



HAL
open science

Argumentation-based Dialogue

Elizabeth Black, Nicolas Maudet, Simon Parsons

► **To cite this version:**

Elizabeth Black, Nicolas Maudet, Simon Parsons. Argumentation-based Dialogue. Dov Gabbay; Massimiliano Giacomin; Guillermo R. Simari; Matthias Thimm. Handbook of Formal Argumentation, Volume 2, College Publications, 2021. <hal-03429859>

HAL Id: hal-03429859

<https://hal.science/hal-03429859v1>

Submitted on 16 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Argumentation-based Dialogue

ELIZABETH BLACK, NICOLAS MAUDET, AND SIMON PARSONS

ABSTRACT. Dialogue is fundamental to argumentation, providing a dialectical basis for establishing which arguments are acceptable. Argumentation can also be used as the basis for dialogue. In such “argumentation-based” dialogues, participants take part in an exchange of arguments, and the mechanisms of argumentation are used to establish what participants take to be acceptable at the end of the exchange. This chapter considers such dialogues, discussing the elements that are required in order to carry out argumentation-based dialogues, giving examples, and discussing open issues.

1 Introduction

Anyone these days who develops an interest in formal models of argumentation is likely to encounter [Dung, 1995] early in their studies. For many readers the key elements of this paper will be the abstract argumentation framework that it introduces, or its account of the formal process for establishing the acceptability of arguments, work which has led to much subsequent work on argumentation semantics. However, there is another aspect to [Dung, 1995] that interests us here. That is the notion of dialogue. More precisely, it is the idea of dialogue as an exchange between two or more individuals, an exchange which captures features of what would be informally called an “argument”. That is, dialogue as the exchange of reasons for or against some matter. This idea crops up right at the start of [Dung, 1995], where readers are asked to imagine the following exchange between two individuals:

A: My government can not negotiate with your government because your government doesn't even recognize my government.

B: Your government doesn't recognize my government either.

A: But your government is a terrorist government.

Later in the introduction, [Dung, 1995] motivates the work in the paper by saying that:

understanding of the structure and acceptability of arguments is essential for a computer system to be able to engage in exchanges of arguments

and it is this latter idea, or a slight modification of it, that interests us in this chapter. We are interested in how computer *systems* might be able to engage in exchanges of reasons for or against some matter (where such reasons are themselves often referred to as “arguments”¹).

In keeping with current terminology in artificial intelligence [Russell and Norvig, 2020], we will refer to these computer systems as “agents”, considering them to be embedded in a multiagent system [Wooldridge, 2009]. While in this chapter we will only be interested in providing formal models for exchanges between artificial agents, by which we mean that we do not consider aspects of natural language understanding or generation, we acknowledge that some of the agents may be human. That is, the multiagent system may be a combination of software agents and human agents. Indeed, the long-term vision for work on argumentation-based dialogue, as this area of study has become known, is for seamless interaction between humans and artificial agents. Further, the motivation for considering that it makes sense to ground such communication on models of argumentation is exactly that human reasoning and communication appears to naturally align with the process of constructing and evaluating arguments [Mercier and Sperber, 2011].

While the formal developments in [Dung, 1995], and in the large body of work on abstract argumentation that has been built on it, do not directly address exchanges between agents, there is a clear connection to such exchanges. The exchange between *A* and *B* that we quoted from [Dung, 1995] can be written as a *dialogue*, that is, as a conversation between multiple individuals. It can also be viewed as part of a *dialectical* process — the process of investigating the truth of an opinion. In this reading, the process is one in which an argument is put forward, followed by a counter-argument, followed by a counter to the counter-argument, and so on. This sequence can then be analysed to determine which of the arguments are acceptable, and variations on the process provide an “argument game” proof theory for different approaches to establishing acceptability [Modgil and Caminada, 2009]. Investigating such proof theories is not our topic here, though, as we will see, there are clear parallels between this work and the kinds of inter-agent dialogue that we are interested in.

Having said what we are not directly concerned with in this chapter, we should say in detail what we are directly concerned with. We have already sketched this, describing a focus on dialogues between agents in a multiagent system. The following example, from [Prakken, 2006], provides more detail on the kind of dialogue we mean:

- (1) Paul : My car is very safe. (making a claim)
- (2) Olga : Why is your car safe? (asking grounds for a claim)
- (3) Paul : Since it has an airbag. (offering grounds for a claim)

¹Of course, this double use of the term “argument” is common in natural language, where it is used both to refer to a reason for or against a conclusion, and an exchange of such reasons.

- (4) Olga : That is true. (conceding a claim) but I disagree that this makes your car safe: the newspapers recently reported on airbags expanding without cause. (stating a counter-argument)
- (5) 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 counter-argument)
- (6) Olga : Still your car is not safe, since its maximum speed is very high. (alternative counter-argument)

Thus, it is clear that a focus of this line of work is the development of software systems that can engage in interactions which involve the exchange — and hence the generation and interpretation of — arguments. And we have also suggested that humans might also be part of these interactions. However, this description gives a rather narrow view of the full scope of the work.

For a start, the use of the word “dialogue” can be misleading, suggesting that the focus is only on interactions between just two entities², a misunderstanding that can be exacerbated by the focus of much of the literature on dialogues that deal with just two participants. (Sometimes this restriction is explicit, sometimes merely implicit or implied by the examples used.) However, there is no such restriction in general. In other words, the aim of work on argumentation-based dialogue is to support interactions between arbitrarily large numbers of agents.

Second, agents may play a number of different roles in interactions, and these may have an effect on aspects of the dialogue. In a lot of work, agents take part in dialogues as equal participants. That means that interactions are, in some sense, symmetrical, though there is asymmetry that arises because of what utterances are made, not because of the nature of the participants. For example, in the simple persuasion model of [Parsons *et al.*, 2003], which we will discuss below, the first agent to make an utterance is the one that accepts the burden of doing the persuading — making the case — and is also insulated from being persuaded in the sense that the rules of the dialogue do not allow it to change its position. In other cases, one distinguishes between the agents engaged in a dialogue, and those observing it. For example, in [Prakken, 2001a] a distinction is made between the two agents who are arguing their positions, and the arbitrator which observes them, decides on the legality of the arguments that they make, and ultimately reaches a decision about which party has won the dispute. In this case, the two arguing agents are not attempting to persuade each other, but to convince the arbitrator. The arbitrator thus plays the role of the judge in some legal systems, settling the dispute based not on its own view, but on the arguments that have been made. In this kind of legal case, in front of a judge, the agents will be looking to make arguments that have legal impact

²Perhaps because of a misunderstanding of the dichotomy between monologue and dialogue, or a misreading of the prefix “dia” as meaning “two”.

