*The Knowledge Engineering Review*, Vol. 21:2, 163–188. © 2006, Cambridge University Press
doi:10.1017/S0269888906000865 Printed in the United Kingdom

# Formal systems for persuasion dialogue

HENRY PRAKKEN

*Department of Information and Computing Sciences, Universiteit Utrecht,*
*Centre for Law & ICT, Faculty of Law, University of Groningen, The Netherlands;*
*e-mail: henry@cs.uu.nl*

**Abstract**

This article reviews formal systems that regulate persuasion dialogues. In such dialogues two or more participants aim to resolve a difference of opinion, each trying to persuade the other participants to adopt their point of view. Systems for persuasion dialogue have found application in various fields of computer science, such as non-monotonic logic, artificial intelligence and law, multi-agent systems, intelligent tutoring and computer-supported collaborative argumentation. Taking a game-theoretic view on dialogue systems, this review proposes a formal specification of the main elements of dialogue systems for persuasion and then uses it to critically review some of the main formal systems for persuasion. The focus of this review will be on regulating the interaction between agents rather than on the design and behaviour of individual agents within a dialogue.

## 1 Introduction

This article reviews formal dialogue systems for persuasion. In persuasion dialogues two or more participants try to resolve a difference of opinion by arguing about the tenability of one or more claims or arguments, each trying to persuade the other participants to adopt their point of view. Dialogue systems for persuasion regulate what utterances the participants can make and under which conditions they can make them, what the effects of their utterances are on their propositional commitments, when a dialogue terminates and what the outcome of a dialogue is. Good dialogue systems regulate all this in such a way that conflicts of view can be resolved in a way that is both fair and effective (Loui, 1998).

Systems for persuasion dialogues were already studied in medieval times (Angelelli, 1970). The modern study of formal dialogue systems for persuasion probably started with two publications by Hamblin (1970, 1971). Initially, the topic was studied only within philosophical logic and argumentation theory; see, for example, Mackenzie (1979, 1990), Woods & Walton (1978) and Walton & Krabbe (1995). From the early nineties the study of persuasion dialogues was taken up in several fields of computer science. In Artificial Intelligence logical models of commonsense reasoning have been extended with formal models of persuasion dialogue as a way to deal with resource-bounded reasoning (Loui, 1998; Brewka, 2001). Persuasion dialogues have also been used in the design of intelligent tutoring systems (Moore, 1993; Yuan, 2004) and have been proposed as an element of computer-supported collaborative argumentation (Maudet & Moore, 1999). In artificial intelligence and law, interest in dialogue systems arose when researchers realized that legal reasoning is bound not only by the rules of logic and rational inference but also by those of fair and effective procedure. See, for example, Gordon (1994), Hage *et al.* (1994), Bench-Capon (1998), Lodder (1999) and Prakken (2001a) and see Alexy (1978) for a philosophical underpinning of this view. Persuasion was seen by these researchers as an appropriate model of legal procedures (cf. Walton, 2003). Finally, in the field of multi-agent systems (MAS) dialogue systems have been incorporated into models of rational agent interaction, for instance, by Kraus *et al.* (1998), Parsons

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

*et al.* (1998), Amgoud *et al.* (2000), McBurney & Parsons (2002) and Parsons *et al.* (2003a,b). To fulfil their own or joint goals, intelligent agents often need to interact with other agents. When they pursue joint goals, the typical modes of interaction are *information seeking* and *deliberation* and when they self-interestedly pursue their own goals, they often interact by way of *negotiation.* In all of these cases the dialogue can shift to *persuasion.* For example, in information-seeking, a conflict of opinion could arise on the credibility of a source of information, in deliberation the participants may disagree about likely effects of plans or actions and in negotiation they may disagree about the reasons why a proposal is in one's interest; also, in all three cases the participants may disagree about relevant factual matters.

The term ''persuasion dialogue'' was introduced into argumentation theory by Walton (1984) as part of his influential classification of dialogues into six types according to their goal (see also e.g. Walton & Krabbe 1995). While *persuasion* aims to resolve a difference of opinion, *negotiation* tries to resolve a conflict of interest by reaching a deal, *information seeking* aims at transferring information, *deliberation* wants to reach a decision on a course of action, *inquiry* is aimed at ''growth of knowledge and agreement'' and *quarrel* is the verbal substitute of a fight. This classification is not meant to be exhaustive and leaves room for dialogues of mixed type.

To delineate the precise scope of this review, it is useful to discuss what is the subject matter of dialogue systems. According to Carlson (1983) dialogue systems define the principles of coherent dialogue. In his words, whereas logic defines the conditions under which a proposition is true, dialogue systems define the conditions under which an utterance is appropriate. And the leading principle here is that an utterance is appropriate if it furthers the goal of the dialogue in which it is made. So, for instance, an utterance in a persuasion should contribute to the resolution of the conflict of opinion that triggered the persuasion, and an utterance in a negotiation should contribute to reaching agreement on a reallocation of scarce resources. 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. Carlson therefore proposes a game-theoretic approach to dialogues, in which speech acts are viewed as moves in a game and rules for their appropriateness are formulated as rules of the game. Virtually all work on formal dialogue systems for persuasion follows this approach and therefore the discussion in this review will assume a game format of dialogue systems. It should be noted that the term *dialogue system* as used in this review only covers the rules of the game, i.e. which moves are allowed; it does not cover principles for playing the game well, i.e. strategies and heuristics for the individual players. Of course, the latter are also important in the study of dialogue, but they will be treated as being external to dialogue systems and instead of aspects of models of dialogue participants.

This review is organized as follows. First in Section 2 an example persuasion dialogue will be presented, to give a feel for what persuasion dialogues are and to provide material for illustration and comparison in the subsequent discussions. Then in Section 3 a formal specification is proposed of common elements of dialogue game systems, which in Section 4 is made more specific for persuasion. The objective of this specification is to provide a precise and unambiguous basis for the discussion of persuasion systems in the remainder of this review. This discussion starts in Section 5 with the most obvious instantiations of the elements of Section 4 and continues with defining some distinctions related to persuasion (Section 6), discussing some features of and issues in the design of persuasion systems (Section 7) and reviewing some properties of persuasion systems that can be formally investigated (Section 8). Then in Section 9 a number of persuasion systems proposed in the literature are critically reviewed and in Section 10 the review is concluded with some main topics for further research.

## 2 An example persuasion dialogue

The following example persuasion dialogue, adapted from Prakken (2005), exhibits some typical features of persuasion and will be used in this review to illustrate different degrees of expressiveness and strictness of the various persuasion systems.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

`Paul:` My car is 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 this does not make your car safe (*stating a counterclaim*).

`Paul:` Why does that not make my care safe? (*asking grounds for a claim*).

`Olga:` Since the newspapers recently reported on airbags expanding without cause (*stating a counterargument by providing grounds for the counterclaim*).

`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 still not safe, since its maximum speed is very high (*alternative counterargument*).

`Paul:` OK, I was wrong that my car is safe.

This dialogue illustrates several features of persuasion dialogues.

(a) Participants in a persuasion dialogue not only exchange arguments and counterarguments but also express various propositional attitudes, such as claiming, challenging, conceding or retracting a proposition.

(b) As for arguments and counterarguments it illustrates the following features.
   (i) An argument is sometimes attacked by constructing an argument for the opposite conclusion (as in Olga's two counterarguments) but sometimes by saying that in the given circumstances the premises of the argument do not support its conclusion (as in Paul's counterargument). This is Pollock's (1995) well-known distinction between rebutting and undercutting counterarguments.
   (ii) Counterarguments are sometimes stated at once (as in Paul's undercutter and Olga's last move) and are sometimes introduced by making a counterclaim (as in Olga's second and third move).
   (iii) Natural-language arguments sometimes leave elements implicit. For example, Paul's second move arguably leaves a commonsense generalization ''Cars with airbags usually are safe'' implicit.

(c) As for the structure of dialogues, the example illustrates the following features.
   (i) The participants 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 her first, to leave Paul with two attacks to counter.
   (ii) The participants 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.

## 3 Elements of dialogue systems

In this section a formal specification is proposed of the common elements of dialogue systems. To summarize, dialogue systems have a *dialogue goal* and at least two *participants*, who can have various *roles*. Dialogue systems have two languages, a *topic language* and a *communication language*. Sometimes, dialogues take place in a *context* of fixed and undisputable knowledge. Typical examples of contexts are the relevant laws in a legal dispute or a system description in a dialogue about a diagnostic problem. The heart of a dialogue system is formed by a *protocol*, specifying the allowed moves at each point in a dialogue, the *effect rules*, specifying the effects of utterances on the participants' commitments, and the *outcome rules*, defining the outcome of a

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

dialogue. Two kinds of protocol rules are sometimes separately defined, viz. *turntaking* and *termination* rules.

Let us now specify these elements more formally. In the rest of this review this specification will be used when describing systems from the literature; in consequence, their appearance in this text may differ from their original presentation. The definitions below of dialogues, protocols and strategies are based on Chapter 12 of Barwise & Moss (1996) as adapted in Prakken (2005). Among other sources of inspiration are Maudet & Evrard (1998) and Parsons & McBurney (2003), and similar definitions of some elements can be found in Dunne & McBurney (2003). As for notation, the complement $\overline{\varphi}$ of a formula $\varphi$ is $\neg\varphi$ if $\varphi$ is a positive formula and $\psi$ if $\varphi$ is a negative formula $\neg\psi$.

