their outcomes are formally investigated. Finally, this research strand is interested in the combination of argumentation with other types of dialogues, such as information exchange, negotiation and inquiry. In [23] a formal framework is presented for combining systems for different types of dialogues.

I end with a discussion of two studies in general AI.

Influential early work applying dialogue systems to defeasible reasoning was done by Ronald Loui [20] (written in 1992). My framework extends or adapts several notions of Loui's approach, in particular the idea that the dialogical status of the initial move can influence turntaking. Loui also explores how protocols can vary in their structural aspects. On the other hand, he pays less attention to the speech act aspects of argumentation dialogues, which makes his work essentially a study of argument games for defeasible reasoning.

Finally, Brewka [8] has studied in detail how protocols for multi-party dialogues with a referee can be formalised in Reiter's variant of the situation calculus [32]. Brewka uses his prioritised default logic [7] in a 'black box' way to formulate an 'any-time' winning criterion. The fact that his protocols are completely logically formalised allows him to model within a protocol dialogues on whether a move is legal according to that protocol. His formalisation method for protocols paves the way for formal verification of their properties.

9 Conclusion

This article has aimed to make two main contributions to the study of dialogues for argumentation.

The first is a framework for expressing and studying various degrees of flexibility and coherence of dialogues for argumentation, especially concerning relevance and focus of moves. By assuming an explicit reply structure on dialogues, by distinguishing between attacking and surrendering replies, and by defining an any-time outcome notion in terms of a move's dialogical status, various more or less strict protocols could be formulated as regards turntaking, relevance and focus. The framework of this article also allows for some freedom on the choice of underlying logic and the communication language.

The second contribution of this article has been a study of the relation between dialogue protocols and their underlying logic. The presented protocols all implicitly incorporate an argument-game style proof theory for a large class of nonmonotonic logics. This has made it possible to regard an argumentative dialogue as implicitly building a structure of arguments and counterarguments related to the dialogue topic, which in turn allowed the study of various soundness and fairness properties of protocols with respect to the underlying logic. More generally, this article has thereby provided a link between the (static) inferential aspects and the (dynamic) investigative aspects of defeasible reasoning.

The present approach also has some limitations. Most importantly, the requirement that each dialogue move indicates its target excludes some natural disputational strategies, such as lines of questioning in cross-examination of witnesses with the goal of revealing an inconsistency in the witness testimony. Typically, such lines of questioning do not indicate from the start what they are aiming at, as in

Witness: Suspect was at home with me that day.
Prosecutor: Are you a student?
Witness: Yes.
Prosecutor: Was that day during summer holiday?
Witness: Yes.
Prosecutor: Aren't all students away during summer holiday?

It may be that an approach with an explicit reply structure is more suitable for 'verbal struggles' but less suitable for dialogues where investigation and inquiry are more important than 'winning'.

Secondly, as remarked in Section 7, in the present approach there is a tension between the (local) reply structure on dialogue moves and the global nature of commitments. There are several styles to formulate a dialogue system. In this paper a style has been studied in which the reply structure of moves is made explicit, and in which move legality is to a large extent defined in terms of the reply structure of a dialogue. Another style is to define move legality largely in terms of the players' commitment sets (e.g. [21, 13, 34]). Then, for instance, a move *why* φ is not legal if the hearer is not committed to φ or *retract* φ is illegal if the speaker is not committed to φ. The reply approach is especially suitable for allowing degrees of flexibility of coherence, while the commitment approach is especially suitable for maintaining global coherence of dialogical behaviour. It remains to be investigated to which extent the strong points of both approaches can be combined.

As for future research, specific suggestions were made throughout this article. I now briefly summarise them and add some new topics.

Firstly, within the present approach extensions of the communication language with new locutions could be investigated, as well as additional contextual protocol rules and dialogue conventions. Also, more principles of sound, fair and effective protocols for persuasion could be formulated and formally verified. It would also be interesting to extend the present approach with a neutral third party (see [29] for initial ideas) or to dialogues with an arbitrary number of participants. Secondly, pros and cons of the present approach should be compared in more detail with those of other ways to formulate dialogue systems, especially those without a clear (implicit or explicit) reply structure on the communication language. Finally, once the rules of the 'argumentation game' have been defined, the question arises how this game can be played well. This amounts to a study of persuasive strategies and tactics for argumentative dialogue (traditionally called 'rhetoric').

A Appendix: Proofs

Proposition 13 For each finite dialogue d there is a unique dialogical status assignment.

Proof The proof is by induction on the tree structure of d. Note first that all leaves of d are either attacking replies or surrenders, and all non-surrendering leaves are *in* since they have no replies. Then for each parent of a leaf there is just one possible status: it is *in* if it is surrendered, otherwise it is *out*. Consider next any node n in the tree such that the status of all its children is uniquely determined: then clearly the same holds for n.

Proposition 18 For any dialectical graph $g_d = (Args_d, R)$ the relation R is acyclic.