— that are, in some sense the strongest arguments from a legal perspective. In contrast, in front of a jury, the arguers might instead aim to come up with the arguments that they think will best influence jury members, for example seeking to make arguments with emotional force. In either case, the arguers take their audience into account, and evaluate what they might say, choosing between possible arguments, by using some model of the audience. (Section 5.2 discusses how an agent might use such a model to determine which argument to assert.)

In the two cases just described, the audience to the dialogue is sitting in judgement. There is another case in which the audience is not judging, but has the aim of helping the parties in the dialogue reach a solution. Thus they act as a mediator rather than judge. We can also imagine other scenarios, for example where the audience will be reporting back on the dialogue to some external party. In such a situation the participants in the dialogue may be aware of the reporter, or they may be unaware. And if they are aware, they may shade their utterances to influence the external party (that is their aim may be less about their overt goal of “winning” the dialogue and more about securing a favorable report)³.

All of these facets will influence the detail of the strategies that agents employ in a dialogue, and they may influence the design and structure of the dialogue mechanism. However, all these different dialogues and scenarios point to the same underlying questions, for example;

- What are the main components of a dialogue?

In other words, when we design a dialogue, what are the major elements that we have to consider?

- How are dialogues represented?

If we want to formally reason about dialogues, how do we represent the various components?

- What makes a good dialogue?

Once we can represent and reason about (or implement and experiment with) dialogues, we need to establish what makes a good (or better) dialogue.

- How do we design dialogue strategies?

The questions so far have looked at dialogues as opposed to the participants in the dialogue. When we look at dialogues from the participants’ perspective, how do we design the elements that allow dialogue participants to participate?

³[Black and Sklar, 2016] considers further some of the issues that arise when there are multiple dialogue participants of different types.

The remainder of this chapter examines these questions, and discusses work that has been done on them to date. To do so, we have drawn heavily on our own work to illustrate the study of argumentation-based dialogue, making this chapter much more of a position piece than an exhaustive survey. This does not, of course, mean that we are the only researchers to have contributed to this line of work. We have attempted to identify other work that is relevant, and thus to present a relatively balanced picture. Where we have failed to mention work that we should have included, we apologise.

In detail, the contents of the rest of the chapter are as follows. Section 2 provides an introduction to the basic ideas in argumentation-based dialogue, and gives an example of such a dialogue. Then, Section 3 discusses a couple of approaches to representing dialogues, and Section 4 provides an in-depth description of one specific formal dialogue system, illustrating many of the ideas that have been introduced. Section 5 expands on several of the key unresolved issues in argumentation-based dialogue, including how to handle multiparty dialogues, the development of dialogue strategies, and how to handle enthymemes⁴ in the context of dialogues. Section 6 concludes.

2 Basic notions

Since work on argumentation-based dialogue pulls together a number of different strands of work, there are several different ways that one might reasonably organise this section. We choose to start with [Walton and Krabbe, 1996], primarily because it provides a widely used classification of different kinds of dialogue.

2.1 Types of dialogue

In particular, [Walton and Krabbe, 1996] suggests that dialogues can be classified by what the participants know, what the participants seek to get from the dialogue, and what the dialogue rules are intended to bring about⁵. The basic classification is often written as follows:

Information-seeking Dialogues are dialogues in which one participant seeks the answer to some question or questions, and looks to obtain answers from another participant, who is believed by the first to know what these answers are.

Inquiry Dialogues are dialogues in which the participants collaborate to answer some question or questions whose answers are not known to either participant.

⁴An enthymeme is an incomplete argument, where some of the premises and/or the claim of the intended complete argument are omitted. One of the mismatches between formal argumentation, and argumentation as used by people, is that while formal arguments are complete, people usually deal in enthymemes, eliding the “obvious” bits of the arguments that they present.

⁵[Walton and Krabbe, 1996] makes use of the notion of the goal of the dialogue, which we find a problematic concept since it suggests intentionality on the part of the rules that specify how the dialogue proceeds.

Persuasion Dialogues involve one participant seeking to persuade another to accept a proposition they do not hold at the beginning of the dialogue. This can mean that the persuadee is agnostic about the position, or it can mean that the persuadee holds the opposite position.

Negotiation Dialogues are dialogues in which the participants bargain over the division of some scarce resource. If a negotiation dialogue terminates with an agreement, then the resource has been divided in a manner acceptable to all participants. Though this may seem to be a rather specific notion of negotiation, it can be made quite broad by a suitably careful choice of the scarce resource in question.

Deliberation Dialogues are dialogues in which the participants collaborate to decide what action or course of action should be adopted in some situation. Here, participants share a responsibility to decide the course of action, or, at least, they share a willingness to discuss whether they have such a shared responsibility. Participants may have only partial or conflicting information, and conflicting preferences. As with negotiation dialogues, if a deliberation dialogue terminates with an agreement, then the participants have decided on a mutually-acceptable course of action.

Eristic Dialogues are those in which participants quarrel verbally as a substitute for physical fighting, aiming to vent perceived grievances. At the time of writing, eristic dialogues have not been widely studied in computer science.

A number of these forms of dialogue have been studied in detail by the argumentation community, for example: information-seeking [Fan and Toni, 2012], inquiry [Black and Hunter, 2009; Hitchcock, 1991; Hulstijn, 2000], persuasion⁶ [Gordon, 1994; Prakken, 2006; Walton and Krabbe, 1996], negotiation [Rahwan *et al.*, 2003] and deliberation [McBurney *et al.*, 2007].

[Walton and Krabbe, 1996] makes no claim that this set of dialogues is exhaustive, and plenty of other examples can be found in the literature. [Walton and Krabbe, 1996] also suggests that dialogues can be formed from combinations of these basic kinds of dialogue, and this is a point that we will return to below. Since the connection was made between this work — which has its roots in the philosophical study of argumentation — and the work that had been developed in within the AI and multiagent systems communities, these definitions have become widely adopted.

