

On Dialogue Systems with Speech Acts, Arguments, and Counterarguments

Henry Prakken

Institute of Information and Computing Sciences
Utrecht University, The Netherlands
<http://www.cs.uu.nl/staff/henry.html>

Abstract. This paper proposes a formal framework for argumentative dialogue systems with the possibility of counterargument. The framework allows for claiming, challenging, retracting and conceding propositions. It also allows for exchanging arguments and counterarguments for propositions, by incorporating argument games for nonmonotonic logics. A key element of the framework is a precise definition of the notion of relevance of a move, which enables flexible yet well-behaved protocols.

1 Introduction

In recent years, dialogue systems for argumentation have received interest in several fields of artificial intelligence, such as explanation [2], AI and law [4, 6], discourse generation [5], multi-agent systems [10, 1], and intelligent tutoring [9]. These developments justify a formal study of such dialogue systems; this paper contributes to this study by an attempt to integrate two relevant developments in the fields of argumentation theory and artificial intelligence.

In argumentation theory, formal dialogue systems have been developed for so-called ‘persuasion’ or ‘critical discussion’; see e.g. [8, 14]. In persuasion, the initial situation is a conflict of opinion, and the goal is to resolve this conflict by verbal means. The dialogue systems regulate the use of speech acts for such things as making, challenging, accepting, withdrawing, and arguing for a claim. The proponent of a claim aims at making the opponent concede his claim; the opponent instead aims at making the proponent withdraw his claim. A persuasion dialogue ends when one of the players has fulfilled their aim. Logic governs the dialogue in various ways. For instance, if a participant is asked to give grounds for a claim, these grounds have to logically imply the claim. Or if a proponent’s claim is logically implied by the opponent’s concessions, the opponent is forced to accept the claim, or else withdraw some of her concessions.

Although such dialogue systems make an interesting link between the (static) logical and (dynamic) dialogical aspects of argumentation, they have one important limitation. The underlying logic is deductive, so that players cannot reply to an argument with a counterargument, since such a move presupposes a nonmonotonic, or defeasible logic. Yet in actual debates it is very common to attack one’s opponent’s arguments with a counterargument. This is where a recent development in AI becomes relevant, viz. the modelling of nonmonotonic, or defeasible

reasoning in the form of dialectical argument games; e.g. [7, 13, 11]. Such games model defeasible reasoning as a dispute between a proponent and opponent of a proposition. The proponent starts with an argument for it, after which each player must attack the other player's previous argument with a counterargument of sufficient strength. The initial proposition is provable if the proponent has a winning strategy, i.e., if he can make the opponent run out of moves in whatever way she attacks. Clearly, this dialectical setup fits well with the above-mentioned dialogue system applications. The main aim of this paper is to incorporate these argument games in protocols for persuasion dialogue. This results in a subtype of persuasion dialogues that in [11] were called 'disputes'.

The following example illustrates these observations.

Paul: My car is safer than your car. (*persuasion: making a claim*)
Olga: Why is your car safer? (*persuasion: asking grounds for a claim*)
Paul: Since it has an airbag. (*persuasion: offering grounds for a claim; dispute: stating an initial argument*)
Olga: That is true, (*persuasion: conceding a claim*) but I disagree that this makes your car safe: the newspapers recently reported on airbags expanding without cause. (*dispute: stating a counterargument*)
Paul: I also read that report (*persuasion: conceding a claim*) but a recent scientific study showed that cars with airbags are safer than cars without airbags, and scientific studies are more reliable than sporadic newspaper reports. (*dispute: rebutting a counterargument, and arguing about strength of conflicting arguments*)
Olga: OK, I admit that your argument is stronger than mine. (*persuasion: conceding a claim*) However, your car is still not safer, since its maximum speed is much higher. (*dispute: alternative counterargument*)

A second aim of this paper is to study the design of argumentative dialogue systems. Although most current systems are carefully designed, their underlying principles are often hard to see. Therefore, I shall in Section 2 propose a general framework for disputational protocols, based on intuitive principles. In Section 3 I shall instantiate it with a particular protocol (illustrated in Section 4), after which I conclude with a discussion in Section 5.

2 A Framework for Disputational Protocols

2.1 Elements and Variations

In the present framework, the initial situation of a persuasion dialogue is a conflict of opinion between two rational agents about whether a certain claim is tenable, possibly on the basis of shared background knowledge. The goal of a persuasion dialogue is to resolve this conflict by rational verbal means. The dialogue systems should be designed such that they are likely to promote this goal. Differences between the various protocols might be caused by different opinions on how this goal can be promoted, but also by, for example, different contexts

in which dialogues take place (e.g. legal, educational, or scientific dispute), or by limitations of such resources as time or reasoning capacity.

The present framework fixes the set of participants; two players are assumed, a *proponent* and an *opponent* of an initial claim. According to [14], dialogue systems regulate four aspects of dialogues:

- *Locution rules* (what moves are possible)
- *Structural rules* (when moves are legal)
- *Commitment rules* (The effects of moves on the players' commitments);
- *Termination rules* (when dialogues terminates and with what outcome).