- A *topic language* $\mathscr{L}_t$, closed under classical negation.
- A *communication language* $\mathscr{L}_c$.
- The set of *dialogues*, denoted by $M^{\leq\infty}$, is the set of all sequences from $\mathscr{L}_c$, and the set of *finite dialogues*, denoted by $M^{<\infty}$, is the set of all finite sequences from $\mathscr{L}_c$. For any dialogue $d=m_1, \ldots, m_n, \ldots$, the subsequence $m_1, \ldots, m_i$ is denoted with $d_i$.
- A *dialogue purpose.*
- A set $\mathscr{A}$ of *participants*, and a set $\mathscr{R}$ of *roles*, defined as disjoint subsets of $\mathscr{A}$. A participant $a$ may or may not have a, possibly inconsistent, *belief base* $\Sigma_a \subseteq Pow(\mathscr{L}_t)$, which may or may not change during a dialogue. Furthermore, each participant has a, possibly empty set of *commitments* $C_a \subseteq \mathscr{L}_t$, which usually changes during a dialogue.
- A *context* $K \subseteq \mathscr{L}_t$, containing the knowledge that is presupposed and must be respected during a dialogue. The context is assumed consistent and remains the same throughout a dialogue.
- A *logic* $L$ for $\mathscr{L}_t$, which may or may not be monotonic and which may or may not be argument-based.
- A set of *effect rules* $E$ for $\mathscr{L}_c$, specifying for each utterance $\varphi \in \mathscr{L}_c$ its effects on the commitments of the participants. These rules are specified as functions:

  $$C_a : M^{<\infty} \rightarrow Pow(\mathscr{L}_t).$$

  Changes in commitments are completely determined by the last move in a dialogue and the commitments just before making that move:

  $$\text{If } d=d \text{ then } C_a(d, m)=\mathrm{C_a}(d',m).$$

- A *protocol* $P$ for $\mathscr{L}_c$, specifying the legal moves at each stage of a dialogue. Formally, a *protocol* on $\mathscr{L}_c$ is a function $P$ with as domain the context plus a non-empty subset $D$ of $M^{<\infty}$ taking subsets of $\mathscr{L}_c$ as values. That is:

  $$P : Pow(\mathscr{L}_t) \times D \rightarrow Pow(\mathscr{L}_c).$$

  such that $D \subseteq M^{<\infty}$. The elements of $D$ are called the *legal finite dialogues.* The elements of $P(d)$ are called the moves allowed after $d$. If $d$ is a legal dialogue and $P(d)=\emptyset$, then $d$ is said to be a *terminated* dialogue. $P$ must satisfy the following condition: for all finite dialogues $d$ and moves $m$, $d \in D$ and $m \in P(d)$ if and only if $d, m \in D$.

  It is useful (although not strictly necessary) to explicitly distinguish elements of a protocol that regulate turntaking and termination.

  > A *turntaking* function is a function $T : D \times Pow(\mathscr{L}_t) \rightarrow Pow(\mathscr{A})$. A *turn* of a dialogue is defined as a maximal sequence of moves in the dialogue in which the same player is to move. Note that $T$ can designate more than one player as to-move next.
  >
  > *Termination* is defined above as the case where no move is legal. Accordingly, an explicit definition of termination should specify the conditions under which $P$ returns the empty set.

- *Outcome rules* $O$, defining the outcome of a dialogue. For instance, in negotiation the outcome is an allocation of resources, in deliberation it is a decision on a course of action, and in persuasion dialogue it is a winner and a loser of the persuasion dialogue.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

Note that no relations are assumed between a participant's commitments and belief base. Commitments are an agent's publicly declared points of view about a proposition, which may or may not agree or coincide with the agent's internal beliefs. For instance, an accused in a criminal trial may very well publicly defend his innocence while he knows he is guilty.

Finally, as explained in the Introduction, participants in a dialogue can have strategies and heuristics for playing the dialogue, given their individual dialogue goal. The notion of a *strategy* for a participant $a$ can be defined in the game-theoretical sense, as a function from the set of all finite legal dialogues in which $a$ is to move into $\mathscr{L}_c$. A strategy for $a$ is a *winning strategy* if in every dialogue played in accord with the strategy $a$ realises his dialogue goal (for instance, winning in persuasion). *Heuristics* generalize strategies: a heuristic for $a$ is a function from a subset of the set of all finite legal dialogues in which $a$ is to move into $Pow(\mathscr{L}_c)$.

More formally, let $D_a$, a subset of $D$, be the set of all dialogues where $a$ is to move, and let $D'_a$ be a subset of $D_a$. Then a strategy and a heuristic for $a$ are defined as functions $s_a$ and $h_a$ as follows:

$s_a : D_a \rightarrow \mathscr{L}_c$;
$h_a : D'_a \rightarrow Pow(\mathscr{L}_c)$.

## 4 Persuasion

Let us now become more precise about persuasion. In Walton & Krabbe (1995) persuasion dialogues are defined as dialogues where the goal of the dialogue is to resolve a conflict of points of view between at least two participants by verbal means. A *point of view* with respect to a proposition can be positive (for), negative (against) or merely one of critical doubt. The participant's individual aim is to persuade the other participant(s) to take over its point of view. According to Walton & Krabbe a conflict of points of view is resolved if all parties share the same point of view on the proposition that is the topic of the conflict.

Walton & Krabbe distinguish *disputes* as a subtype of persuasion dialogues where two parties disagree about a single proposition $\varphi$, such that at the start of the dialogue one party has a positive ($\varphi$) and the other party a negative ($\neg\varphi$) point of view towards the proposition.Walton & Krabbe then extend this notion to *conflicts of contrary opinions*, where the participants have a positive point of view on, respectively, $\varphi$ and $\psi$ such that $\models \neg(\varphi \wedge \psi)$.

Some of the elements of dialogue systems need to be instantiated for persuasion dialogues, viz. the dialogue goal, the participant roles, and the dialogue outcomes.

- The *dialogue purpose* is resolution of a conflict of opinion about one or more propositions, called the *topics* $T \subseteq \mathscr{L}_t$. This dialogue purpose gives rise to the following participant roles and outcome rules.
- The participants can have the following *roles*. $prop(t) \subseteq \mathscr{A}$, the *proponents* of topic $t$, is the (nonempty) set of all participants with a positive point of view towards $t$. Likewise, $opp(t) \subseteq \mathscr{A}$, the *opponents* of $t$, is the (non-empty) set of all participants with a doubtful point of view toward a topic $t$. Together, the proponents and opponents of $t$ are called the *adversaries* with respect to $t$. For any $t$, the sets $prop(t)$ and $opp(t)$ are disjoint but do not necessarily jointly exhaust $\mathscr{A}$. The remaining participants, if any, are the *third parties* with respect to $t$, assumed to be neutral towards $t$.

  Note that this allows that a participant is a proponent of both $t$ and $\neg t$ or has a positive attitude towards $t$ and a doubtful attitude towards a topic $t'$ that is logically equivalent to $t$. Since protocols can deal with such situations in various ways, I choose not to exclude them by definition.
- The *Outcome rules* of systems for persuasion dialogues define for each dialogue $d$, context $K$ and topic $t$ the winners and losers of $d$ with respect to topic $d$. More precisely, $O$ consists of two functions $w$ and $l$:

  $w : D \times Pow(\mathscr{L}_t) \times \mathscr{L}_t \rightarrow Pow(\mathscr{A})$;
  $l : D \times Pow(\mathscr{L}_t) \times \mathscr{L}_t \rightarrow Pow(\mathscr{A})$.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

These functions will be written as $w_t^K(d)$ and $l_t^K(d)$ or, if there is no danger for confusion, as $w_t(d)$ and $l_t(d)$. They are defined for each dialogue $d$ but only for those $t$ that are a topic of $d$. They further satisfy the following conditions for arbitrary but fixed context $K$:

$w_t(d) \cap l_t(d) = \emptyset$;
$w_t(d) = \emptyset$ if and only if $l_t(d) = \emptyset$;
if $|\mathcal{A}| = 2$, then $w_t(d)$ and $l_t(d)$ are at most singletons.

Strictly speaking, the win and lose functions are needed only for defining strategies for the individual participants, since participants want to win.

## 5 Instantiations: obvious choices

Some instantiations of the elements from Sections 3 and 4 are quite obvious. They are discussed below and are assumed to hold throughout the remainder of this review.

First, to make sense of the notions of proponent and opponent, their commitments at the start of a dialogue should not conflict with their points of view:

- If $a \in prop(t)$ then $\bar{t} \notin C_a(\emptyset)$;
- If $a \in opp(t)$ then $t \notin C_a(\emptyset)$.

Furthermore, in persuasion at most one side in a dialogue gives up, i.e.

- $w_t(d) \subseteq prop(t)$ or $w_t(d) \subseteq opp(t)$ ; and
- If $a \in w_t(d)$ then

  if $a \in prop(t)$ then $t \in C_a(d)$;
  if $a \in opp(t)$ then $t \notin C_a(d)$.

These conditions ensure that a winner did not change its point of view. Note that the only-ifs of the two latter winning conditions do not hold in general. This will be explained further below when the distinction between pure persuasion and conflict resolution is made. Note also that these conditions mean that two-person persuasion dialogues are zero-sum games. Perhaps this is the main feature that sets persuasion apart from information seeking, deliberation and inquiry.

Next the most common speech acts that can be found in the literature are listed, with their informal meaning and the various terms with which they have been denoted in the literature. To make this survey more uniform, the present terminology will be used even if the original publication of a system uses different terms.

- *claim* $\varphi$ (assert, statement, . . .). The speaker asserts that $\varphi$ is the case.
- *why* $\varphi$ (challenge, deny, question, . . .). The speaker challenges that $\varphi$ is the case and asks for reasons why it would be the case.
- *concede* $\varphi$ (accept, admit, . . .). The speaker admits that $\varphi$ is the case.
- *retract* $\varphi$ (withdraw, no commitment, . . .). The speaker declares that he is not committed (any more) to $\varphi$. Retractions are ''really'' retractions if the speaker is committed to the retracted proposition, otherwise it is a mere declaration of non-commitment (e.g. in reply to a question).
- $\varphi$ *since* $S$ (argue, argument, . . .). The speaker provides reasons why $\varphi$ is the case. Some protocols do not have this move but require instead that reasons be provided by a *claim* $\varphi$ or *claim* $S$ move in reply to a *why* $\psi$ move (where $S$ is a set of propositions). Also, in some systems the reasons provided for $\varphi$ can have structure, for example, of a proof three or a deduction.
- *question* $\varphi$ (. . .). The speaker asks another participant's opinion on whether $\varphi$ is the case.

*Paul and Olga (continued)*: In this communication language our example from Section 2 can be more formally displayed as follows:

$P_1$: *claim* safe;
$O_2$: *why* safe;

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

$P_3$: safe *since* airbag;
$O_4$: *concede* airbag;
$O_5$: *claim* ¬ safe;
$P_6$: *why* ¬ safe;
$O_7$: ¬ safe *since* newspaper: ''explode'';
$P_8$: *concede* newspaper: ''explode'';
$P_9$: so what *since* ¬ newspapers reliable;
$O_{10}$: ¬ safe *since* high max. speed;
$P_{11}$: *retract* safe.

As for the commitment rules, the following ones 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 $s(m)=\varphi$ *since* $S$ then $C_s(d,m)\supseteq C_s(d)\cup prem(A)$.

The rule for *since* uses $\supseteq$ for two reasons. In some systems, such as Prakken (2000, 2005), the move also commits to $\varphi$, since arguments can also be moved as counterarguments instead of as replies to challenges of a claim. In other systems, such as Walton & Krabbe (1995), the move also commits the speaker to the material implication $S\rightarrow\varphi$.