2.2 Hamblin-style dialectical games

In addition to the typology, [Walton and Krabbe, 1996] is also notable for describing a persuasion dialogue. This is done in the form of a *dialogue game*

⁶Persuasion is perhaps the most widely studied form of dialogue, and these are just a few of the papers to have studied it.

between two or more players, the participants in the dialogue. The game specifies the moves that the players can make, and the rules that dictate which moves are possible when. This is in line with the tradition of Hamblin’s *dialectical games* [1970; 1971]. Under this model, each agent holds a *commitment store*, accessible to both agents, which contains statements made during the dialogue. Since commitment stores evolve over time, we index them both by agent and by the step in the dialogue game. Thus we write the commitment stores for x and y at step k as $CS_{x,k}$ and $CS_{y,k}$ respectively⁷. The function of commitment stores varies between dialogue systems. The name stems from the idea that agents only put forward statements that they believe are in some sense true, and so making the statements is a commitment to defend that position if challenged. In the dialogue system we shall give as an illustration here, an agent’s commitment store is just a subset of what can be inferred from its knowledge base, a subset that it chooses to make public. In other work, [Amgoud *et al.*, 2000b] for example, one agent is allowed to place commitments on another by placing statements into that agent’s commitment store. It is perhaps the work of Mackenzie [1979b; 1979a; 1990] which made such models popular outside the formal dialectics community. His game DC is perfect representative of such dialectical games, whose primary objective was to study and avoid fallacies in dialogues. The game is defined through several rules, eg.:

“No statement may occur if it is a commitment of both speaker and hearer at that stage”

or

“After a statement p , unless the preceding event was a challenge, p is included in both participants’ commitments.”

A bit more formally, this means that we can specify for the dialogue move $\langle x, \text{assert}, p \rangle$, meaning agent x asserts formula p , the following conditions and updates:

| | |
|---|--|
| $\langle x, \text{assert}, p \rangle$ conditions update | where p is a formula. $p \notin CS_{x,k}$ and $p \notin CS_{y,k}$ $CS_{x,k+1} = CS_{x,k} \cup \{p\}$ and $CS_{y,k+1} = CS_{y,k}$ |
|---|--|

Other dialogue rules specify what moves can follow what other moves. However, these specifications often conflate different components which should be distinguished.

2.3 Components of a dialogue

We describe in turn the different components of dialogue systems: locutions, dialogues, protocols, and finally agents’ strategies.

⁷Note, we refer to agents using the variables x and y , such that $x \neq y$.

Locutions and dialogue moves. We start with the set of *moves* available to the agents. We can think of these as being made up of a set of *locutions*, capturing the kinds of things that agents are allowed to say, and which can be instantiated with certain propositions. The instantiated locutions make up the possible dialogue moves. There is no restriction on the locutions that can be used, however a common set of locutions found in such games is:

$$\{\text{assert, accept, challenge, question, retract, argue}\}$$

We call the set of locutions L . They allow for claims and their supporting arguments to be stated (or indeed retracted), arguments to be requested and questions to be asked. The set of locutions L can be instantiated to create a set of moves M . Assuming only two agents, we simply write dialogue moves as

$$\langle x, \text{locution}, \phi \rangle$$

where x is the speaker, $\text{locution} \in L$ is a locution, and ϕ is a formula built from some content language. For instance, these can be instantiated with propositions (and, in the case of **assert** and **accept**, may be instantiated with sets of propositions). But note that agents are not completely free to combine locutions and propositions — for example rules like that above will specify what is possible.

Dialogue history. A *dialogue* can be seen as a non-empty sequence of dialogue moves. We index the set of moves by the step of the dialogue, and denote by d_k a dialogue history of length k .

$$d_k = [m_1, m_2, \dots, m_k]$$

where each m_k is a well-formed dialogue moves as previously defined. We will write the set of all sequences of dialogue moves as D .

Dialogue protocols. Now, as already discussed, not all sequences of dialogues are *legal*. In general, it could be possible to specify, for any dialogue history, the set of dialogue moves that can be uttered. That is, we can think of a protocol as a function π that defines a subset of M that can be uttered at a particular point in a dialogue.

$$\pi : D \mapsto 2^M$$

Typically, it is infeasible to specify the allowed moves for every possible dialogue state that can occur. Often, protocols will instead be specified thanks to simple rules: turn-taking, moves allowed after another move, etc. For instance, for persuasion dialogues we may specify the set of possible responses given in Table 1 (adapted from [Prakken, 2006]).

Also, as we saw earlier, it is possible to specify commitment rules stating conditions upon which moves can be made, thus using commitment stores as (part of) the “dialogue state”. This approach should be familiar to the AI-oriented reader: such locutions can be considered to be a set of speech acts in

| move | possible replies |
|---|---|
| $\langle x, \text{assert}, \phi \rangle$ | $\langle y, \text{accept}, \phi \rangle, \langle y, \text{assert}, \neg\phi \rangle, \langle y, \text{challenge}, \phi \rangle$ |
| $\langle x, \text{challenge}, \phi \rangle$ | $\langle y, \text{argue}, (\Phi, \phi) \rangle, \langle y, \text{retract}, \phi \rangle$ |
| $\langle x, \text{argue}, (\Phi, \phi) \rangle$ | $\langle y, \text{accept}, \psi \rangle (\psi \in \Phi), \langle y, \text{challenge}, \psi \rangle (\psi \in \Phi)$ |
| $\langle x, \text{retract}, \phi \rangle$ | |
| $\langle x, \text{accept}, \phi \rangle$ | |

Table 1. Possible replies in persuasion games (where (Φ, ϕ) is an argument with support Φ and claim ϕ)

the sense of [Austin, 1975; Searle, 1969]. What this means is that although these locutions are merely utterances, they have the properties of actions, in that they change the state of the world. The “state of the world” in this case is the state of the dialogue, and it is changed because the dialogue moves change the state of the commitment stores. Note that the definition of the locutions in terms of conditions and updates rules also provides them with an operational semantics [Plotkin, 1981] — conditions and update rules give pre- and post-conditions for the instantiated locutions, thus describing the transitions in world state that they enable. The public nature of the commitment stores also provide the locutions with a social semantics [Singh, 2000].