For present purposes a fifth element must be distinguished, viz. the *underlying logic for defeasible argumentation*. On all five points the framework must allow for variations. In particular, the framework should leave room for:

- allowing one or allowing several moves per turn (*unique-move* vs. *multi-move* protocols);
- different choices on whether players can move alternatives to their earlier moves (*unique-response* vs. *multi-response* protocols);
- different underlying argument games (but all for justification);
- various sets of speech acts (but always including claims and arguments);
- different rules for legality of dialogue moves. In particular,
 - different views on inconsistent commitments
 - automatic vs forced commitment to implied commitments
- different rules for the effects of moves on the commitments of the players;
- different termination and winning criteria.

On the other hand, some conditions are hardwired in the framework. Most importantly, every move must somehow have a bearing to the main claim. This is realised by two other principles: every move must be a *reply* to some other move, being either an *attack* or a *surrender*, and every move should be *relevant*.

2.2 The Framework

The framework defines the notion of a protocol for dispute (*PPD*).

Definition 1. [*Protocols for persuasion with dispute*]. A protocol for persuasion with dispute (*PPD*) consists of the following elements. (*L*, *Players*, *Acts*, *Replies*, *Moves*, *PlayerToMove*, *Comms*, *Legal*, *Disputes*, *Winner*), as defined below.

I now define and comment on each of the elements of a protocol for dispute.

- *L* is a notion of [11], viz. a protocol for disputes based on a logic for defeasible argumentation. $wff(L)$ is the set of all well-formed formulas of *L*'s language and $Args(L)$ the set of all its well-formed arguments. For any set $T \subseteq wff(L)$, $Args_L(T) \subseteq Args(L)$ are all *L*-arguments constructible on the basis of the input information *T*. Below, *L* will often be left implicit.

Logics for defeasible argumentation (cf. [12]) formalise nonmonotonic reasoning as the construction and comparison of possibly conflicting arguments. They define the notions of an argument and of conflict between arguments, and assume or define standards for comparing arguments. The output is a classification of arguments as, for instance, ‘justified’, ‘defensible’ or ‘overruled’. One way to define argumentation logics is, as noted above, in the form of argument games. In [11] I showed how these games can be ‘dynamified’ in that the information base is not given in advance but constructed during the dispute. For present purposes this is very important, since in persuasion dialogues this typically happens.

The format of both arguments games and protocols for dispute is very similar to that of *PPD*’s. The main elements missing are the set *Act* and the functions *Replies* and *Comms*, since these formalisms have no room for speech acts.

- *Players* = $\{P, O\}$. $\overline{Player} = O$ iff $Player = P$, and P iff $Player = O$.
- *Acts* is the set of speech acts. $\{claim\ \varphi, argue(\Phi, so\ \varphi)\} \subseteq Acts$ (here, $\Phi \subseteq wff(L)$, $\varphi \in wff(L)$ and $(\Phi, so\ \varphi) \in Args(L)$). Acts have a performative and a content part. Note that each protocol has a *claim* and an *argue* act.
- *Replies* : $Acts \longrightarrow Pow(Acts)$

is a function that assigns to each act its possible replies. It is defined in terms of two other functions of the same type, *Attacks* and *Surrenders*. These functions jointly satisfy the following conditions. For any $A, B \in Acts$:

1. $B \in Replies(A)$ iff $B \in Attacks(A)$ or $B \in Surrenders(A)$;
2. $Attacks(A) \cap Surrenders(B) = \emptyset$;
3. If $B \in Surrenders(A)$, then $Replies(B) = \emptyset$;
4. If $B \in Attacks(A)$, then $Replies(B) \neq \emptyset$.

Intuitively, an attacking reply is a challenge to the replied-to act, while a surrendering reply gives up the possibility of attack. For instance, challenging a claim, responding to a challenge with an argument for the claim, and stating a counterargument are attacking replies, while retracting a proposition in reply to a challenge and conceding a proposition in reply to a claim are surrenders.

- *Moves* is the set of all well-formed moves. All moves are initial or replying moves. An *initial move* is of the form $M_1 = (Player, Act)$, and a *replying move* is of the form $M_i = (Player, Act, Move)$ ($i > 1$). $Player(M_i)$ denotes the first element of a move M_i , $Act(M_i)$ its second element and $Move(M_i)$ its third element. If $Move(M_i) = M_j$, we say that M_i is a *reply to*, or *replies to* M_j , and that M_j is the *target* of M_i .
Now the set *Moves* is recursively defined as the smallest set such that if $Player \in Players$, $Act \in Acts$ and $M_i \in Moves$, then $(Player, Act) \in Moves$ and $(Player, Act, M_i) \in Moves$.

- *PlayerToMove* determines the player to move at each stage of a dialogue. Let $Pow^*(Moves)$ be the set of all finite sequences of subsets of *Moves*. Then

$$PlayerToMove: Pow^*(Moves) \longrightarrow Players$$

such that $PlayerToMove(D) = P$ if $D = \emptyset$; else

1. $PlayerToMove(D) = P$ iff the dialogical status of M_1 is ‘out’;
2. $PlayerToMove(D) = O$ iff the dialogical status of M_1 is ‘in’.