*Paul and Olga (continued)*: According to these rules, the commitment sets of Paul and Olga at the end of the example dialogue are;

$C_P(d_{11})\supseteq$ {airbag, newspaper: ''explode'', ¬ newspapers reliable};
$C_O(d_{11})\supseteq$ {¬ safe, airbag, newspaper: ''explode'', high max. speed}.

## 6 Instantiations: some distinctions

The above list of elements allows us to define some further distinctions between dialogue systems. With respect to outcomes, a distinction can be made between so-called *pure persuasion* and *conflict resolution.* The outcome of pure persuasion dialogues is fully determined by the participants' points of view and commitments:

- A dialogue system is for *pure persuasion* if and only if for any terminated dialogue $d$ it holds that $a\in w_t(d)$ if and only if:

  either $a\in prop(t)$ and $t\in C_{a'}(d)$ for all $a'\in prop(d)\cup opp(d)$;
  or $a\in prop(t)$ and $t\notin C_{a'}(d)$ for all $a'\in prop(d)\cup opp(d)$.

  Otherwise, it is for *conflict resolution.*

In addition, pure persuasion dialogues are assumed to terminate as soon as the right-hand side conjuncts of one of these two winning conditions hold.

*Paul and Olga (continued)*: In our running example, if the dialogue is regulated by a protocol for pure persuasion, it terminates after Paul's retraction.

In conflict resolution dialogues the outcome is not fully determined by the participant's points of view and commitments. In other words, in such dialogues it is possible that, for instance, a proponent of $\varphi$ loses the dialogue about $\varphi$ even if at termination he is still committed to $\varphi$. A typical

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

**Table 1** Locutions and typical replies

| Locutions | Replies |
|---|---|
| *claim* $\varphi$ | why $\varphi$, *claim* $\overline{\varphi}$, *concede* $\varphi$ |
| *why* $\varphi$ | $\varphi$ *since* $S$ (alternatively: *claim* $S$), *retract* $\varphi$ |
| *concede* $\varphi$ | |
| *retract* $\varphi$ | |
| $\varphi$ since S | why $\varphi$ ($\varphi \in S$), *concede* $\varphi$ ($\varphi \in S$) |
| *question* $\varphi$ | *claim* $\varphi$, *claim* $\overline{\varphi}$, *retract* $\varphi$ |

example is legal procedure, where a third party can determine the outcome of the case. For instance, a crime suspect can be convicted even if he maintains his innocence throughout the case.

A protocol has a *public semantics* if and only if the set of legal moves is always independent from the agents' belief bases.

A protocol is *context-independent* if the set of legal moves and the outcome are always independent of the context, so if $P(K,d)=P(\emptyset,d)$, $w_t^K(d)=w_t^\emptyset(d)$ and $l_t^K(d)=l_t^\emptyset(d)$, for all $K$, $d$ and $t$.

A protocol $P$ is *fully deterministic* if $P$ always returns a singleton or the empty set. It is *deterministic in* $\mathscr{L}_c$ if the set of moves returned by $P$ at most differ in their propositional content.

A protocol is *unique-move* if the turn shifts after each move; it is *multiple-move* otherwise.

*Paul and Olga (continued)*: The protocol in our running example clearly is multiple-move.

If the win and lose functions are defined on all legal dialogues instead of on terminated dialogues only, then another distinction can be made (Loui, 1998): a protocol is *immediate-response* if the turn shifts just in case the speaker is the ''current'' winner and if it then shifts to a ''current'' loser.

## 7 Instantiations: general characteristics and some issues

In this section some general characteristics of persuasion systems are discussed, as well as some design choices and related issues.

### *7.1 Dialectical obligations*

At first sight, it would seem natural to call the expectancies created by commitments ''(dialectical) obligations''. For instance, in many systems committing oneself to a proposition requires the speaker to support the proposition with an argument when it is challenged or else retract it. However, this can be called an obligation only in a loose sense. Some protocols allow points of order (e.g. that the challenger has the burden of proving that the claim is unwarranted), and some protocols allow that under certain conditions a challenge can be ignored, such as when an answer would be irrelevant, or when an answer can be postponed since it may become irrelevant because of some other way of continuing the dialogue. Strictly speaking, the only dialectical obligation that a participant has is making an allowed move when it is one's turn.

On the other hand, it still seems useful to systematize the loose sense of dialectical obligation. One way in which this can be done is by listing the typical replies of speech acts. Table 1 lists the typical replies of the common speech acts listed above. The list seems valid for all systems surveyed in this review, except that some systems, such as Gordon (1994, 1995) and Prakken (2000, 2005) do not allow *claim* $\varphi$ to be replied-to with *claim* $\overline{\varphi}$. Also, some systems with a non-monotonic and argument-based logic, such as Gordon (1994, 1995), Prakken (2000, 2005) and Parsons & McBurney (2003) allow arguments to be replied to with counterarguments and/or with moves that concede the argument (thus conceding only that its premises imply its conclusion, and reserving the right to challenge the premises; obviously the latter makes sense only for non-deductive arguments).

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

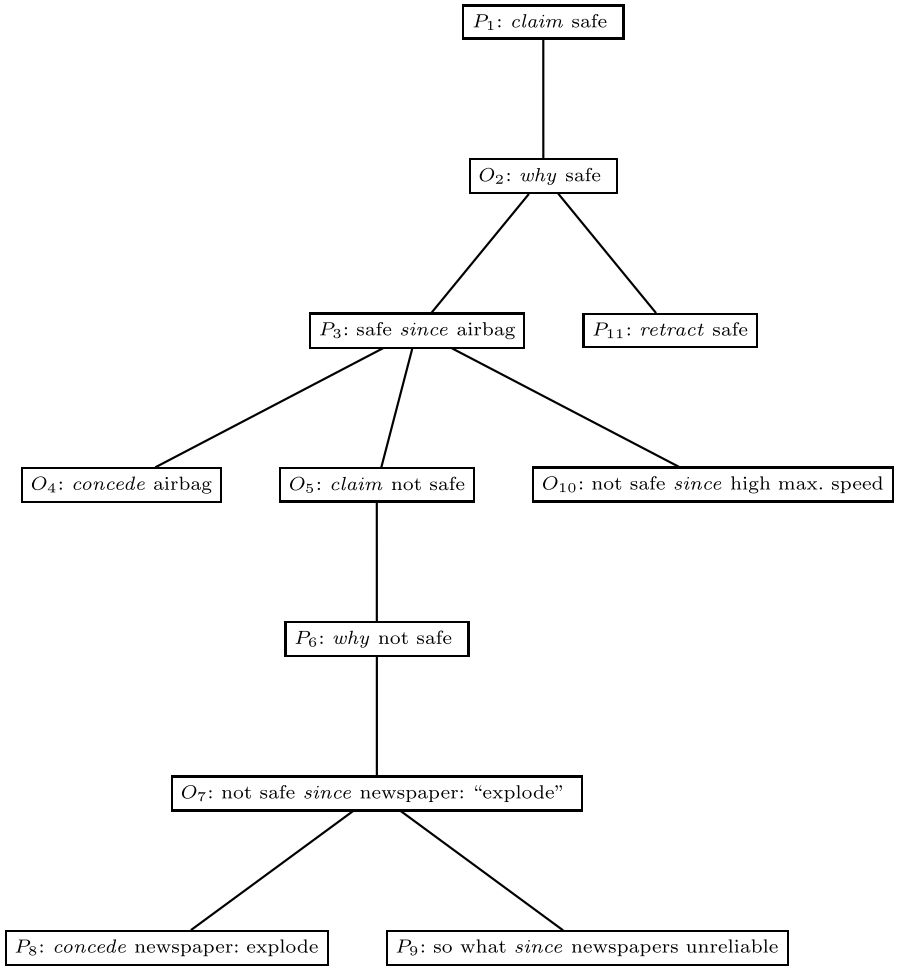



**Figure 1** Reply structure of the example dialogue

*Paul and Olga (continued)*: In terms of this table, our running example can now be displayed as in Figure 1, where the boxes stand for moves and the links for reply relations. A table such as the above one induces another distinction between dialogue protocols: a protocol is *unique-reply* if at most one reply to a move is allowed throughout a dialogue; otherwise it is *multiple-reply*. Of course, this distinction can be made fully precise only for systems that formally incorporate the notion of replies.

*Paul and Olga (continued)*: The protocol governing our running example is multiple-reply, as illustrated by the various branches in Figure 1.

### *7.2 Types of protocol rules*

According to their subject matter, several types of protocol rules can be distinguished. Some rules regulate a participant's *consistency*. This can be about *dialogical* consistency, such as Prakken (2000)'s rule that each move must leave the speaker's commitments consistent or Mackenzie (1979)'s rule that upon demand a speaker must resolve such an inconsistency. Or it can be about a participant's *internal* consistency, such as the use in Amgoud *et al.* (2000) and Parsons *et al.* (2003a, b) of assertion and acceptance attitudes. For instance, one of their protocol rules says that a participant may claim or accept a proposition only if his belief base contains an argument (or an acceptable argument) for the claim (see below). Such protocol rules are called ''rationality rules'' by Maudet & Evrard (1998).

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

Other rules are about *dialogical coherence*, such as the rules that require a non-initial move to be an appropriate reply to some earlier move (see e.g. the table above).

Yet other rules are about the *dialogical structure*, such as the termination rules and the rules that make the protocol a unique- or multiple move protocol, a unique- or multiple reply protocol, or an immediate- or non-immediate-response protocol.

### *7.3 Roles of commitments*

Commitments can serve several purposes in dialogue systems. One role is in enforcing a participant's dialogical consistency, for instance, by requiring him to keep his commitments consistent at all times or to make them consistent upon demand. Another role is to enlarge the hearer's means to construct arguments. For instance, the assertion and acceptance attitudes of Parsons *et al.* (2003b) assume that the participants reason from their own belief bases plus the other participants' commitments. A third role of commitments is to determine termination and outcome of a dialogue, such as in the above definition of pure persuasion. For example, in two-party pure persuasion the proponent wins as soon as the opponent concedes his main claim while the opponent wins as soon as the proponent has retracted his main claim.

A final role of commitments is to determine certain ''dialectical obligations''. Dialectical obligations concerning propositions can be formulated in two ways. One way is to attach the obligations to the speech acts that create a commitment, as was done in the reply-table above. Alternatively, the obligations can be directly attached to the content of one's commitment set. For instance, Walton & Krabbe (1995) allow the challenge of any commitment of another participant. In this approach, a distinction must be made between two kinds of commitments, viz. *assertions* and *concessions*, to reflect the different ways in which they were created (Walton & Krabbe, 1995): claims create asserted commitments and concessions create conceded commitments. Only assertions have a dialectical obligation attached to them.