At this stage it is useful to inspect a bit more closely, following [Fernández and Endriss, 2007], how these various types of dialogue protocols compare. Do they have different expressive power, for instance (in the sense that they allow more dialogues to be captured)? At the basic level lies *deterministic finite-state automaton* (DFA) (aka. *finite state machines*). After each move, a given set of moves is specified, as in our example in Table 1. It is sufficient to store the last dialogue move in a dialogue state to operate. Now, adding commitment stores (sets of propositions agents are committed to, as in Hamblin-like dialogue games) to this state does not make protocols more expressive than DFA. But they can be much more concise and convenient to specify. On the contrary, using some *stacks* (like questions under discussions, which allows embedded sub-questions) would improve the expressivity of the protocol. Often, in practice, protocols are thus specified based on a combination of the last dialogue move and some commitment stores.

A general remark that can be made about the resulting protocols is that they remain mostly “syntactically regulated”. An alternative approach, advocated in particular by [Prakken, 2005], is to also base the legality of moves on the current dialectical status of some propositions. In particular, in the context of persuasion dialogues, we may allow moves as long as they modify the dialectical status of the main topic of the discussion. We call such protocols *relevance-based*. This requires of course a reasoning machinery to evaluate this dialectical status of argument.

Agents' strategy. The last remaining component differs a bit, in the sense that it is not part of the dialogue game *per se*. However, it is certainly required to understand or automate agents playing such games. How does an agent decide which dialogue move to play? An agent might make its decision with an overall aim in mind. It might want to comply to rationality principles. It might want to end the dialogue quickly. It might want to drag the dialogue out as long as possible [Gabbay and Woods, 2001b; Gabbay and Woods, 2001a]. It might want to “win”, however that is defined for the type of dialogue, if it possibly can. It might want to avoid giving away information [Oren *et al.*, 2006]. The mechanism for making a choice of move that takes this overall aim (or indeed a combination of these aims) into account is a *strategy*.

We can think of the strategy for an agent x as being a function \mathcal{S}_x which takes a dialogue history, the agent's private knowledge base Σ_x , and which returns a subset of dialogue moves. Each agent has its own strategy:

$$\mathcal{S}_x : D \times \Sigma_x \mapsto 2^M$$

More generally, it would be possible to define *probabilistic strategies*, which return probability distributions over the possible dialogue moves of agents. But the strategies we discuss in this chapter are mostly not probabilistic, i.e. they are *deterministic*.

Finally, we say that a strategy is *decisive* if it returns a single move. We note also that nothing at this stage guarantees that the agent's strategy complies with a given protocol. This is known as the *conformance problem*. One simple way to guarantee this is to filter out illegal moves returned by the strategy. We shall denote \mathcal{S}_x^π the strategy of agent x made compliant for protocol π .

As for protocols, we see that specifying a strategy can be tedious, as it may require specifying decisions for each possible dialogue history. When the strategy remains the same at each stage of the dialogue, we talk of a *stationary strategy*, and can simply drop the reference to the dialogue history. More often, strategies may only depend on the previous move of the other agent, and thus can be specified for each possible locution uttered by the other agent.

In the rest of this chapter, we will sometimes use simple *ordered strategies*, that is, a simple preference ranking of the dialogue moves of the agents, assuming those moves comply with rationality constraints. For example, the following simplistic strategy corresponds to an agent x which would, in any circumstance during the dialogue, prefer (if it is allowed by the protocol π): first to challenge a proposition of the other agent, then to assert some proposition, then concede, and finally to retract some proposition.

$$\mathcal{S}_x^\pi = \begin{cases} 1 : \langle x, \text{challenge}, \phi \rangle \\ 2 : \langle x, \text{assert}, \phi \rangle \\ 3 : \langle x, \text{accept}, \phi \rangle \\ 4 : \langle x, \text{retract}, \phi \rangle \end{cases}$$

Suppose the previous move was an assert by the other agent, and that only challenging or accepting are possible replies. In that case the agent would

prefer to challenge. Of course, there could be many dialogue moves satisfying a given priority level (e.g. the agent could have many propositions to possibly assert). In that case, further selection functions would be required to make the strategy decisive. There a lot more to say about strategies, and in particular we may want to take advantage of some information we may have on the other agent(s), in other words consider an *opponent model* on top of the sole dialogue history as discussed. We devote a full section to this recent and lively research question (Section 5.2).

2.4 Evaluation of dialogues

Now comes a difficult question. How can we evaluate the quality of the resulting dialogues? We start with the most obvious properties that one might require.

- *termination*: requires that the dialogue does not run forever. A stronger guarantee is that the dialogue is *deadlock-free*, that is, that there is always a legal move to play at any state of the dialogue. Even these seemingly very basic properties are sometimes difficult to guarantee.
- *successfulness*: whether the goal of the dialogue, or the goal of the participants is fully/partially attained. This criterion can be a simple boolean test, but much finer-grained approaches are possible—think of outcomes in a negotiation for instance.
- *efficiency*: what was the length of the dialogue? Would it have been possible to reach an equally good outcome in fewer communication steps?
- *comparison to full merging of knowledge, or variants of the protocol*: it is often very useful to compare the protocol with (even idealized) alternative solutions, like simply merging all the knowledge of agents (that is, the outcome an omniscient agent would obtain). Or, on the contrary, to a less expressive version of the protocol (for instance, a protocol not allowing the exchange of arguments).
- *relevance of dialogue moves*: was the dialogue coherent overall, were all the moves relevant? These are important aspects to consider when humans are involved in the dialogue system. It is possible to design heuristics evaluating how “close” to the topic under discussion is each move, see for example the work of Rosenfeld and Kraus introducing several heuristics for that purpose in the context of dialogues for persuasion [Rosenfeld and Kraus, 2016a].

There are also criteria that are more difficult to pin down formally, but may still be very important in some contexts—for instance, how cognitively difficult are the reasoning tasks involved in the dialogue? Depending on the application, we may thus require expert feedback to evaluate some of these criteria, which may be difficult to assess automatically.