The *PlayerToMove* function is completely defined by the framework: proponent begins a dispute and then a player keeps moving until s/he has changed the ‘dialogical status’ of the initial claim (to be defined below) his or her way. This function is hardwired in the framework since the *Legal* function of the framework requires moves to be relevant, and a move will (roughly) be defined to be relevant iff it can change the dialogical status of the initial move. Clearly, this does not leave room for other *PlayerToMove* functions than the above one.

- *Comms* 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.

$$Comms: Pow^*(Moves) \times Players \longrightarrow Pow(wff(L)).$$

$$\text{such that } Comms_\emptyset(P) = Comms_\emptyset(O).$$

Note that $Comms_\emptyset(p)$ can be nonempty (although it must have the same content for P and for O). This allows for an initially agreed or assumed basis for discussion. Note also that the framework does not require consistency of a player’s commitments. This is since some protocols allow inconsistency, after which the other player can demand retraction of one of the sources of inconsistency.

- *Legal* is a function that for any dialogue specifies the legal moves at that point, given the dialogue so far and the players’ commitments. Let C_p ($p \in Players$) stand for $Pow(wff(L)) \times p$. Then

$$Legal: Pow^*(Moves) \times C_P \times C_O \longrightarrow Pow(Moves)$$

(Below I will usually leave the commitments implicit).

This function is constrained as follows. For all $M \in Moves$ and all $D \in Pow^*(Moves)$, if $M_i \in Legal(D)$, then:

1. If $D = \emptyset$, then M_i is an initial move and $Act(M_i)$ is of the form $claim(\varphi)$;
2. $Move(M_i) \in D$;
3. $Act(M_i)$ is a reply to $Act(Move(M_i))$;
4. If M_i and M_j ($j < i$) are both replies to $M_k \in D$ and $M_j \in D$, then $Act(M_{i+1}) \neq Act(M_j)$;
5. If $Act(M_i)$ is of the form $Argue(A)$ then M_i ’s counterpart in the L -dispute L_i associated with D_i is legal in L_i ;
6. M_i is relevant in D .

Condition 1 says that a dispute always starts with a claim. Condition 2 says the obvious thing that a replied-to move must have been moved in the dialogue. Condition 3 says that an act can only be moved if it is a reply to the act moved in the replied-to move. Condition 4 states the obvious condition that if a player backtracks, the new move must be different from the first move.

The last two conditions are crucial. Condition 5 incorporates the underlying disputational protocol L , by requiring *argue* moves to conform to the legality

rules of this protocol. L_i is the proof-theoretical ‘subdispute’ of D in which the *argue* move occurs. Note that thus the framework assumes that with each sequence of *PPD*-moves an *L*-dispute can be associated. Particular protocols must specify the details.

Finally, Condition 6 every move to be relevant. Relevance, to be defined below, is the framework’s key element in allowing maximal freedom (including backtracking and postponing replies) while yet ensuring focus of a dispute.

- *Disputes* is the set of all sequences M_1, \dots, M_n of moves such that for all i :
 1. $Player(M_i) = PlayerToMove(M_1, \dots, M_{i-1})$,
 2. $M_i \in Legal(M_1, \dots, M_{i-1})$.
- *Winner* is a function that determines the winner of a dialogue, if any:

Winner: Disputes \rightarrow *Players*

The winning function is constrained by the following condition.

- If $Winner(D) = p$, then $PlayerToMove(D) = \bar{p}$ and $Legal(D) = \emptyset$;

Thus, to win it must hold that the other player has run out of moves. The rationale for this is the relevance condition (to be defined next); as long as a player can make relevant moves, s/he should not be losing. Note that termination is defined implicitly, as the situation where a player-to-move has no legal moves.

I now turn to relevance. This notion is defined in terms of the dialogical status of a move (either ‘in’ or ‘out’), which captures whether its mover has been able to ‘defend’ the move against attacks. A move can be in in two ways: the other player can have conceded it, or all attacks of the other player have been successfully replied to (where success is determined recursively). As for conceding a move, the general framework only states two necessary conditions:

- If a move M is conceded in D , then it has a surrendering reply in D .
- If M is conceded in D , it is conceded in all continuations of D .

The reason why these conditions are not sufficient lies in the most natural treatment of replies to arguing moves. In Section 3 we shall see that an arguing move has several elements (premises, conclusion, inference rule), some of which can be surrendered but others attacked at the same time. Therefore the notion of conceding a move must be fully defined in particular dialogue systems.

Definition 2. [*Dialogical status of moves*] A move M of a dialogue D is either in or out in D . It is in in D iff

1. M is conceded in D ; or else
2. all attacking moves in D that reply to it are out in D .

Now a move is relevant iff any attacking alternative would change the status of the initial move of the dialogue. This can be captured as follows.

Definition 3. [*Relevance.*] A move in a dialogue D is a relevant target iff any attacking reply to it changes the dialogical status of D ’s initial move. A move is relevant in D iff it replies to a relevant target in D .