### *7.4 Assertion and acceptance attitudes*

The Toulouse-Liverpool approach incorporates so-called ''assertion and acceptance attitudes'' into their protocols (see e.g. Parsons *et al.*, 2002, 2003a,b). The following assertion attitudes are distinguished: a *confident* agent can assert any proposition for which he can construct an argument, a *careful* agent can do so only if he can construct such an argument and cannot construct a stronger counterargument (in terms of some binary relation of strength between arguments), and a *thoughtful* agent can do so only if he can construct an ''acceptable'' argument for the proposition (in terms of the inference relation of the underlying argumentation system). The corresponding acceptance attitudes also exist: a *credulous* agent accepts a proposition if he can construct an argument for it, a *cautious* agent does so only if in addition he cannot construct a stronger counterargument and a *skeptical* agent does so only if he can construct an acceptable argument for the proposition. In this approach protocols are parameterized by the attitudes to be adopted by the participants, and the protocols require each participant to comply with his attitudes.

In all publications just-mentioned the attitudes are verified assuming that the participants reason with their internal belief base plus the commitments undertaken by the other participants. However, this is not an essential feature of these attitudes; in Section 9.3 we will come across a reason not to take the other participants' commitments into account.

It can be debated whether rules on respecting attitudes must be part of a protocol or of a participant's heuristics. According to one approach, a dialogue protocol should only enforce coherence of dialogues (e.g. Walton & Krabbe, 1995; Prakken, 2000); according to another approach, it should also enforce rationality of the agents engaged in a dialogue (e.g. Parsons & McBurney, 2003). The second approach allows protocol rules to refer to an agent's internal belief base and therefore such protocols do not have a public semantics (in the sense defined above in Section 6). The first approach does not allow such protocol rules and instead studies assertion and

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

acceptance attitudes as an issue of interaction between dialogue systems and the participant's strategies and heuristics.

### *7.5 The role of the logic*

The logic of most philosophical persuasion-dialogue systems is monotonic (usually standard propositional logic), while of most AI & Law and MAS-systems it is non-monotonic. Most of these are argument-based, but some are not. For instance, Lodder (1999)'s DiaLaw is based on reason-based logic (Hage, 1997; Verheij, 1996) and Brewka (2001)'s system is based on prioritized default logic (Brewka, 1994).

The logic of a persuasion-dialogue system can serve several purposes. Firstly, it can be used in determining consistency of a participant's commitments. For this purpose a monotonic logic must be used. Secondly, it can be used to determine whether the reasons given by a participant for a challenged proposition indeed imply the proposition. When the logic is monotonic, the sense of ''imply'' is obvious; when the logic is non-monotonic, ''imply'' means ''being an argument'' in argument-based logics and something like ''being a defeasible consequence from the premises alone' in other non-monotonic logics. Not all protocols require the reasons to be ''valid'' in these senses. For instance, Mackenzie (1979), Walton & Krabbe (1995) and Gordon (1994) allow the moving of enthymemes (but in Mackenzie (1979) and Walton & Krabbe (1995) this still commits the speaker to the material implication *premises*→*conclusion*). In Gordon (1995), an incomplete argument can be conceded by the hearer, which gives up the right to move a counterargument (but not the right to challenge its premises).

Note that this second use of a non-monotonic logic does not yet exploit the non-monotonic aspects of the logic. In argument-based terms, it only focuses on the construction of arguments, not on their defeat by counterarguments. This is different in a third use of the logic, viz. to determine whether a participant respects his assertion or acceptance attitude: as we have just seen, most of these attitudes are defined in terms of counterarguments or defeasible consequence. In particular, Parsons *et al.* (2003b)'s sceptical and thoughtful attitudes apply Dung (1995)'s grounded semantics: an argument is ''acceptable'' if it is in the grounded extension of the set of propositions the agent is supposed to reason with.

However, even if the full power of a non-monotonic logic is used, it is still possible to distinguish between internal and external use of the logic. In the Toulouse-Liverpool approach, the non-monotonic aspects of their (argument-based) logic are only used in verifying compliance with the assertion and acceptance attitudes; no other protocol rule refers to the notion of a counterargument. In particular, there is no rule allowing the attack of a moved argument by a counterargument. Also, the logic is not used in defining the outcome of a dialogue. Consequently (if the attitudes are regarded as heuristics and therefore external to a dialogue system), in these systems defeasible argumentation takes place only within an agent. In contrast, in the systems of e.g. Gordon (1994), Loui (1998) and Prakken (2000), the moving of counterarguments is explicitly allowed. And in Gordon (1994) and Loui (1998), the outcome of a dialogue is partly or wholly defined in terms of the logic (by checking whether the topic of the dialogue is defeasibly implied by the participants' joint commitments).

## 8 Investigation of properties of dialogue systems

So far the formal study of persuasion dialogue has been mainly definitional and application-oriented; the metatheory of persuasion protocols still largely awaits exploration. A notable exception is the work in the Toulouse-Liverpool approach, which in the publications discussed here makes strong assumptions on the participants' beliefs and reasoning behaviour, so that several interesting properties can be formally proven.

The most general property to be proven about dialogue systems for persuasion is that they induce coherent dialogues, that is, that they promote the resolution of conflicts of opinion. This is,

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

of course, too vague and general to be formally proven, so it must be decomposed into more specific properties. Loui (1998) states that dialogue protocols must be *fair* and *effective* given their goal. This is a good start for a decomposition, but much more is needed. McBurney *et al.* (2002) propose a list of 13 desiderata for argumentation protocols.

Complexity results can be proven of computational tasks of the participants, such as determining whether a move complies with the protocol. Parsons *et al.* (2002, 2003a,b) prove a number of complexity results, but only of reasoning tasks in the logic. Also, results can be proven on the minimal or maximal length of dialogues and on optimal choices of moves to minimize the length of a dialogue (cf. Parsons *et al.* 2003b; Dunne & McBurney, 2003). Furthermore, termination results can be proven. Generally, termination can be proven only if (but not if and only if!) premises of arguments are required to be taken from the dialogue context or the participants' belief bases. Without this condition, endless challenging is often possible: W: *claim p*, B: *why p*, W: *p* since *q*, B: *why q*, and so on.

Results can also be proven on the correspondence between the outcome of a dialogue and what is logically implied by the ''theory'' jointly constructed during the dialogue. For instance, it might be verified whether, if a (two-party) dialogue $d$ is won by the proponent, the topic $t$ is implied by $C_{pro}(d) \cup C_{opp}(d)$ (this holds by definition for pure persuasion), or by the premises of all arguments exchanged in the dispute, or by those such premises that were not challenged by the other party. In Prakken (2001b, 2005), such results are proven for several classes of persuasion games.

Similar correspondence results between dialogue outcomes and the participants' beliefs can be proven if assumptions are made about the participants' belief bases and heuristics (including the assertion and acceptance attitudes). For instance Parsons & McBurney (2003) prove for a protocol that if a dialogue terminates with both parties committed to the topic, then the topic is defeasibly implied by the participants' joint belief bases. Parsons *et al.* (2003a) prove for another protocol that this is not always the case. A weaker property that can be verified is that if at the start of a dialogue the joint belief bases of the participants defeasibly imply the topic, then there exists a legal dialogue that results in all adversaries being committed to the topic (and the minimal and maximal length of such a dialogue could be investigated).

When a dialogue protocol is fully specified in some formal language, then their metatheory can be investigated with the help of automated reasoning tools. Brewka (2001) specifies his protocols in a dialect of the situation calculus and Artikis *et al.* (2003) formalize variations of Brewka's protocols in the $C^+$ language of Giunchiglia *et al.* (2004). They then use implemented tools to verify various properties, such as the minimal length of dialogues that reach a certain state given a certain initial state.

## 9 Some systems

To illustrate the general discussion and some of the main design options, some persuasion protocols proposed in the literature will now be discussed. Several systems will be discussed in some detail. Among other things, they will be applied to our running example. Some other systems will be discussed more briefly at the end. In example dialogues of two-party dialogues the participants will be denoted with *W* and *B* or, if they have proponent/opponent roles throughout the dialogues, with *P* and *O*.

### *9.1 Mackenzie (1979)*

Much work applying persuasion dialogues is claimed to be based on Mackenzie (1979). In reality, however, usually all that is used from his system is a subset of his set of speech acts. In fact, we shall see that, with hindsight, the rest of his system is rather non-standard. Nevertheless, because of its historic role this system deserves a detailed discussion.

Mackenzie's system is for two parties with symmetric roles, i.e. they are subject to the same rules. Dialogues have no context. The communication language consists of claims, challenges, questions,

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

retractions and resolution demands (''resolve''). It has no explicit reply structure. The logical language is that of propositional logic (PL) but the logic is not full PL but instead a restricted notion of ''immediate consequence'', to capture resource-bounded reasoning (e.g. $p$, $p \rightarrow q$ and $q \rightarrow r$ immediately imply $q$ but not $r$). Arguments are moved implicitly as *claim* replies to challenges of another claim. An argument may be incomplete but its mover becomes committed to the material implication *premises*→*conclusion*. The logic is further mainly used for managing the participants' dialogical consistency, but in an indirect way: when one party challenges $\varphi$ but his commitments ''immediately'' imply $\varphi$, the other party may demand resolution: the challenger must then either retract one of his implying commitments or claim $\varphi$. Likewise if his commitments are ''immediately'' inconsistent: upon demand of resolution he must then retract one of the $\perp$-implying commitments.

As for commitments, they are mainly used to manage the participants' dialogical consistency, in the manner just explained. Mackenzie's commitment rules are quite non-standard since they are intended to capture the *silence implies consent* principle. Firstly, *claim* $\varphi$ commits not only the speaker but also the hearer to $\varphi$ (the hearer can end this commitment only by challenging $\varphi$). Also, if a *claim* $\varphi$ is moved as an argument for another claim $\psi$, then the implication $\varphi \rightarrow \psi$ is not only added to the speaker's but also to the hearer's commitments. Finally, a challenge commits the hearer to $\varphi$ and it commits the speaker to *why* $\varphi$, as part of Mackenzie's mechanism to avoid circular dialogues. This mechanism is further implemented by the prohibition to claim propositions to whose challenge the hearer is committed.

Mackenzie's protocol is unique-move and unique-reply but not deterministic. 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. Also, combined with the absence of a reply structure on $\mathcal{L}_c$ this makes that dialogues can become rather unfocused: only the moves required after questions and challenges and the conditions under which resolution demands may be made are constrained; apart from this the participants may freely exchange unrelated claims, and may freely challenge, retract or question. For instance, the following dialogue between Wilma and Bob is legal: *W*: *claim p*, *B*: *claim q*, *W*: *question r*, *B*: *claim* $\neg r$, *W*: *retract s*.