Proof Let for any argument $A \in g_d$ the first part of A moved in d be denoted by *top*(A). Consider a sequence $A_1, \ldots, A_n$ of arguments from g_d such that $(A_{i+1}, A_i) \in R$ whenever $1 \leq i < n$. We first prove that *top*(A_i) was moved before *top*(A_{i+1}) in d. Let *conc*(A_{i+1}) be labelled with m_j. Then some node in A_i is labelled m_k where $k < j$, so *top*(A_i) is labelled m_l such that $l \leq k$, so *top*(A_i) was moved before *top*(A_{i+1}) in d. Generalising this, we have that *top*(A_1) was moved before *top*(A_n) in d, so *top*(A_1) cannot have been moved in reply to a part of A_n. Observe next that all extending parts of A_1 were moved in reply to *why* moves, so no part of A_1 can have been moved in reply to a part of A_n. Hence $(A_1, A_n) \notin R$.

Proposition 23 A dialogue d contains a winning part for p just in case p currently wins d.

Proof The if part is obvious. For the only-if part: since all non-surrendering moves by $\overline{p}$ included in the construction of d^p are *out* in d, they have an attacking reply that is *in* in d.

Proposition 24 Of any winning part d^p the leaves are either surrenders by $\overline{p}$ or attackers by p.

Proof Obvious from the construction and finiteness of d^p and the fact that m_1 favours p.

Proposition 25 Of any winning part d^p all moves of p are *in* and all moves of $\overline{p}$ are *out* in both d and d^p.

Proof This follows immediately from Proposition 24 and the construction of d^p.

Proposition 26 Let d be a logically completed finite liberal dialogue without surrenders. Then P currently wins d iff g_d contains an argument for the dialogue topic that is justified on the basis of C^d.

Proof Note first that from the the definition of g_d and the *argue* move in L_c and the assumption on the defeat relation in Definition 2 it follows that when A replies to B in g_d, then A defeats B. This allows us to proceed as follows.

($\Rightarrow$) Consider a logically completed finite liberal dialogue d without surrenders and in which m_1 favours P. I prove that g_d contains as a subgraph a winning strategy of P on the basis of C^d. First, by Proposition 23 there exists a winning part d_P. Call it W and let G be the dialectical graph of W. Clearly, G is a subgraph of g_d. Since W does not split after opponent moves, the only nodes of G with multiple parents are proponent nodes. Then G can be converted into a tree S by splitting all such nodes. I prove that S is a winning strategy for P on the basis of C^d. Suppose first that S is not a strategy. If any node at even depth in G contains more than one reply, then W is not a winning part. So some proponent node in S does not have all defended defeaters from C^d as its children. But then d is not logically closed, which contradicts the assumption that it is. So S is a strategy for P on the basis of C^d. Suppose next that S is not a winning strategy. Then since d is finite, S is also finite so a node $l : B$ of even depth (i.e., an opponent node) in S is a leaf node of S. Since B is defended in d and has no counterarguments in S, all moves m of d labelling any node of B in S are *in* in d^P. But this contradicts Proposition 25.

($\Leftarrow$) Consider a logically closed finite liberal dialogue d without surrenders, and where some argument A for the topic of d is justified on the basis of C^d. Then since d is logically closed, d_d contains as a subgraph a graph that can be converted into a winning strategy S for A by splitting all nodes with multiple parents. A winning part d^P can be constructed as follows. First 'unravel' every node in S according to the stepwise construction of the arguments in d, by adding the appropriate *why* moves between parts of the arguments; join nodes corresponding to the same move in d. This results in d_1. Since all arguments in S are defended, no node in d_1 has an unanswered *why* reply; moreover, since S is a winning strategy for P, all *argue* moves of O in d_1 have an *argue* reply. Then for each *argue* A move of P in d_1 add *argue* B attacks for each undefended argument B in g_d that is a child of a current version of A in S. Unravel these arguments with the corresponding *why* moves and again join nodes corresponding to the same move; this results in d_2. Observe that ultimately for all the arguments added to d_2 some extension has an unanswered *why* reply. So d_2 is a winning part d^P.

Proposition 29 Let d be a finite liberal dialogue without surrenders. If P has a winning strategy at the start of a logical completion d' of d, then C^d contains a justified argument for the dialogue topic.

Proof Consider any logical completion d, s of d according to a winning strategy s of P. Since O has no legal *argue* moves from $C^{d,s}$ and since no protocol rule for liberal dialogues prevents the moving of any argument, d, s is logically completed. Then the result follows from Proposition 26.

Proposition 30 Let d be a finite liberal dialogue without surrenders such that C^d is finitary. If C^d contains a justified argument for the dialogue topic and during any logical completion of d the set C^d remains constant, then P has a winning strategy at the start of a logical completion of d.

Proof Consider a finitary C^d that contains a justified argument for the dialogue topic. Then C^d contains a winning strategy S for this argument. I prove that at any point in a logical completion of d if O is the current winner then P can make himself the current winner by following S. Since C^d is finitary so that any winning strategy for P is finite, this implies that P can make himself win a logical completion.