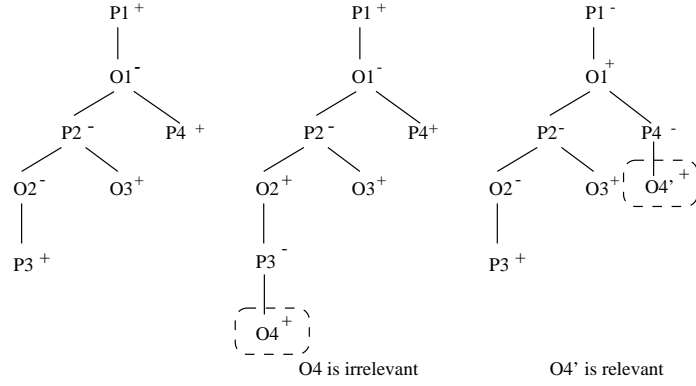


Fig. 1. Dialogical status of moves.

Note that a reply to a conceded move is never relevant.

To illustrate these definitions, consider figure 1. The first dispute tree shows the situation after P_4 . The next tree shows the dialogical status of the moves when O has continued with replying to P_3 : this move does not affect the status of P_1 , so O_4 is irrelevant. The final tree shows the situation where O has instead replied to P_4 : then the status of P_1 has changed, so O'_4 is relevant.

3 An Instantiation of the Framework

To illustrate the general framework, I now instantiate it with a specific protocol.

The underlying disputational protocol The disputational protocol L is that of liberal disputes as defined in [11], instantiated with proof-theoretical rules for sceptical argumentation. Liberal disputes allow an argument as long as it is relevant. In [11] it is shown that this protocol satisfies certain ‘soundness’ and ‘fairness’ properties with respect to the underlying argumentation logic.

Besides a set $Args$ of constructible arguments, L also assumes a binary relation of *defeat* among arguments. An argument *strictly defeats* another if the first defeats the second but not the other way around. Now Dung’s argument game says that proponent begins with an argument and then players take turns as follows: proponent’s arguments strictly defeat their targets, while opponent’s arguments defeat their targets. In addition, proponent is not allowed to repeat his moves in one ‘dialogue line’ (a dispute without backtracking moves). The precise definition of the notions of an argument, conflict and comparison of arguments are not essential, and therefore I keep these elements semiformal, using obvious logical symbols in the examples, with both material (\supset) and defeasible (\Rightarrow) implication. But the protocol assumes that arguments can be represented as a premises-conclusion pair $\Phi, \text{ so } \varphi$, where $\Phi \subseteq wff(L)$ are the premises and $\varphi \in wff(L)$ is the conclusion of the argument.

Speech acts The set of speech acts is defined as follows.

Acts	Attacks	Surrenders
<i>claim</i> φ	<i>why</i> φ	<i>concede</i> φ
<i>why</i> φ	<i>argue</i> $\Phi, so \varphi$	<i>retract</i> φ
<i>concede</i> φ		
<i>retract</i> φ		
<i>argue</i> ($\Phi, so \varphi$)	<i>why</i> φ_i ($\varphi_i \in \Phi$) <i>argue</i> ($\Phi', so \varphi'$)	<i>concede</i> φ_i ($\varphi_i \in \Phi$) <i>concede</i> (Φ implies φ)
<i>concede</i> (Φ implies φ)		

Here $\Phi, \Phi' \subseteq wff(L)$, $\varphi, \varphi' \in wff(L)$, and $(\Phi, so \varphi)$ and $(\Phi', so \varphi') \in Args(L)$.

The *claim*, *why*, *retract* and *concede* φ moves are familiar from MacKenzie-style dialogue systems. The *argue* move is present in e.g. [4] and [14]. The conceding an inference move is adapted from [4]. Its effect is to give up the possibility of counterargument. Note that an argument can be replied to by replying to one of its premises or to its inference rule, or by a counterargument.

Commitment rules The commitment rules are as follows. Let $D_i = M_1, \dots, M_i$ be any sequence of moves, and let $Player(M_i) = p$.

- If $Act(M_i) = claim \varphi$ or *concede* φ , then $Comms_{D_i}(p) = Comms_{D_{i-1}}(p) \cup \{\varphi\}$.
- If $Act(M_i) = argue(\Phi, so \varphi)$, then $Comms_{D_i}(p) = Comms_{D_{i-1}}(p) \cup \Phi \cup \{\varphi\}$.
- If $Act(M_i) = retract \varphi$ then $Comms_{D_i}(p) = Comms_{D_{i-1}}(p) / \{\varphi\}$.
- In all other cases the commitments remain unchanged.

The effects of claims, concessions and retractions are obvious. As for the effects of moving arguments, note that their conclusion is not also added to the mover's commitments. This is since some dialectical proof theories, including the present-used one, sometimes allow a player to attack himself. In [14] the material implication is also added to the commitments of the argument's mover. Although this works fine if the underlying is monotonic, in the present approach, which allows defeasible arguments, this is different.