Mackenzie proves no properties about his system, although his system is explicitly meant to avoid circular dialogues. The following dialogue illustrates how Mackenzie may have achieved this goal. W: *claim p*, B: *why p*, W: *claim q*, B: *why q*, [W: *claim p*]. W's last move is not allowed, since B is committed to *why p*.

*Paul and Olga (continued)*: Our running example can be handled in Mackenzie's system as follows. The structural features of giving alternative replies and postponing replies give no problems since the protocol's lack of focus, but turns of more than one move cannot be modelled. Also, all counterarguments must now be introduced by a claim, which can be supported only after they have been challenged. Finally, since arguments can be incomplete, they can be modelled as stated in Section 2. The following legal dialogue then comes the closest to our version in Section 2:

$P_1$: *claim* safe;
$O_2$: *why* safe;
$P_3$: *claim* airbag;
$O_4$: *claim* $\neg$ safe;
$P_5$: *why* $\neg$ safe;
$O_6$: *claim* newspaper: ''explode'';
$P_7$: *claim* $\neg$(newspaper: ''explode'' $\rightarrow \neg$ safe);
$O_8$: *why* $\neg$(newspaper: ''explode'' $\rightarrow \neg$ safe);
$P_9$: *claim* $\neg$ newspapers reliable;
$O_{10}$: *claim* high max. speed;
$P_{11}$: *retract* safe.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

**Table 2** An example PPD dialogue

| Turn | Moves | $A_W$ | $C_W$ | $A_B$ | $C_B$ |
|---|---|---|---|---|---|
| | | $\{a\}$ | $\{b, c\}$ | $\{d, e\}$ | $\{f, g\}$ |
| $W_1$ | *why d*<br>*concede e* | | <br>$+e$ | | |
| $B_1$ | *why a*<br>*d since h, i,*<br>*i since j, k* | | | <br><br>$+h, h \wedge i \rightarrow d,$<br>$j, k, j \wedge k \rightarrow i$ | |
| $W_2$ | *concede j*<br>*concede k*<br>*why* $j \wedge k \rightarrow i$<br>*concede* $h \wedge i \rightarrow d$<br>*why h*<br>*a since f, j* | <br><br><br><br><br>$+f, j, f \wedge j \rightarrow a$ | $+j$<br>$+k$<br><br>$+h \wedge i \rightarrow d$ | | |
| $B_2$ | *h since l,* $l \rightarrow h$<br><br>$j \wedge k \rightarrow i$ *since m*<br>*concede* $f \wedge j \rightarrow a$<br>*retract*$_C$ *f* | | | $+l, l \rightarrow k,$<br>$l \wedge (l \rightarrow k) \rightarrow k$<br>$+m, m \rightarrow (j \wedge k \rightarrow i)$ | <br><br><br>$+f \wedge j \rightarrow a$<br>$-f$ |

Note that Paul's undercutter is modelled as a counterclaim of the ''hidden premise'' of $O_6$, to which Olga became committed by making this move in reply to $P_5$. Note also that since arguments cannot be built explicitly in support of a claim and the protocol has no explicit reply structure, $O_{10}$ cannot easily be recognized as an alternative support of $O_4$'s claim. Finally, note that the unique-move character of the protocol prevents a natural modelling of the various concessions.

### *9.2 Walton & Krabbe (1995)*

Walton & Krabbe define two dialogue systems for pure persuasion, PPD for ''permissive'' (Table 2) and RPD for ''rigorous'' persuasion dialogues. PPD is, like all other systems discussed in this review, for substantive discussions, where the parties try to identify and challenge each other's views on a certain topic. RPD is for discussions about the logical implications of the participant's commitments, which makes RPD more like a Lorenzen-style system of dialogue logic. Since the focus of this review is on substantive discussions, only PPD will be described here.

Dialogues have no context. It is assumed that in a (undefined) preparatory phase of a dialogue each player has declared zero or more assertions and concessions. The players are called White ($W$) and Black ($B$). Each participant is a proponent of his own and an opponent of the other participant's initial assertions. $B$ must have declared at least one assertion, and $W$ starts a dialogue. The communication language consists of challenges, (tree-structured) arguments, concessions, questions, resolution demands (''resolve''), and two retraction locutions, one for assertion-type and one for concession-type commitments (see below). The communication language has no explicit reply structure but the protocol reflects the reply table of Section 7.

The logical language is that of propositional logic and the logic consists of an incomplete set of deductively valid inference rules: they are incomplete to reflect that for natural language no complete logic exists. Although an argument may thus be incomplete, its mover becomes committed to the material implication *premises*→*conclusion.* The logic is used for managing the participants' dialogical consistency, in two indirect ways. Firstly, the system contains Mackenzie's *resolve* speech act for demanding resolution of two inconsistent commitments. Secondly, if a participant's commitments logically imply an assertion of the other participant but do not contain that assertion,

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

then the initial participant must either concede the assertion or retract one of the implying commitments.

The commitment rules are standard but Walton & Krabbe (1995) distinguish between several kinds of commitments for each participant, viz. *assertions*, *concessions* and *dark-side* commitments. Initial assertions and premises of arguments are placed in the assertions while conceded propositions are placed in the concessions. Only assertions carry dialectical obligations with them. Dark-side commitments, which remain fixed throughout the dialogue, are hidden or veiled commitments, which the participants bring to a dialogue and of which they (even the one who has them) are often unaware. They seem similar to the agents' private belief bases in the Toulouse-Liverpool approach but the fact that agents may not even be aware of their own dark-side commitments makes them hard to model computationally, for which reason they will be ignored below.

The protocol is driven by two main factors: the contents of the commitment sets and the content of the last turn. *W* starts and in their first turn both *W* and *B* either concede or challenge each initial assertion of the other party. Then the basic structure of a dialogue is that each turn consists of replies to all non-conceding moves in the other player's last turn; in particular, for *since* moves each premise must be conceded or challenged, including the hidden premise of incomplete arguments. Multiple replies are allowed, for instance, alternative arguments for the same assertion. Counter-arguments are not allowed. In sum, the PPD protocol is non-deterministic, multi-move and multi-reply but postponement of replies is not allowed. Dark-side commitments prevent the protocol from having a public semantics. Walton & Krabbe do not prove properties of the protocol.

The commitments constrain move legality in obvious ways to enforce dialogical coherence. For instance, a speaker cannot challenge or concede his own commitments, and *question* $\varphi$ and $\varphi$ *since* *S* may not be used if the listener is committed to $\varphi$. Finally, retractions must be successful in that the retracted proposition is not still implied by the speaker's commitments. The commitments also determine dialectical obligations in the way explained in Section 7, by the distinction between assertions and concessions. Finally, the commitments determine the outcome of a dialogue: dialogues terminate after a predetermined number of turns, and the outcome of terminated dialogues is defined as for pure persuasion.

Let us illustrate the system with the following example dialogue. The first column numbers the turns, and the second contains the moves made in each turn. The other columns contain the assertions and concessions of *W* and *B*: the first row contains the initial commitments and the other rows indicate changes in these sets: $+\varphi$ means that $\varphi$ is added and $-\varphi$ that it is deleted. If the dialogue terminates here, there is no winner, since neither player has conceded any of the other player's assertions or retracted any of his own.

Several points are worth noting about this example. Firstly, *B* in his first turn moves a complex argument, where the second argument provides an argument for a premise of the first; for this reason *i* is not added to *B*'s assertions. Next, in his second turn, *W* first concedes *j* and then asserts *j* as a premise of an argument; only after the second move has *W* incurred a burden of proof with respect to *j*. However, *B* in his second turn cannot challenge *j* since *B* is itself committed to *j*: if *B* wants to challenge *j* , he must first retract *j*. Another point to note is that after *B* concedes $f \wedge j \rightarrow a$ in his second turn, his commitments logically imply *a*, which is an assertion of *W*. Therefore *B* must in the same turn either concede *a* or retract one of the implying commitments. *B* opts for the latter, retracting *f*. Next we look at *B*'s second move of his second turn: remarkably, *B* becomes committed to a tautology but *W* still has the right to challenge it at his third turn. Finally, the example illustrates that the protocol only partly enforces relevance of moves. For instance, at any point, a participant could have moved *question* $\varphi$ for any $\varphi$ not in the commitments of the listener. For instance, *W* could have added *question n* to his first turn. Walton & Krabbe remark that relevance of PPD-dialogues partly depends on the cooperativeness of the participants.

*Paul and Olga (continued)*: Let us finally reconstruct our running example in PPD. To start with, Paul's initial claim must now be modelled as an initial assertion. Apart from this, since PPD has the

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

same underlying logic, the same treatment of hidden premises and the same communication language as Mackenzie's system, the moves are much the same as in that system. Two features of PPD make a straightforward modelling of the example impossible. The first is that PPD requires that every claim or argument is replied to in the next turn and the second is that explicit counterarguments are not allowed. To deal with the latter, it must be assumed that Olga has also declared an initial assertion, viz. that Paul's car is not safe. Then:

$O_1$: *why* safe;
$P_2$: safe *since* airbag;
$P_3$: *why* ¬ safe;
$O_4$: *concede* airbag;
$O_5$: ¬ safe *since* newspaper: ''explode''.

Here a problem arises, since Olga now has to either concede or challenge Paul's hidden premise airbag→safe. If Olga concedes it, she is forced to also concede Paul's initial claim, since it is now implied by Olga's commitments. If, on the other hand, Olga challenges the hidden premise, then at his next turn Paul must provide an argument for it, which he does not do in our original example. Similar problems arise with the rest of the example. To proceed with the example, let us now ignore this ''completeness'' requirement of turns:

$P_6$: *concede* newspaper: ''explode''.

Here another problem arises, since PPD does not allow Paul to move his undercutting counter-argument against $O_5$. The only way to attack $O_5$ is by challenging its unstated premise (newspaper: ''explode''→¬safe). In Mackenzie's system this problem did not arise since it allows the participants to make any claim at any moment, including ¬ (newspaper: ''explode''→¬safe).

In sum, two features of PPD prevent a fully natural modelling of our example: the monotonic nature of the underlying logic and the requirement to reply to each claim or argument of the other participant.

### *9.3 The Toulouse-Liverpool approach*

In a series of papers, researchers from Toulouse and Liverpool have developed an approach to specify dialogue systems for various types of dialogues, notably in Amgoud *et al.* (2000), Parsons *et al.* (2002, 2003a, b), McBurney & Parsons (2002) and Parsons & McBurney (2003). I here focus on their systems for persuasion, taking the most recent papers as the basis for discussion.