Consider the first stage d' in a logical completion where P cannot construct a turn with S that makes him the current winner. By assumption $C^{d'} = C^d$. Since P cannot reply to an argument in W from C^d while no protocol rule restricts the moving of such arguments, d' is logically completed. But then by Proposition 26 C^d contains no justified argument for the dialogue topic, which contradicts the assumption that there is such an argument.

Proposition 33 Every relevant dialogue d contains a unique winning part.

Proof This follows from the fact that a reply to a move that is *out* is never relevant, so that p cannot have moved more than one successful reply to a move.

Proposition 35 Let d be a finite relevant dialogue without surrenders. If P has a winning strategy at the start of a logical completion d' of d, then C^d contains a justified argument for the dialogue topic.

Proof The proof cannot proceed as for liberal dialogues, since a terminated logical completion of a relevant dialogue may not be logically completed in the sense of Definition 21: this is since an implied move may be irrelevant. Instead, the proof adapts the proof of the only-if part of Proposition 26 as follows. Consider any logical completion d, s of d according to a winning strategy s of P. Consider again S as constructed from a winning part W of d, s. By Proposition 33 there exists only one such S. Let A be any defeater of a P-argument in S from $C^{d,s}$ that is not in S. Since W is P's only winning part, *argue* A is a relevant move in d, s and since no other protocol condition restricts the moving of arguments, this move is legal in d, s. But then d, s is not terminated. So there exists no such A. Then the proof can be completed as for Proposition 26.

Proposition 36 Let d be a finite relevant dialogue without surrenders such that C^d is finitary. If C^d contains a justified argument for the dialogue topic and during any logical completion of d the set C^d remains constant, then P has a winning strategy at the start of a logical completion of d.

Proof The proof is (almost) a special case of the proof of Proposition 30. Note first that only a single winning part W of d for O needs to be considered. Then any *argue* reply for P from C^d is relevant so no such reply is illegal, and the proof can be completed as before.

References

[1] L. Amgoud and C. Cayrol. A model of reasoning based on the production of acceptable argument. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002.

[2] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the Fourth International Conference on MultiAgent Systems*, pages 31–38, Boston, MA, 2000.

[3] I. Angelelli. The techniques of disputation in the history of logic. *The Journal of Philosophy*, 67:800–815, 1970.

[4] T. Ballmer and W. Brennenstuhl. *Speech Act Classification. A Study in the Lexical Classification of English Speech Activity Verbs*. Springer Verlag, Berlin, 1981.

[5] T.J.M. Bench-Capon, T. Geldard, and P.H. Leng. A method for the computational modelling of dialectical argument with dialogue games. *Artificial Intelligence and Law*, 8:233–254, 2000.

[6] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[7] G. Brewka. Reasoning about priorities in default logic. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 247–260, 1994.

[8] G. Brewka. Dynamic argument systems: a formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation*, 11:257–282, 2001.

[9] L. Carlson. *Dialogue Games: an Approach to Discourse Analysis*. Reidel Publishing Company, Dordrecht, 1983.

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

[11] H. Geffner and J. Pearl. Conditional entailment: bridging two approaches to default reasoning. *Artificial Intelligence*, 53:209–244, 1992.

[12] T.F. Gordon. The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.

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

[14] J.C. Hage, R.E. Leenes, and A.R. Lodder. Hard cases: a procedural approach. *Artificial Intelligence and Law*, 2:113–166, 1994.

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

[16] C.L. Hamblin. Mathematical models of dialogue. *Theoria*, 37:130–155, 1971.

[17] E.C.W. Krabbe. The problem of retraction in critical discussion. *Synthese*, 127:141–159, 2001.

[18] A.R. Lodder. *DiaLaw. On Legal Justification and Dialogical Models of Argumentation*. Law and Philosophy Library. Kluwer Academic Publishers, Dordrecht/Boston/London, 1999.

[19] P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1978.

[20] R.P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.

[21] J.D. Mackenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

[22] J.D. Mackenzie. Four dialogue systems. *Studia Logica*, 51:567–583, 1990.

[23] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 13:315–343, 2002.

[24] S. Parsons, M. Wooldridge, and L. Amgoud. On the outcomes of formal interagent dialogues. In *Proceedings of the Second International Conference on Autonomous Agents and Multiagent Systems (AAMAS-03)*, pages 616–623, 2003.

[25] S. Parsons, M. Wooldridge, and L. Amgoud. Properties and complexity of some formal interagent dialogues. *Journal of Logic and Computation*, 13, 2003. 347-376.

[26] J.L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

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

[28] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'2000)*, number 1919 in Springer Lecture Notes in AI, pages 224–238, Berlin, 2000. Springer Verlag.

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

[30] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.

[31] H. Prakken and G.A.W. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Günthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht/Boston/London, second edition, 2002.

[32] R. Reiter. *Knowledge in Action: Logical Foundations for Describing and Implementing Dynamical Systems*. MIT Press, Cambridge, MA, 2001.

[33] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

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