Legality of moves The definition of the *Legal* function is completed as follows. For all $M \in Moves$ and all $D \in Pow^*(Moves)$, $M_i \in Legal(D)$ iff the above conditions and the following conditions are satisfied.

7. Each move must leave the mover's commitments classically consistent;
8. If $Act(M_i) = concede \varphi$, then
 - (a) $Comms_{D_{i-1}}(Player_i) \not\vdash \varphi$;
 - (b) $Comms_{D_{i-1}}(Player_i)$ do not justify $\neg\varphi$;
9. If $Act(M_i) = retract \varphi$, then
 - (a) $\varphi \in Comms_{D_{i-1}}(Player_i)$; and
 - (b) φ was explicitly added to $Comms_{D_{i-1}}(Player_i)$.

10. If $Act(M_i) = \textit{why } \varphi$, then $Comms_{D_{i-1}}(Player_i)$ do not justify φ .
11. If $Act(M_i) = \textit{argue}(\Phi, \textit{so } \varphi)$, then
 - (a) all preceding moves $M_j \in D$ with $Act(M_j) = \textit{why } \varphi_i$ ($\varphi_i \in \Phi$) are *out*;
 - (b) If M_i replies to an *argue* move M_j , then M_j has no child *concede*($\Phi, \textit{so } \varphi$).

As for Condition 7, note that a commitment set which supports two conflicting defeasible arguments does not have to be classically inconsistent. Whether it is, depends on the underlying logic for constructing arguments. Many logics allow the consistent expression of examples like ‘Tweety is a bird, birds generally fly, but Tweety does not fly’. This enables such moves as “I concede your argument as the general case, but in this case I have a counterargument ..”

Condition 8a says that a proposition may only be conceded if the mover is not committed to it. (This allows conceding a proposition that is defeasibly implied by the player’s own commitments.) Condition 8b forbids conceding a proposition if the opposite is justified by the player’s own commitments.

Condition 9 is obvious. Condition 10 allows retractions of ‘explicit’ commitments only. This forces a player to explicitly indicate how an implied commitment is retracted. Condition 11a forbids moving arguments of which the premises are under challenge. This is [8]’s way to avoid arguments that “beg the question”. Finally, Condition 11b says that if an argument was already conceded, no counterargument can be stated any more.

Conceding a move Next I complete the definition of conceding a move.

Definition 4 (Conceding a move). *A move M in a dialogue D has been conceded iff*

- $Act(M) \neq \textit{argue}(A)$ and M has a surrendering child; or
- $Act(M) = \textit{argue}(A)$ and both all premises and the inference rule of A have been conceded.

Associated L -disputes Next the notion of an L -dispute associated with a PPD -dispute must be defined. This notion is used in determining legality of counterarguments, but it can also serve to study logical properties of winning criteria. The idea is that during a PPD -dispute an L -dispute of arguments and counterarguments is constructed. A technical problem is that *argue* replies to *why* moves extend an argument backwards, by replacing one of its premises with an argument for this premise. To account for this, we must first define the notions of a combination of two arguments and of a modification of an argument.

Definition 5. *[Combinations of arguments.] Let $(A = S, \textit{so } \varphi)$ and $(B = S', \textit{so } \psi)$ be two arguments such that $\psi \in S$. Then $A \otimes B = (S/\{\psi\}) \cup S', \textit{so } \varphi$.*

Definition 6. *[Modification of arguments.] For any arguments A, B and C , A is a modification of A ; if B is a modification of A and $B \otimes C$ is defined, then*

$B \otimes C$ is a modification of A ; nothing else is a modification of A . We also say that A modifies A , and B modifies A in $A \otimes B$. And an argue move modifies another argue move if the argument moved by the first modifies the argument moves by the second move.

For any move $M = (p, \alpha, m)$ and arguments a and b , $M[a/b] = (p, \text{argue}(b), m)$ if $\alpha = \text{argue}(a)$; otherwise $M[a/b] = M$. Likewise for initial moves.

Now the notion of the L -part of a dispute can be defined.

Definition 7. [*L-disputes of a PPD-dispute.*] For any PPD-dispute D , the associated L -dispute $L(D)$ is a sequence of argue moves defined as follows.

1. $T(\emptyset) = \emptyset$;
2. If $\text{Act}(M_{i+1}) \neq \text{argue}(A)$ for any A , then $T(D_{i+1}) = T(D_i)$;
3. If $\text{Act}(M_{i+1}) = \text{argue}(A)$ for some A , then
 - (a) If M_{i+1} replies to an argue move M_j , then $T(D_{i+1}) = T(D_i), M'_{i+1}$, where M'_{i+1} is M_{i+1} except that it replies to the move in $T(D_{i+1})$ modified by M_j ;
 - (b) If $M_{i+1} = (p, \alpha, m)$ replies to a why move replying to a claim, then $T(D_{i+1}) = T(D_i), (p, \alpha)$;
 - (c) If M_{i+1} replies to a why φ move replying to an $\text{argue}(B)$ move M_j , then
 - i. If T_i contains any argue moves M_k resulting from modifications of M_j such that their arguments C still have a premise φ , then $T(D_{i+1}) = T^*(D_i)$, where $T^*(D_i)$ is obtained from $T(D_i)$ by replacing C in all such M_k with $C \otimes A$, and then adjusting the targets of moves when these targets have been changed.
 - ii. Else $T(D_{i+1}) = T(D_i), M'_j$, where M'_j is obtained from M_j by replacing B with $B \otimes A$.