The systems are for two-party dialogues, and the participants appear to have a proponent and opponent role (but the roles may switch during a dialogue). Dialogues have no context but the participants have their own, possibly inconsistent belief base. Participants are assumed to adopt an assertion and acceptance attitude, which they must respect throughout the dialogue. The communication language consists of claims, challenges, concessions and questions; it has no explicit reply structure but the protocols largely conform to the table specified above in Section 5. Claims and concessions can concern both individual propositions and sets of propositions. The logical language is that of propositional logic. The logic is Amgoud & Cayrol (2002)'s argument-based non-monotonic logic in which arguments are classical proofs from consistent premises and in which counterarguments negate a premise of their target. Defeat relations between counterarguments are defined in terms of a priority relation on the premises and defeasible inference is then defined with Dung (1995)'s grounded semantics. Arguments are moved implicitly as *claim* replies to challenges of another claim and they must be complete. The logic is used to verify whether moved arguments are internally valid and whether the participants comply with their assertion and acceptance attitudes. The logic is not used externally, except in a protocol of Parsons & McBurney (2003), which explicitly allows the moving of counterarguments.

The commitment rules are standard and commitments are only used to enlarge the participant's belief base with the other participant's commitments. The protocols all appear to be unique-move

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

(but clause (4b) of the example protocol below seems ambiguous). They are multiple-reply in that each premise of an argument can be separately challenged or conceded, but otherwise multiple replies are not allowed. The example protocol discussed below appears to be deterministic in $\mathcal{L}_c$. Since agents have to comply with their assertion and acceptance attitudes, the semantics of the protocols are not public. Termination is defined in various ways in specific protocols. No explicit win and lose functions are defined, but in Parsons *et al.* (2003a) the possible outcomes are defined in terms of the propositions claimed by one participant and conceded by the other; and Parsons *et al.* (2003b) informally speak of ''having to concede the dialogue'' if no legal move can be made. As indicated above in Section 8, the strong assumptions made about the participants' beliefs and behaviour allow the formal verification of quite a number of properties of the systems.

The following example protocol is taken from Parsons *et al.* (2003b):

1. $W$ claims $\varphi$;
2. $B$ concedes $\varphi$ if its acceptance attitude allows, if not $B$ asserts $\neg\varphi$ if it is allowed to, or otherwise challenges $\varphi$;
3. if $B$ claims $\neg\varphi$, then goto 2 with the roles of the participants reversed and $\neg\varphi$ in place of $\varphi$;
4. if $B$ has challenged, then there is just one legal dialogue:
   (a) $W$ claims $S$, an argument for $\varphi$;
   (b) goto 2 for each $s \in S$ in turn.
5. $B$ concedes $\varphi$ if its acceptance attitude allows, or the dialogue terminates.

Dialogues terminate under a specific condition (condition 5), or when the move required by the protocol cannot be made, or when the player-to-move has conceded all claims made by the hearer, or when a locution is repeated by the same participant. Although this protocol does not forbid repetition of moves with the same content, such a prohibition still seems to be assumed in all papers of this approach and therefore this assumptions will be made below also.

Let us consider some simple dialogues that fit this protocol. First let $\Sigma_W=\{p\}$ and $\Sigma_B=\emptyset$. Then the only legal dialogue is:

$W_1$: *claim* $p$, $B_1$: *concede* $p$.

$B_1$ is $B$'s only legal move, whatever its acceptance attitude (note that, after $W_1$, $B$ must reason from $\Sigma_B \cup C_W(W_1)=\{p\}$). Here $B$ ''has to concede the dialogue'', since there are no claims of $W$ it has not accepted.

Consider next $\Sigma_W=\{q, q\rightarrow p\}$ and $\Sigma_B=\{\neg p\}$, where $q$ and $q\rightarrow p$ are preferred over $\neg p$. Then:

- $W_1$: *claim* $p$. It is now not entirely clear what preference relations hold for $p$; if all commitments made receive lowest priority (as proposed in Morge, 2004), then a credulous agent must concede $p$. A cautious and skeptical agent must instead proceed with $B_1$: *claim* $\neg p$; then $W$ must move $W_2$: *why* $\neg p$ after which $B$ replies with the trivial support *claim* $\{\neg p\}$, after which the non-repetition rule makes the dialogue terminate without agreement.

This example illustrates that even if a proposition is defeasibly implied by $\Sigma_W \cup \Sigma_B$, it may not be agreed upon by the participants. In fact, it also illustrates that sometimes there are no legal dialogues that agree upon such an implied proposition.

Together these examples suggest that if a claim is accepted, it is accepted in the first ''round'' of moves (but this should be formally verified). Also, since all arguments must be constructed ''in one shot'' from the participants' belief bases and commitments, dialogues tend to be short.

*Paul and Olga (continued)*: Finally, our running example can be modelled in this approach as follows. To do this, we must make some assumptions about the agents' internal knowledge bases. Let us assume that they believe in advance all propositions they state in the original version of the

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

example and that all propositions are equally preferred. Let us further first assume that Paul has a careful assertion attitude and a cautious acceptance attitude while Olga has a confident assertion attitude and a cautious acceptance attitude,. Note that arguments now have to be propositionally valid:

$P_1$: *claim* safe.

At first sight, it now seems that Olga's challenge of Paul's main claim can be straightforwardly modelled. However, this is not the case since Olga can construct an argument for the opposite claim and since her assertion attitude allows her to state it, her only legal move is:

$O_2$: *claim* ¬ safe.

Now since players may not repeat moves, Paul can only challenge Olga's counterclaim:

$P_3$: *why* ¬ safe;
$O_4$: *claim* {newspaper: ''explode'', newspaper: ''explode''→¬safe};
$P_5$: *concede* newspaper: ''explode''.

Paul had to concede Olga's first premise since he cannot construct a counterargument. However, he can and must attack Olga's second premise with a counterclaim:

$P_6$: *claim* ¬ (newspaper: ''explode''→¬safe).

Olga can build an argument for the opposite and her assertion attitude allows her to state it, therefore she must state it:

$O_7$: *claim* (newspaper: ''explode''→¬safe);
$P_8$: *why* (newspaper: ''explode''→¬safe);
$O_9$: *claim* {(newspaper: ''explode''→¬safe)}.

The non-repetition rule now makes the dialogue terminate without agreement. In this dialogue only Olga could develop her arguments (although she could not state her second counterargument). To change this, assume now that Olga has a thoughtful assertion attitude:

$P_1$: *claim* safe.

Now Olga's assertion attitude does not allow her to state her argument for ¬ safe, so she can only challenge Paul's claim:

$O_2$: *why* safe;
$P_3$: *claim* {airbag, airbag→safe};
$O_4$: *concede* airbag;
$O_5$: *why* airbag→safe;
$P_6$: *claim* {airbag→safe}.

Again the non-repetition rule makes the dialogue terminate without agreement. This time only Paul could develop his arguments (but not his counterargument).

Together, the two variants of the example illustrate that in this protocol it is hard, if not impossible, to have dialogues where arguments for and against the same claim are exchanged. It should be noted, however, that earlier systems in this approach, notably Amgoud *et al.* (2000), do not have this problem since they only weakly enforce focus of the dialogue, much in the style of Mackenzie's system. Of course, the downside of is that thus dialogues between non-cooperative participants may contain many irrelevant moves and may not even terminate.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

**Table 3** An example *Lc* in Prakken's framework

| Acts | Attacks | Surrenders |
|---|---|---|
| *claim* $\varphi$ | *why* $\varphi$ | *concede* $\varphi$ |
| $\varphi$ *since* $S$ | *why* $\psi$ ($\psi \in S$)<br><br>$\varphi'$ *since* $S'$<br>($\varphi'$ *since* $S'$ defeats $\varphi$ *since* $S$) | *concede* $\psi$ ($\psi \in S$)<br>*concede* $\varphi$ |
| *why* $\varphi$ | $\varphi$ *since* $S$ | *retract* $\varphi$ |
| *concede* $\varphi$ | | |
| *retract* $\varphi$ | | |

## 9.4 Prakken's framework

In Prakken (2000, 2005), a framework for specifying two-party persuasion dialogues is presented, which is then instantiated with some example protocols. The participants have proponent and opponent roles, and their beliefs are irrelevant to the protocols. Dialogues have no context. The framework largely abstracts from the logical language, the logic and the communication language but the logic is assumed to be argument-based and to conform to Dung (1995)'s grounded semantics. Also, arguments are assumed to be trees of deductive and/or defeasible inferences. The logic is used externally, to verify whether an argument of a *since* move is valid, and to allow for explicit counterarguments.

A main motivation of the framework is to ensure focus of dialogues while yet allowing for freedom to move alternative replies and to postpone replies. This is achieved with two main features of the framework. Firstly, an explicit reply structure on $\mathscr{L}_c$ is assumed, where each move either *attacks* or *surrenders to* its target. An example $\mathscr{L}_c$ of this format is displayed in Table 3. Secondly, winning is defined for each dialogue, whether terminated or not, and it is defined in terms of a notion of *dialogical status* of moves. The *dialogical status* of a move is recursively defined as follows, exploiting the tree structure of dialogues. A move is *in* if it is surrendered or else if all its attacking replies are *out.* (This implies that a move without replies is *in.*) And a move is *out* if it has a reply that is *in.* (Actually, this has to be refined to allow that some premises of an argument are conceded while others are challenged; see Prakken (2005) for the details.) Then a dialogue is (currently) won by the proponent if its initial move is *in* while it is (currently) won by the opponent otherwise.

Together, these two features of the framework allow for a notion of relevance that ensures focus while yet leaving the desired degree of freedom: a move is *relevant* just in case making its target *out* would make the speaker the current winner. Termination is defined as the situation where a players is to move but has no legal moves.

As for dialogue structure, the framework allows for all kinds of protocols. The instantiations presented in Prakken (2005) are all multi-move and multi-reply; one of them has the communication language of Table 3 and is constrained by the requirement that each move be relevant. This makes the protocol immediate-response, which implies that each turn consists of zero or more surrenders followed by one attacker. Within these limits postponement of replies is allowed, sometimes even indefinitely. As noted above in Section 8, various ''fairness'' and ''completeness'' properties are proven about this protocol.