2.5 Artificial agents playing dialectical games

We now give an example of a specific dialogue game, showing how the components presented above can come together. The game we present is taken from [Parsons *et al.*, 2003], which contains a set of basic dialogue games that are instantiations of the Walton and Krabbe concepts of information-seeking, inquiry and persuasion, dialogue games that are refinements⁸ of the more general dialogues introduced in [Amgoud *et al.*, 2000a; Amgoud *et al.*, 2000b], and so are in the Hamblin/MacKenzie tradition. In fact, this dialogue is a further refinement of the persuasion dialogue from [Parsons *et al.*, 2003]⁹.

As before, we assume that dialogues take place between two agents, and we refer to these with the variables x and y (where $x \neq y$). Each agent has a private knowledge base, Σ_x and Σ_y respectively, containing their beliefs. We assume that agents reason using some form of structured argumentation system, and so the knowledge base is a collection of formulae in some logical language. We aren't going to impose any conditions on the knowledge base other than that the underlying language contains at least a set of ground literals, or propositions, and the only condition that we will place on the argumentation system is that it can construct arguments from the elements of the knowledge base, and can apply the grounded semantics. (The choice of the grounded semantics is somewhat arbitrary, though were we going to examine the properties of this dialogue system, adopting a semantics that is guaranteed to have an extension, and to only ever have one extension, would be helpful.) Thus, an agent x can construct arguments using¹⁰:

$$\Sigma_x \cup \text{CS}_{x,k} \cup \text{CS}_{y,k}$$

Here we will assume that there are no degrees of belief to be accounted for. That is, we assume that both agents believe all the elements of Σ_x and Σ_y , and hence $\text{CS}_{x,k}$ and $\text{CS}_{y,k}$ equally. This is, of course, a big assumption, and it is easy to incorporate different levels of belief [Amgoud *et al.*, 2000a; Amgoud *et al.*, 2000b], to relate belief an agent has in a proposition to the level of trust it has in the agent who utters it [Parsons *et al.*, 2011], or to allow

⁸In the sense of being refined down to the minimal dialogue games that could be considered instances of the relevant type.

⁹The dialogue games in [Parsons *et al.*, 2003] tied the game to a specific underlying argumentation system, so the version here is more general, and allowed for a range of agent *attitudes* that specified the conditions under which agents could make utterances. The game defined here allows just one attitude and ties that to the use of the grounded semantics, so in that sense it is more specific than that in [Parsons *et al.*, 2003]. The description here also separates protocol and strategy, providing a much more declarative version of the dialogue game, and adopts the *argue* locution.

¹⁰Since, as we will see, what gets placed into an agent's commitment store are the conclusions of arguments, these will include elements that are not present in the relevant knowledge base. As a result, we can think of a commitment store as being an extension of the corresponding knowledge base rather than a subset, and so consider an agent to have access to the union of the commitment store and that knowledge base though, anything in the commitment store can be inferred from the corresponding knowledge base.

| move | conditions |
|--|---|
| $\langle x, \text{assert}, p \rangle$ | $p = \text{Conc}(A)$, s. t. $A \in E_{\mathcal{GR}}(\text{Ar}(\Sigma_x \cup \text{CS}_{x,k} \cup \text{CS}_{y,k}))$ |
| $\langle x, \text{assert}, S \rangle$ | $\forall s \in S, s = \text{Conc}(A)$, s. t. $A \in E_{\mathcal{GR}}(\text{Ar}(\Sigma_x \cup \text{CS}_{x,k} \cup \text{CS}_{y,k}))$ |
| $\langle x, \text{accept}, p \rangle$ | $p = \text{Conc}(A)$, s. t. $A \in E_{\mathcal{GR}}(\text{Ar}(\Sigma_x \cup \text{CS}_{x,k} \cup \text{CS}_{y,k}))$ |
| $\langle x, \text{accept}, S \rangle$ | $\forall s \in S, s = \text{Conc}(A)$, s. t. $A \in E_{\mathcal{GR}}(\text{Ar}(\Sigma_x \cup \text{CS}_{x,k} \cup \text{CS}_{y,k}))$ |
| $\langle x, \text{challenge}, p \rangle$ | |

Table 2. Conditions in persuasion games

agents to discuss the level of belief of some proposition that is being discussed [Amgoud and Parsons, 2001].

The protocol used will be similar to the one described in Table 1, at the sole exception that no retract moves are considered. In addition, moves affect the state of the commitment stores, by basically committing agents to the content of the moves they assert or accept (other moves have no consequences on the commitment stores).

The strategy of agents is first based on *rationality conditions*. Agents abiding by these rationality conditions can assert any proposition that is the conclusion of an argument that is in the grounded extension of that agent’s argumentation framework¹¹. In other words, if agent x can construct an argument from $\Sigma_x \cup \text{CS}_{x,k} \cup \text{CS}_{y,k}$, and that argument is acceptable under the grounded semantics, then its conclusion can be asserted. Similarly, an agent can accept any proposition asserted by another agent, provided that the proposition is the conclusion of an argument that is in the grounded extension of its argumentation framework. As [Parsons *et al.*, 2003] discusses, varying the rationality conditions varies the way that agents behave in a given dialogue game.

Table 2 describes each move as being uttered by x and addressed to y ; since x and y are variables that can refer to either participant (where $x \neq y$) this is sufficient to define the rationality and update rules for both of the participating agents. We use the notation $\text{Ar}(\Sigma)$ to denote the set of all arguments constructed from the set Σ , $\text{Conc}(A)$ to denote the conclusion of the argument A , and $E_{\mathcal{GR}}(\text{Ar})$ to denote the grounded extension of the set of arguments Ar . We frame each move as taking place at step k of the dialogue.

Now in terms of protocol, each move has an *update rule* which specifies how the commitment stores of the agents are modified by the move. (An *assert* places the formula with which it is instantiated into the agent’s commitment store, other locutions do not change the commitment store.) There are also conditions regarding what types of moves can follow other moves, which are in line with those given in Table 1.

How about strategy? As discussed above, in a persuasion dialogue, one party seeks to persuade another party to adopt a belief or point-of-view he or she does not currently hold. We slightly depart from the description given in

¹¹This is same condition as adopted in [Amgoud *et al.*, 2000a; Amgoud *et al.*, 2000b], restated in terms of the Dung semantics.