So the construction of an L -dispute starts with the empty set, and each PPD-move other than an *argue* move leaves its content unchanged. As for *argue* PPD-moves, two cases must be distinguished, whether it replies to another *argue* move or to a *why* move. In the first case the argue move can simply be added to the L -dispute, but the second case is more complex. Again two cases must be considered. If the replied-to *why* move itself replied to the initial claim, then the *argue* is the root of a new dialectical tree, so the move to which it replies must be omitted, to turn it into an initial move. Finally, if the replied-to *why* φ move challenged the premise of an argument B , then again two cases must be considered. If the L -dispute contains modifications of A that still contain premise φ , then these modifications (if not equal to B itself) were triggered by a *why* attack on another premise of B . In that case φ must in all these modifications be replaced with the premises of A . Note that if no such other *why* attacks were made, this boils down to modifying B itself. (Note also that if in T , M_i replies to M_j , and M_j is then modified by M_k , from then on M_i replies to the modified move.) If, however, no modification of B in the L -dispute contains a premise φ , then moving A was an alternative to an earlier reply to the *why* φ move. Then we must add to the L -dispute an alternative modification of the original *argue*(B) move, with $B \otimes A$ (note that the original move was not in T_i any more).

In general an L -dispute is a collection of trees, since a *why* reply to the initial move can be answered with alternative arguments. So the condition of the general framework that an *argue* move M is legal in the associated L -dispute means that M is legal in the tree contained in this dispute that itself contains $Move(M)$.

Winning As for winning, several definitions are conceivable. Part of the aim of the present framework is to provide a setting in which the alternatives can be compared. In the present protocol I simply turn the necessary conditions of the general framework into a necessary-and-sufficient condition.

Definition 8. For any dispute D , $Winner(D) = p$ iff $PlayerToMove(D) = \bar{p}$ and $Legal(D) = \emptyset$.

It is immediate that if p wins D , then M_1 in D is labelled p 's way. However, the same does not always hold for the associated L -dispute. Consider the following dispute D (In the examples below I leave the replied-to move implicit if it is the preceding move, and P_i and O_i stand for turns of a player.)

P_1 : <i>claim</i> p	O_1 : <i>why</i> p
P_2 : <i>argue</i> ($q, q \Rightarrow p$, <i>so</i> p)	O_2 : <i>argue</i> ($q, r, q \wedge r \Rightarrow \neg p$, <i>so</i> $\neg p$)
P_3 : <i>why</i> r	O_3 : <i>concede</i> p (to P_1)

Now P has won, but $T(D) = P_2, O_2$, in which P_2 is *out* and O_2 is *in*. So a player can lose by unforced surrenders.

It also holds that if O has won, P is not committed to his main claim any more. This is since if all other moves have become illegal for P , he can still surrender to O 's initial *why* attack. However, it does not hold that if P has won, O is always committed to P 's main claim φ . This is since O might have moved an argument with premise $\neg\varphi$ and in the course of the dispute retracting φ may have become irrelevant and thus illegal, so that conceding M_1 has also become illegal. Future research should reveal whether this is a problematic property of the protocol.

4 Examples of Dialogues

Example 1. Most argumentation logics do not allow counterarguments to deductively valid arguments. If such a logic underlies our protocol, then conceding the premises of such an argument can cause a loss. Consider

P_1 : <i>claim</i> p	O_1 : <i>why</i> p
P_2 : <i>argue</i> ($q, q \supset p$, <i>so</i> p)	O_2 : <i>concede</i> q , <i>concede</i> $q \supset p$

Now O is still to move, and her only legal moves are *concede*($\{q, q \supset p\}$ *implies* q) and *concede* p , after which moves P_1 is still *in* so O cannot move.

Example 2. The next example (on the Nixon diamond) shows that a player can lose with a poor move even if the player's own commitments support a valid counterargument. Suppose $Comms_0(p) = \{Qx \Rightarrow Px, Qn\}$ and consider

$$\begin{array}{ll} P_1: \textit{claim } \neg Pn & O_1: \textit{why } \neg Pn \\ P_2: \textit{argue}(Rn, Rx \Rightarrow \neg Px, \textit{so } \neg Pn) & O_2: \textit{concede } Rn, \textit{concede } Rx \Rightarrow \neg Px, \\ & \textit{concede } \neg Pn \textit{ (to } P_1\text{)}. \end{array}$$

Now P wins while O could instead of conceding P_1 have attacked it with $\textit{argue}(Qn, Qx \Rightarrow Px, \textit{so } Pn)$. Note also that if O had not conceded P_2 's premises, then conceding $\neg Pn$ would have violated condition 8b on move legality.