In Prakken (2000), the protocol is further instantiated with Prakken & Sartor (1997)'s argument-based version of prioritized extended logic programming, which supports arguments about rule priorities. Let us now illustrate this instantiation with some examples. Consider two agents with the following belief bases (rule connectives are tagged with a rule name, which is needed to express rule priorities in the object language):

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

$\Sigma_P = \{q,\ q \Rightarrow r_1\ p,\ q \wedge s \Rightarrow_{r_3} r_1 > r_2\}$;
$\Sigma_O = \{r,\ r \Rightarrow_{r_2} \neg p\}$.

Then the following is a legal dialogue:

$P_1$ : *claim* $p$, $O_1$ : *why* $p$, $P_2$ : $p$ *since* $q$, $q \Rightarrow p$, $O_2$ : *concede* $q \Rightarrow p$, $O_3$ : *why* $q$.

At this point $P$ has three allowed moves, viz. retracting $p$, retracting $q$ or giving an argument for $q$. Note that the set of allowed moves is not constrained by $P$'s belief base. If the dialogue terminates here since $P$ withdraws from it then $O$ has won since $P_1$ is *out*.

The dialogue may also evolve as follows. The first three moves are as above and then:

$O_2$ : $\neg p$ *since* $r$, $r \Rightarrow \neg p$;
$P_3$ : $r_1 > r_2$ *since* $q$, $s$, $q \wedge s \Rightarrow r_1 > r_2$.

$P_3$ is a priority argument which in the underlying logic makes $P_2$ strictly defeat $O_2$ (note that the fact that $s$ is not in $P$'s own knowledge base does not make the move illegal). At this point, $P_1$ is *in*; the opponent has various allowed moves, viz. challenging or conceding any premise of $P_2$ or $P_3$, moving a counterargument to $P_3$ or a second counterargument to $P_2$, conceding one of these two arguments, and conceding $P$'s initial claim.

This example shows that the participants have much more freedom in this system than in the Toulouse-Liverpool approach. The downside of this is that dialogues can be much longer, and that the participants can prevent losing by simply continuing to challenge premises of arguments of the other participant. One way to tackle such ''filibustering'' is to introduce a context; another way (explored in Prakken, 2001a) is to introduce a third party who may reverse the burden of proof after a challenge: the challenger of $\varphi$ then has to provide an argument for $\overline{\varphi}$.

Another drawback of Prakken's approach is that not all dialogues that can be found in natural language conform to an explicit reply structure. For instance, in cross-examination dialogues (cf. Fulda, 2000; see also Angelelli, 1970), the purpose of the cross-examiner is to reveal an inconsistency in the testimony of a witness. Typically, questions by cross-examiners 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?

In the systems of Mackenzie and Walton & Krabbe, such dialogues can be modelled with the *question* locution, but at the price of decreased coherence and focus.

*Paul and Olga (continued)*: Let us finally model our running example in this protocol. Figure 2 displays the dialogue tree, where moves within solid boxes are *in* and moves within dotted boxes are *out*. As can be easily checked, this formalization captures all aspects of our original version, except that arguments have to be complete and that counterarguments cannot be introduced by a counterclaim. (But other instantiations of the framework may be possible without these limitations.)

### *9.5 Other systems*

I end this section with a briefer discussion of some other systems.

#### *9.5.1 The Pleadings Game*

Gordon's Pleadings Game (Gordon, 1994, 1995) was to my knowledge the first attempt to model legal procedure as persuasion dialogue. It was intended as a normative model of civil pleading in

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

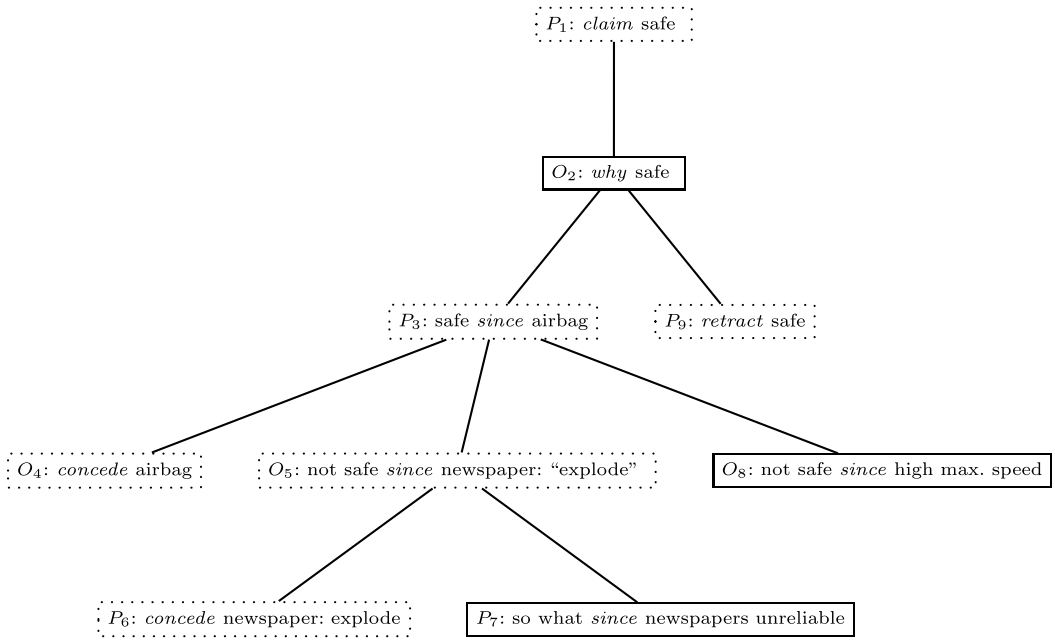



**Figure 2** The example dialogue in Prakken's approach

Anglo-American legal systems. The purpose of civil pleading is to identify the issues that must be decided in court. The game is formally defined and implemented as a system that verifies compliance of participants' moves with the protocol.

The logic is Geffner & Pearl (1992)'s conditional entailment, which has a model-theoretic semantics and an argument-based proof theory. Within this logic, Gordon has defined a method to reason about rule priorities. Gordon restricts inference in the logic to a notion of ''known implication'', similar in spirit to Mackenzie (1979)'s ''immediate consequence''. The logic is used externally, to define the notion of arguments and counterarguments (but arguments may be incomplete) and to determine the winner at termination by checking which of the participants' claims are defeasibly implied by the background theory jointly constructed during a dialogue. Each game starts with an initial background theory shared by the parties. During the game, it is continuously updated with each claim or premise of a party that is conceded by the other party.

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. Prakken's 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 Prakken) 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. The game does not have an explicit notion of commitments, but special protocol conditions enforce the participants' dialogical consistency. The protocol does not refer to the participants' internal beliefs and is not deterministic. Gordon does not explore the metatheory of his system.

The result of a terminated game is twofold: a list of issues identified during the game (i.e. the claims on which the players disagree), and a winner, if there is one. Winning is defined relative to the background theory constructed during a game. If issues remain, there is no winner and the case

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

**Table 4** Attackers and surrenders in TDG

| Acts | Attacks | Surrenders |
|---|---|---|
| *claim* $\varphi$ | *why* $\varphi$ | *concede* $\varphi$ |
| *why* $\varphi$ | *supply data*$_{\varphi}$ $\psi$ | *retract* $\varphi$ |
| *concede* $\varphi$ | | |
| *supply data*$_{\psi}$ $\varphi$ | so$_{\psi}$ $\varphi$<br>*why* $\varphi$ | *concede* $\varphi$ |
| so$_{\psi}$ $\varphi$ | *supply warrant* $\psi \Rightarrow \varphi$ | |
| *supply warrant w* | *presupposing w*<br>*on account of w* | *OK w* |
| *presupposing w* | *supply presupposition*$_{w}$ $\varphi$ | *retract w* |
| *on account of w* | *supply backing*$_{w}$ *b* | *retract w* |
| *supply backingw b* | | |

must be decided by the court. If no issues remain, then the plaintiff wins if his main claim is defeasibly implied by the final background theory, while the defendant wins otherwise. Thus, the Pleadings Game is a zero-sum game with three possible outcomes: win, lose or draw.

*9.5.2 Toulmin Diagram Game*

The Toulmin Diagram Game (TDG) of Bench-Capon (1998) and Bench-Capon *et al.* (2000) is intended to produce more natural dialogues than the ''stilted'' ones produced by systems such as those reviewed thus far. To this end, its speech acts are based on an adapted version of Toulmin (1958)'s well-known argument scheme. In this scheme, which can be regarded as TDG's (informal) underlying logic, a *claim* is supported by *data*, which support is *warranted* by an inference licence, which possibly has *presuppositions*, and which is *backed* by grounds for its acceptance; finally, a claim can be attacked with a *rebuttal*, which itself is a claim and thus the starting point of a counterargument. Arguments can be chained by regarding data also as claims, for which further data can be provided.

TDG's communication language is summarized in Table 4. For ease of comparison, this table has the reply format of Table 3 although the original system does not make such a reply structure explicit (TDG also has four dialogue control moves, which are not listed in the table).

The protocol is multi-move and multi-reply. The proponent starts with a claim and then the parties can reply to each other's moves according to Table 4. Backtracking moves are possible at any stage. To handle this, propositions are added to a ''claim stack'' when stated and removed from it when retracted or conceded; each move replies to the top claim in the stack, but with a *switch focus* $\varphi$ control move a player can make an earlier claim top of the stack and so reply to it. After a surrendering move or when an argument structure has been completed the turn switches to the referee, who then (in the first case) assigns the turn to the opponent of the top of the claim stack or (in the second case) invites the opponent to move a rebuttal. If the claim stack is empty, i.e. when all claims have been either conceded or retracted, the referee terminates the game. (In fact, the referee always has just one admissible move, so his role could have been hardwired in the protocol rules.) Finally, TDG has the usual commitment rules. The game has no explicit definition of a winner.

The idea to generate natural dialogues by defining the communication language in terms of some argumentation scheme has been applied to practical reasoning by Atkinson *et al.* (2005). They propose a more elaborate version of Walton (1996)'s argumentation scheme from consequences and define locutions in terms of the various premises and critical questions of this scheme.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

*9.5.3 Lodder's Dialaw*

The DiaLaw game of Lodder (1999) is the final development of the research of Hage *et al.* (1994). It is motivated by applications to legal disputes, reflected by is its underlying logic, which was especially developed for modelling legal reasoning. This logic is *reason-based logic*, a non-monotonic logic developed jointly by Hage (1997) and Verheij (1996).

The DiaLaw game has two participants, with symmetric dialectical roles. They can use locutions for claiming a proposition and for challenging, conceding and retracting a claimed proposition. A claim can also be attacked by claiming its negation. Arguments are constructed implicitly, with a *claim* $\psi$ reply to a *why* $\varphi$ attack on a *claim* $\varphi$ move. A supporting claim is not required to logically imply the supported claim. However, the game does not provide a means to attack arguments on their invalidity. Discussions about procedural correctness of claims are modelled with a special first-order predicate ''illegal'' ranging over dialogue moves.