[Parsons *et al.*, 2003] and describe a possible strategy, which — when combined with its rationality conditions as discussed above — allows one agent, x to try to persuade another, y , to accept a proposition p :

$$\mathcal{S}_x^\pi = \begin{cases} 1 : \langle x, \text{accept}, \phi \rangle \\ 2 : \langle x, \text{challenge}, \phi \rangle \\ 3 : \langle x, \text{assert}, \neg\phi \rangle \end{cases}$$

An interesting case emerges with the challenge move. Indeed, as procedurally described in [Parsons *et al.*, 2003], the behaviour of the agent should be that “If y has challenged p , then x asserts S , the premises of its argument for p ”. Then agent y should inspect in turn each premise. But as the premises of an argument can in turn be challenged, we see that embedded dialogue structures (as in the Paul and Olga example) may occur. To account for this, there are two options:

- (i) either the language allows agent x to assert a *collection* of premises in a single move. In that case, the protocol should allow to accept or challenge any proposition in the commitment store of the other agent (thus departing from the simple last move move-reply protocol);
- (ii) or the language only allows **assert** moves involving a single proposition at-a-time. In that case, the protocol should be equipped to keep track of propositions challenged, so as to make sure that agents are enforced to respond appropriately to those challenges.

Note also that in both cases, enforcing agents to address the latest **challenge** or **assert** raised requires the use of stack structures, instead of mere sets.

If at any point an agent cannot make the indicated move, it has to concede the dialogue game. An agent also concedes the game if at any point there are no propositions made by the other agent that it hasn’t accepted. These rules give two conditions under which the dialogue will terminate with an agent conceding. However, on their own, they do not mean that the protocol will terminate. One agent could, for example, keep make the same assertion repeatedly¹². However, termination can be guaranteed by adding a rule specifying that **assert** is not allowed if the proposition is already present in the CS, which is the case in the Hamblin’s protocol, or, by adding a rule specifying that no move may be repeated, as in [Parsons *et al.*, 2003]¹³.

While this makes a useful example, it is clearly extremely simple. Agent x ’s method for persuasion is to assert p , the conclusion that it wants y to accept.

¹²A behaviour that any parent will recognise.

¹³With this protocol, if we rely on Hamblin’s rule, then, together with the rule about conceding if there is no legal move, it is the case that an agent can win by challenging the same proposition multiple times. When its opponent runs out of alternative justifications the opponent will have nothing new to **assert** and will have to concede. If we adopt the rule that repeating the same move is not possible in place of Hamblin’s rule, then the first agent to run out of new moves will have to concede, and that seems a more equitable solution in this case.

Agent y accepts if it can, which it can only do if has no argument against it. If there is a counter-argument, y can assert that, and x may find that more persuasive than its initial argument (particularly if we allow for preferences over beliefs, and hence arguments). If an argument is *challenged*, the response is to put forward the premises for p , and these are considered in turn. If this does not convince the recipient (as it well might not if the recipient does not have the necessary rules in its knowledge base¹⁴), then the dialogue will end inconclusively.

In [Parsons *et al.*, 2003] the reason for looking at such a simple protocol was to be able to establish some baseline results for dialogues on which more complex dialogue systems could be built. As above, we can show termination, and it is straightforward to obtain results about the conditions under which one participant will be persuaded by another. However, we don't have to look far to find examples of persuasion dialogues that cannot be captured by it. Indeed, we have already seen one — the Paul and Olga example from Section 1. The first three first moves of the Paul and Olga dialogue can be readily captured by our protocol, the assertion of a statement about Paul's car, a question from Olga, and the assertion of the supporting argument that the car has an airbag¹⁵. However, the fourth move sees Olga both conceding, and making a counter-argument, and in the fifth, Paul does the same. Then in the sixth move, Olga returns to an earlier choice and proposes a new alternative counter-argument. As Prakken [2006] points out, in the sixth move Olga also postpones replying to Paul's counter-argument (by providing another attack on his argument for safety). This kind of flexibility requires different techniques to maintain dialogue coherence, and the approaches in [Prakken, 2001b; Prakken, 2005] aim to provide exactly this.

2.6 A final word on the basics

As mentioned earlier, the work discussed in the chapter as a whole only scratches the surface of existing work on argumentation-based dialogue. The same goes for this section. It has covered the core ideas of argumentation-based dialogue, and it has illustrated many of them using a simple dialogue game. But there are many other aspects of argumentation-based dialogue that it has not covered. Some of these will be discussed later in the chapter, but there are a couple of specific elements that we'll mention here.

First, the argumentation game described here is monotonic in terms of as-

¹⁴The limitation depends on the argumentation system used by the agents. In [Parsons *et al.*, 2003] the underlying system made no distinction between facts (premises or axioms) and rules, so p could be a rule, in the form of a material implication. However, in systems like ASPIC+ [Modgil and Prakken, 2013], where rules cannot be the conclusion of an argument, what can be asserted is more limited.

¹⁵Note that this argument is an enthymeme — it does not include the connection between airbag and safety that we might write as “a car that has an airbag is safe”. Such an enthymeme can be asserted in our example protocol whereas the full argument in an ASPIC+-like system would not because the rule “a car that has an airbag is safe” can't be the conclusion of an argument.

sertions. Once a proposition has been asserted, it remains so. Taken together with the idea that a proposition that has been asserted, and hence is placed in a commitment store, will be defended by the relevant agent, this monotonicity does not make sense for many dialogue games. Even in the simple persuasion dialogue above, x can assert p and then end up accepting $\neg p$, provided that there is a suitably persuasive argument for it. In such a case it makes no sense to continue to commit to p . As a result, many argumentation-based dialogue systems allow for retraction of prior assertions.

Second, both our discussion and our example, has been for a single kind of dialogue. As we noted when discussing [Walton and Krabbe, 1996], it is possible to imagine dialogues that are combinations of other types. A canonical example is that of a negotiation, for example a negotiation around the purchase of an item, in which one participant transitions into a persuasion dialogue aimed at convincing the other participant that it really needs some feature of the item that justifies a higher price. Now, one could define such combined dialogues — the locutions, protocols, strategies and so on — in addition to the dialogues of which they are composed. One might also look at how dialogues can be combined. Both lines of work have been pursued, with [McBurney and Parsons, 2002] being the first work to explicitly consider the requirements of dialogues which are themselves made up of other dialogues.