Example 3. The next dispute shows that a player can sometimes use the other player's commitments against that player (the commitments are shown each time when they have changed).

Move	$Comms_D(P)$	$Comms_D(O)$
	$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p\}$	$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p\}$
$P_1: \textit{claim } p$	$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p, p\}$	
$O_1: \textit{why } p$		
$P_2: \textit{argue}(r, s, r \wedge s \Rightarrow p, \textit{so } p)$	$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p, p, r, s, r \wedge s \Rightarrow p\}$	
$O_2: \textit{concede } r, \textit{argue}(q, q \Rightarrow t, t \Rightarrow \neg s, \textit{so } \neg s)$		$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p, q, r, q \Rightarrow t, t \Rightarrow \neg s, \neg s\}$

At this point, O 's commitments justify p , since they contain an implicit argument for p . Suppose P next moves this argument. Then O can in turn use a counterargument supported by P 's commitments.

$P_3: \textit{argue}(r, q, q \Rightarrow t, r \wedge t \Rightarrow p, \textit{so } p)$	$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p, p, r, s, r \wedge s \Rightarrow p, r, t, q \Rightarrow t\}$	
$O_3: \textit{argue}(s, s \Rightarrow \neg q, \textit{so } \neg q)$		$\{s \Rightarrow \neg q, r \wedge t \Rightarrow p, q, r, q \Rightarrow t, t \Rightarrow \neg s, \neg s\}$

And the dispute continues.

Example 4. Next I illustrate the construction of an L -dispute. I first list a PPD -dispute and then the construction of the associated L -dispute.

$$\begin{array}{ll} P_1: \textit{claim } p & O_1: \textit{why } p \\ P_2: \textit{argue}(q, q \Rightarrow p, \textit{so } p) & O_2: \textit{argue}(r, r \Rightarrow \neg p, \textit{so } \neg p) \\ P_3: \textit{argue}(s, t, s \wedge t \Rightarrow p, \textit{so } p) & (P_3 \textit{ jumps back to } O_1) \\ & O_3: \textit{why } s \\ P_4: \textit{argue}(u, u \Rightarrow s, \textit{so } s) & O_4: \textit{argue}(v, v \Rightarrow \neg u, \textit{so } \neg u) \end{array}$$

$P_5: \text{argue}(x, x \Rightarrow s, \text{so } s)$ (P_5 jumps back to O_3)
 $O_5: \text{argue}(y, y \Rightarrow \neg x, \text{so } \neg x)$
 $P_6: \text{argue}(z, z \Rightarrow \neg r, \text{so } \neg r)$ (P_6 jumps back to O_2)
 $(O_6 \text{ jumps back to } P_3)$ $O_6: \text{why } t$
 $P_7: \text{argue}(k, k \Rightarrow t, \text{so } t)$

The associated L -dispute is constructed as follows. The first two arguments are added with P_2 and O_2 , so (denoting disputes with their last move and listing the replied-to moves between square brackets):

$$\begin{aligned}
T(P_2) &= P_2 \\
T(O_2) &= P_2, O_2[P_2]
\end{aligned}$$

So far, T contains just one dialectical tree. A second tree is created by P_3 , which is an alternative *argue* reply to O 's *why* attack on P 's main claim. Hence

$$T(P_3) = P_2, O_2[P_2], P_3$$

With P_4 the first modification of an argument in T takes place. P_3 's argument is combined with P_4 's argument for s (displayed with overloaded \otimes).

$$T(P_4) = P_2, O_2[P_2], P_3 \otimes P_4$$

O_4 simply adds a new argument, which replies to P_3 as modified by P_4 .

$$T(O_4) = P_2, O_2[P_2], P_3 \otimes P_4, O_4[P_3 \otimes P_4]$$

P_5 splits the second tree in T into two alternative trees, by giving an alternative backwards extension of its root. Then O_5 simply extends the newly created tree, after which P_6 extends the first tree in T .

$$\begin{aligned}
T(P_5) &= P_2, O_2[P_2], P_3 \otimes P_4, O_4[P_3 \otimes P_4], P_3 \otimes P_5 \\
T(O_5) &= P_2, O_2[P_2], P_3 \otimes P_4, O_4[P_3 \otimes P_4], P_3 \otimes P_5, O_5[P_3 \otimes P_5] \\
T(P_6) &= P_2, O_2[P_2], P_6[O_2], P_3 \otimes P_4, O_4[P_3 \otimes P_4], P_3 \otimes P_5, O_5[P_3 \otimes P_5]
\end{aligned}$$

Finally, P_7 illustrates an interesting phenomenon. It replaces the second premise of O_3 with an argument; however, O_3 was already modified twice in two alternative ways with respect to its first premise, so P_7 actually modifies both of these modifications of O_3 . This results in the following final L -dispute. (Note also that the targets of O_4 and O_5 have been replaced with their extended versions.)