As for turntaking, each dialogue begins with a claim from one player, and then the turn switches after each move, except in a few cases where surrenders are moved. A dialogue terminates if no disagreement remains, i.e. if no commitment of one player is not a commitment of the other. The first player wins if at termination he is still committed to his initial claim, the second player wins otherwise. Thus DiaLaw is for pure persuasion.

DiaLaw contains the usual commitment rules for claims, concessions and retractions. The commitments are not logically closed, but several rules make the making, conceding or retraction of claims obligatory depending on what logically follows from one's commitments (cf. also Walton & Krabbe 1995). The game contains detailed protocol rules for each specific type of move. This makes the system fine-tuned to the intended applications but also less transparent. For instance, the system has no general rule on whether backtracking is admissible.

*9.5.4 Argument games*

A special case of a dialogue system is an *argument game* i.e. a two-party dialogue where a proponent and an opponent move arguments and nothing else. Argument games can be used as a proof theory for argumentation logics (see e.g. Prakken & Vreeswijk, 2002). For instance, the following game of Prakken & Sartor (1997) is sound and complete with respect to Dung (1995)'s grounded semantics:

- Proponent begins with an argument for a claim he wants to defend
- At each other move:
  - opponent replies to the previous move with a counterargument that is at least as strong as its target;
  - proponent replies to the previous move with a counterargument that is stronger than its target (while not repeating his previous arguments).

In addition, Loui (1998), Jakobovits & Vermeir (1999), Jakabovits (2000) and Prakken (2001b) have studied argument games in their own right. Both Jakobovits (2000) and Prakken (2001b) have observed that when a game protocol is ''dynamified'', i.e. when the theory from which the arguments are constructed is not given in advance but consists of the premises of all arguments moved in a dialogue, the properties of the game can change. A positive change is proven by Jakobovits, viz. that certain dynamic argument game protocols prevent the construction of theories whose argument attack graph contains odd loops (it is well known that such theories may have no extensions). A negative result is proven by Prakken, viz. that in the dynamified version of the above game, the initial claim may at a certain state of the dialogue be defeasibly implied by the theory constructed thus far while yet the proponent has no way to proceed the dialogue that makes him win. Prakken also proves a positive version of this result for so-called ''relevant protocols'', which are protocols that make all relevant moves allowed that satisfy the rest of the protocol (see for relevance Section 9.4 above).

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

## 10 Conclusion

In this review, we have critically reviewed a number of systems for persuasion dialogue in terms of a formal specification of the main elements of such systems. Concluding, we can say that the formal study of persuasion dialogue has resulted in many interesting dialogue-game protocols for persuasion, some of which have been applied in insightful case studies or applications, but that a consensus on many issues is still lacking. As a consequence, there is still little work on formally relating the various systems or on a general framework for designing persuasion protocols, and a formal metatheory of systems is still in its early stages. These are some of the main issues that should be tackled in future research. Some other issues are the study of strategies and heuristics for individual participants and how these interact with the protocols to yield certain properties of dialogues, a similar study of varying degrees of cooperativeness of participants, and the integration of persuasion systems with systems for other types of dialogues. Perhaps the main challenge in tackling all of these issues is how to reconcile the need for flexibility and expressiveness with the aim to enforce coherent dialogues. The answer to this challenge may well vary with the nature of the context and application domain, and a precise description of the grounds for such variations would provide important insights in how dialogue systems for persuasion can be applied.

## Acknowledgement

This research was partially supported by the EU under IST-FP6–002307 (ASPIC).

## References

Alexy, R, 1978, *Theorie der juristischen Argumentation. Die Theorie des rationalen Diskurses als eine Theorie der juristischen Begrunding.* Suhrkamp Verlag: Frankfurt am Main.

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

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

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

Artikis, A, Sergot, M and Pitt, J, 2003, An executable specification of an argumentation protocol. In *Proceedings of the Ninth International Conference on Artificial Intelligence and Law.* New York. ACM Press, pp. 1–11.

Atkinson, K, Bench-Capon, T and McBurney, P, 2005, A dialogue game protocol for multi-agent argument over proposals for action. *Journal of Autonomous Agents and Multi-Agent Systems* **11**, 153–171.

Barwise, J and Moss, L, 1996, *Vicious Circles.* No. 60 in CSLI Lecture Notes. Stanford, CA: CSLI Publications.

Bench-Capon, T, 1998, Specification and implementation of Toulmin dialogue game. In *Legal Knowledge-Based Systems. JURIX: The Eleventh Conference.* Nijmegen: Gerard Noodt Instituut, pp. 5–19.

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

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

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

Carlson, L, 1983, *Dialogue Games: An Approach to Discourse Analysis.* Dordrecht: Reidel.

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

Dunne, P and McBurney, P, 2003, Optimal utterances in dialogue protocols. In *Proceedings of the Second International Conference on Autonomous Agents and Multiagent Systems*, pp. 608–615.

Fulda, J, 2000, The logic of ''improper cross''. *Artificial Intelligence and Law* **8**, 337–341.

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

Giunchiglia, E, Lee, J, Lifschitz, V, McCain, N and Turner, H, 2004, Nonmonotonic causal theories. *Artificial Intelligence* **153**, 49–104.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

Gordon, T, 1994, The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law* **2**, 239–292.
Gordon, T, 1995, *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice.* Dordrecht: Kluwer.
Hage, J, 1997, *Reasoning With Rules. An Essay on Legal Reasoning and Its Underlying Logic.* Law and Philosophy Library. Dordrecht: Kluwer.
Hage, J, Leenes, R and Lodder, A, 1994, Hard cases: a procedural approach. *Artificial Intelligence and Law* **2**, 113–166.
Hamblin, C, 1970, *Fallacies.* London: Methuen.
Hamblin, C, 1971, Mathematical models of dialogue. *Theoria* **37**, 130–155.
Jakobovits, H, 2000, *On the Theory of Argumentation Frameworks.* Doctoral dissertation, Free University Brussels.
Jakobovits, H and Vermeir, D, 1999, Dialectic semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law.* New York: ACM Press, pp. 53–62.
Kraus, S, Sycara, K and Evenchik, A, 1998, Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence* **104**, 1–69.
Lodder, A, 1999, *DiaLaw. On Legal Justification and Dialogical Models of Argumentation.* Law and Philosophy Library. Dordrecht: Kluwer.
Loui, R, 1998, Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence* **14**, 1–38.
Mackenzie, J, 1979, Question-begging in non-cumulative systems. *Journal of Philosophical Logic* **8**, 117–133.
Mackenzie, J, 1990, Four dialogue systems. *Studia Logica* **51**, 567–583.
Maudet, N and Evrard, F, 1998, A generic framework for dialogue game implementation. In *Proceedings of the Second Workshop on Formal Semantics and Pragmatics of Dialogue.* Enschede, The Netherlands: University of Twente.
Maudet, N and Moore, D, 1999, Dialogue games for computer-supported collaborative argumentation. In *Proceedings of the Workshop on Computer-Supported Collaborative Argumentation for Learning Communities*, Stanford.
McBurney, P and Parsons, S, 2002, Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information* **13**, 315–343.
McBurney, P, Parsons, S and Wooldridge, M, 2002, Desiderata for agent argumentation protocols. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems*, pp. 402–409.
Moore, D, 1993, *Dialogue Game Theory for Intelligent Tutoring Systems.* PhD thesis, Leeds Metropolitan University.
Morge, M, 2004, Computer supported collaborative argumentation. In *Proceedings of the Fourth International Workshop on Computational Models of Natural Argument*, pp. 69–72.
Parsons, S and McBurney, P, 2003, Argumentation-based communication between agents. In *Communications in Multiagent Systems (Springer Lecture Notes in AI, 2650).* Berlin: Springer, pp. 164–178.
Parsons, S, Sierra, C and Jennings, N, 1998, Agents that reason and negotiate by arguing. *Journal of Logic and Computation* **8**, 261–292.
Parsons, S, Wooldridge, M and Amgoud, L, 2002, An analysis of formal interagent dialogues. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems*, pp. 394–401.
Parsons, S, Wooldridge, M and Amgoud, L, 2003a, On the outcomes of formal interagent dialogues. In *Proceedings of the Second International Conference on Autonomous Agents and Multiagent Systems*, pp. 616–623.
Parsons, S, Wooldridge, M and Amgoud, L, 2003b, Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation* **13**, 347–376.
Pollock, J, 1995, *Cognitive Carpentry. A Blueprint for How to Build a Person.* Cambridge, MA: MIT Press.
Prakken, H, 2000, On dialogue systems with speech acts, arguments, and counterarguments. In *Proceedings of the Seventh European Workshop on Logic for Artificial Intelligence (Springer Lecture Notes in AI, 1919).* Berlin: Springer, pp. 224–238.
Prakken, H, 2001a, Modelling reasoning about evidence in legal procedure. In *Proceedings of the Eighth International Conference, on Artificial Intelligence and Law.* New York: ACM Press, pp. 119–128.
Prakken, H, 2001b, Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese* **127**, 187–219.
Prakken, H, 2005, Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation* **15**, 1009–1040.
Prakken, H and Sartor, G, 1997, Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics* **7**, 25–75.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press

Prakken, H and Vreeswijk, G, 2002, Logics for defeasible argumentation. In Gabbay, D and Günthner, F. (eds.), *Handbook of Philosophical Logic*, 2nd edn., vol. 4. Dordrecht: Kluwer, pp. 219–318.
Toulmin, S, 1958, *The Uses of Argument.* Cambridge: Cambridge University Press.
Verheij, B, 1996, *Rules, Reasons, Arguments: Formal Studies of Argumentation and Defeat.* Doctoral dissertation, University of Maastricht.
Walton, D, 1984, *Logical Dialogue-Games and Fallacies.* Lanham, MD: University Press of America, Inc.
Walton, D, 1996, *Argumentation Schemes for Presumptive Reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates.
Walton, D, 2003, Is there a burden of questioning? *Artificial Intelligence and Law* **11**, 1–43.
Walton, D and Krabbe, E, 1995, *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning.* Albany, NY: State University of New York Press.
Woods, J and Walton, D, 1978, Arresting circles in formal dialogues. *Journal of Philosophical Logic* **7**, 73–90.
Yuan, T, 2004, *Human-Computer Debate, A Computational Dialectics Approach.* PhD thesis, Leeds Metropolitan University.

https://doi.org/10.1017/S0269888906000865 Published online by Cambridge University Press