Finally, we should mention some seminal work that does not fit directly into the Hamlin/Mackenzie line that we have been describing above. A major contribution here is Prakken’s general framework for dialogue [Prakken, 2005], which we mentioned above. An early version of this general approach can be found in [Prakken, 2000], and the basic mechanism, albeit focused on a specific legal dialogue game is in [Prakken and Sartor, 1998]. The key characteristic of this approach is that it is more general than the specific dialogue games that we have been discussing, making it possible to produce results that can evaluate families of dialogue games. [Prakken, 2006] starts from this viewpoint, and indeed in some sense extends it, by identifying and contrasting the key features of many different formal persuasion dialogues. Two of the dialogue systems reviewed in [Prakken, 2006] should also be mentioned. [Walton and Krabbe, 1996] not only classifies different kinds of dialogue, but defines and discusses a dialogue system PPD that is an early example of a general model of persuasion. Finally, Gordon’s “pleadings game” [Gordon, 1994], is an early example of a formal dialogue game, one which captures features of US civil procedure, and which includes many of the elements which have now become standard.

3 Representing dialogues

Given the diversity of dialogue or dialectical games proposed in the literature (as briefly discussed in Section 2), it may seem challenging to come up with a common representation scheme. However, such a scheme would in principle bring many benefits, the first one being to facilitate the design of dialogue games, as well as making it possible to compare them and to easily share them

among different users. But another potential interest is to make these games amenable to formal verification and analysis of properties. In this section we discuss two proposals which illustrate these two aspects: the dialogue specification language of [Wells and Reed, 2012], and the use of *executable logic* as a formal language, as advocated in [Black and Hunter, 2012a].

3.1 Dialogue specification language

Before setting up their framework, [Wells and Reed, 2012] describe in great detail the landscape of extant dialectal games. They exhibit relations between games, and provide an historical perspective which is interesting in its own right. Based on this analysis, they identified a set of features (See Table 3 for details) which form the basis of their *Dialogue Game Description Language* (DGDL). This is a specific instance of a *Domain Specific Language*, that is “a small language developed to be both concise and powerful” [Wells and Reed, 2012].

| | |
|--------------------|---|
| Moves per turn | Either a single, multiple, or a defined maximum number of moves per turn |
| Turn Organization | Strict progression |
| Dialogue magnitude | The maximum number of turns that the dialogue allows |
| Move types | Range of available locutions or performatives |
| Move content | The statements, propositions, variables, tokens, or collections thereof that are moved |
| Openers | A description of the locutional form of the move, eg. “Is it the case that... ” |
| Stores | Collections of in-game artifacts, organized as sets, stacks, or queues |
| Store contents | In addition to move content, stores can contain locutions or arguments |
| Store visibility | Whether a given store is public or private |
| Move legality | The formulation of conditions that must be satisfied for the move to be legal. Inspection of previous moves; store contents; role occupation; store magnitude; store comparisons; the length of the dialogue; correspondence to a given scheme; relation between content elements; form in which the content is presented |
| Move effects | The effect of a successfully played move upon the dialogue and its components. Prescription of mandated responses; operations on stores; update to status of a game or system; assignment of role |
| Participants | The number and identity of players that the game supports |
| Roles | The roles that players can occupy at various stages of a game |
| Rules | Non-move specific rules that can alter the game as a function of game state rather than as a result of the particular move that has been played |

Table 3. Features of dialogue games (from [Wells and Reed, 2012])

They provide a formal specification, under an Extended-Backus Normal Form, and show how several games can be captured in this setting, including DC [Mackenzie, 1979b], which we have already mentioned. Once a game is described, it then becomes possible to proceed to verify whether the description is syntactically correct, for instance.

As dialogue games are also games, a natural question is whether the more

general *Game Description Language*¹⁶ could also be used to describe dialogue games. Until recently, this language was rather restricted in the sense that incomplete information games could not be handled, but recent developments [Thielscher, 2010] mean that it could in principle be possible to use the Game Description Language to reason about the strategical aspects of agents' involved in dialogue games, even in situations where knowledge about the opponent is incomplete (see Section 5.2 for a more detailed treatment of these questions)It would thus be interesting to investigate using this general language to capture and reason about dialogue games, although there is the risk that we might lose the domain specific facilities DGDL offers.

3.2 Representing dialogues with logic

In the early 2000s, a stream of papers [Sadri *et al.*, 2001; Torroni, 2002; Endriss *et al.*, 2004] studied communication among agents in the context of *computational logic*. In general, protocols were interpreted as *constraints* that were integrated in the agent's reasoning.

A related but different approach was taken by [Black and Hunter, 2012a], who built a logic-based language aiming at representing dialogical argumentation systems. More precisely, they proposed viewing these systems as operating on a dialogue state, and make use of different modal operators to account for the modification, either of the private or public state: $\oplus\alpha$ (resp. $\ominus\alpha$) adds the literal to (resp. removes the literal from) the *private* state, while $\boxplus\alpha$ (resp. $\boxminus\alpha$) adds the literal α to (resp. removes the literal from) the public state. The designer is then free to define and use predicates which can be used either in the private or public state. For instance, assuming a predicate $b(\phi)$ to represent that the agent believes ϕ and a predicate $t(x)$ to represent that it is x 's turn in the dialogue, a simple rule like the following could be stated:

$$b(\phi) \wedge t(x) \Rightarrow \boxplus\text{claim}(\phi) \wedge \boxminus t(x) \wedge \boxplus t(y)$$

In words, if the agent believes ϕ and if it is her turn to speak, then she might claim ϕ , which would also release the turn token to the other agent. Following the discussion in Section 2, we see that such a rule integrates both protocol and strategical (under the form of a rationality condition) aspects.