$$\begin{aligned}
P_2: & \quad q, q \Rightarrow p, \text{so } p \\
O_2: & \quad r, r \Rightarrow \neg p, \text{so } \neg p [P_2] \\
P_6: & \quad z, z \Rightarrow \neg r, \text{so } \neg r [O_2] \\
P_3 \otimes P_4 \otimes P_7: & \quad u, u \Rightarrow s, k, k \Rightarrow t, s \wedge t \Rightarrow p, \text{so } p \\
P_3 \otimes P_5 \otimes P_7: & \quad x, x \Rightarrow s, k, k \Rightarrow t, s \wedge t \Rightarrow p, \text{so } p \\
O_4: & \quad v, v \Rightarrow \neg u, \text{so } \neg u [P_3 \otimes P_4 \otimes O_7] \\
O_5: & \quad y, y \Rightarrow \neg x, \text{so } \neg x [P_3 \otimes P_5 \otimes O_7]
\end{aligned}$$

5 Discussion

Alternative Instantiations To discuss some alternative instantiations of the framework, note first that alternative definitions of winning may be possible, for instance, in terms of what is implied by the players' commitments. Secondly, as for maintaining consistency of one's commitments, some protocols allow inconsistency but give the other party the option to demand resolution of the conflict; a similar *resolve* move is possible if a commitment is explicitly retracted but still implied by the remaining commitments [8, 14]. Thus the burden of proving inconsistency or implicit commitment is placed upon the other party. Finally, as for replies to *why* moves, the obligation to reply to it with an argument for the challenged claim could be made dependent on questions of the burden of proof.

Features and Restrictions of the Framework The framework of this paper is flexible in some respects but restricted in some other respects. It is flexible, firstly, since it allows for different sets of speech acts, and different commitment rules, underlying logics and winning criteria. It is also 'structurally' flexible, in that it allows for backtracking, including jumping to earlier branches, and for postponing replies to move (even indefinitely if the move has become irrelevant). This flexibility is induced by the notion of relevance.

However, the framework also has some restrictions. For instance, the condition of relevance prevents the moving in one turn of alternative ways to change the status of the main claim. Further, the requirement that each move replies to a preceding move excludes some useful moves, 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 want to reveal what they are aiming at. The same requirement also excludes invitations to retract or concede [8, 14]. Finally, the framework only allows two-player disputes, leaving no room for, for example, arbiters or judges.

Related Research There have been some earlier proposals to combine formal dialogue systems with argumentation logics. Important early work was done by Loui [7], although he focussed less on speech act aspects. A major source of inspiration for the present research was Tom Gordon's model of civil pleading in anglo-american law [4] (cf. also [6]). Gordon presents a particular protocol rather than a framework. The same holds for a recent proposal in the context of multi-agent negotiation systems [1]. Finally, [3] shows how protocols for multi-party disputes can be formalised in situation calculus. Brewka focuses less on dialectical and relevance aspects but more on describing the 'current state' of a dispute and how it changes. His approach paves the way for, for instance, formal verification of consistency of protocols.

Conclusion This paper has presented a formal framework for persuasion dialogues with counterargument, and has given one detailed instantiation. Unlike

earlier work, the framework is based on some general design principles, notably the distinction of attacking and surrendering replies to a move, and the notions of dialogical status and relevance of moves. The framework's instantiation also provided a still generic notion of an argument-counterargument dispute associated with a persuasion dialogue; I expect that this notion will provide a basis for investigating logical properties of the protocol, especially of its winning conditions.

Being a first attempt to provide a general framework, the focus of this paper has been more on definition than on technical exploration. Much work needs to be done on investigating its properties. In fact, one aim of this paper was to make this further work possible.

References

- [1] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the Fourth International Conference on MultiAgent Systems*, Boston, MA, 2000.
- [2] T.J.M. Bench-Capon, D. Lowes, and A.M. McEnergy. Using Toulmin's argument schema to explain logic programs. *Knowledge Based Systems*, pages 177–183, 1991.
- [3] G. Brewka. Dynamic argument systems: a formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation*, 2000. To appear.
- [4] T.F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.
- [5] F. Grasso, A. Cawsey, and R. Jones. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies*, 2000. To appear.
- [6] J.C. Hage, R.E. Leenes, and A.R. Lodder. Hard cases: a procedural approach. *Artificial Intelligence and Law*, 2:113–166, 1994.
- [7] R.P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.
- [8] J.D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.
- [9] N. Maudet and D. Moore. Dialogue games for computer-supported collaborative argumentation. In *Proceedings of the Workshop on Computer-Supported Collaborative Argumentation for Learning Communities*, Stanford, 1999.
- [10] S. Parsons, C. Sierra, and N.R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8:261–292, 1998.
- [11] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 2000. To appear in special issue on New Perspectives in Dialogical Logics.
- [12] H. Prakken and G.A.W. Vreeswijk. Logical systems for defeasible argumentation. In D. Gabbay, editor, *Handbook of Philosophical Logic*. Kluwer Academic Publishers, Dordrecht/Boston/London, 2000. Second edition, to appear.
- [13] G.A.W. Vreeswijk. The computational value of debate in defeasible reasoning. *Argumentation*, 9:305–341, 1995.
- [14] 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.