With a few more simple rules (in addition to the rule above) we can capture the first part of the dialogue between Paul and Olga (Section 1). We assume the following predicates: $\text{public}(\psi)$ means that ψ is part of the public state; $\text{argue}(a, \phi)$ means that the agent can construct an argument a whose claim is ϕ .

$$\text{public}(\text{claim}(\phi)) \wedge \neg\text{argue}(a, \phi) \wedge t(x) \Rightarrow \boxplus\text{why}(\phi) \wedge \boxminus t(x) \wedge \boxplus t(y)$$

$$\text{public}(\text{why}(\phi)) \wedge \text{argue}(a, \phi) \wedge t(x) \Rightarrow \boxplus\text{assert}(a) \wedge \boxminus t(x) \wedge \boxplus t(y)$$

Referring back to the Paul and Olga example (Section 1) (assuming that Paul can, from his private beliefs, construct an argument for why the car is safe based

¹⁶<http://games.stanford.edu/games/gdl.html>

on the fact that it has an airbag, and that Olga cannot construct any argument for why the car is safe) these rules will generate the following behaviour.

| n | x | action on dialogue state |
|-----|------|--|
| 0 | Paul | $\boxplus\text{claim}(\text{safe}), \boxplus(t(\text{Olga}), \boxminus(t(\text{Paul})))$ |
| 1 | Olga | $\boxplus\text{why}(\text{safe}), \boxplus(t(\text{Paul}), \boxminus(t(\text{Olga})))$ |
| 2 | Paul | $\boxplus\text{assert}(\{\text{airbag} \implies \text{safe}, \text{airbag}\}, \text{safe}), \boxplus(t(\text{Olga}), \boxminus(t(\text{Paul})))$ |

When dialogical systems are *finite* (that is, when for each action rule, there is a finite number of groundings of the rule), then a translation to finite state machines can be obtained, thus allowing to prove properties (like termination, for instance).

4 Example: Value-based deliberation

In this section, we present Black and Atkinson’s dialogue system for deliberation [Black and Atkinson, 2011a; Black and Atkinson, 2011b], which allows two agents to agree on an action to perform to achieve some shared goal and formally delineates the public aspects of the dialogue system (such as protocol and moves) from the individual participants’ strategies. The participating agents each aim to reach an agreement, but individually they may wish to influence the outcome of the dialogue to better suit their preferences. This type of dialogue therefore sits somewhere between persuasion dialogues (which may be zero sum) and entirely cooperative inquiry dialogues (where agents seek to establish whether a claim can be justified), as is typically the case with real life encounters. Consider, for example, a situation where colleagues at a conference would all like to go to a restaurant for dinner together. Each individual has a set of preferences that they wish to be satisfied by the group’s decision about which restaurant to visit. For example, some may require vegetarian options, some may not want to travel far, and some may want to avoid expensive restaurants. In such a scenario, the participants each have their own preferences but are committed to finding an outcome they can all agree to.

4.1 Practical reasoning

The participating agents use the popular *argumentation scheme and critical question* approach [Walton and Krabbe, 1996; Macagno *et al.*, 2018] in order to reason about how to act. Argumentation schemes are stereotypical patterns of reasoning, which are used as presumptive justification for generating arguments. Each scheme has associated with it a set of critical questions, which allow one to identify potential attacks on an argument generated by the scheme. Specifically, the agents make use of the *practical reasoning argument scheme* [Atkinson and Bench-Capon, 2007]:

In the current circumstances R,
we should perform action A,
which will result in new circumstances S,

which will achieve goal G,
 which will promote value V.

The set of characteristic critical questions that are associated with the scheme can be used to identify challenges to proposals for action that instantiate the scheme. An unfavourable answer to a critical question will identify a potential flaw in the argument. Since the scheme makes use of what are termed as ‘values’, this caters for arguments based on subjective preferences as well as more objective facts. Such values represent qualitative interests that an agent wishes (or does not wish) to uphold by realising the goal stated.

Each agent has knowledge about the state of the world, the preconditions and effects of actions, and about which values are promoted or demoted by the transition from one state to another (as caused by an action). This knowledge (which is represented as a *value-based transition system* [Black and Atkinson, 2011b], the definition of which we omit here for brevity) can be used to instantiate the practical reasoning argument scheme in order to construct arguments for (or against) actions to achieve a particular goal because they promote (or demote) a particular value.

Definition 4.1 *An argument constructed by an agent x is a 4-tuple $a = \langle act, p, v, s \rangle$ where:*

- *act is an action;*
- *p is a goal;*
- *v is a value;*
- *$s \in \{+, -\}$;*
- *$s = +$ iff act is an action that will achieve goal p and will promote value v;*
- *$s = -$ iff act is an action that will achieve goal p but will demote value v.*

For any argument $a = \langle act, p, v, s \rangle$: $Act(a) = act$; $Goal(a) = p$; $Val(a) = v$; $Sign(a) = s$.

*If $Sign(a) = +$, then a is a **positive argument for** action $Act(a)$.*

*If $Sign(a) = -$, then a is a **negative argument against** action $Act(a)$.*

The set of all arguments an agent x can construct is denoted $Args^x$; the set of all arguments for a particular goal p that an agent can construct is $Args_p^x = \{a \in Args^x \mid Goal(a) = p\}$.

The set of values for a set of arguments S is $Vals(S) = \{v \mid a \in S \text{ and } Val(a) = v\}$.

Given a particular argument for an action, one can then (as mentioned above) identify attacks on that action by posing the various critical questions associated with the practical reasoning argument scheme. This reasoning (through posing of the different critical questions) is split into three stages [Atkinson and Bench-Capon, 2007]: *problem formulation*, where the agents decide on the facts and values relevant to the particular situation under consideration; *epistemic reasoning*, where the agents determine the current situation with respect to the structure formed at the previous stage; and *action selection*, where the agents develop, and evaluate, arguments and counter arguments

about what to do. The dialogue system we present here deals only with the action selection stage. It assumes that the agents' problem formulation and epistemic reasoning are sound and that they are in agreement on the output of these stages (this agreement could itself be the product of another dialogue, see for example [Black and Atkinson, 2009]). The critical questions associated with the action selection stage (numbered as in [Atkinson and Bench-Capon, 2007]) are:

- CQ5:** Are there alternative ways of realising the same consequences?
- CQ6:** Are there alternative ways of realising the same goal?
- CQ7:** Are there alternative ways of promoting the same value?
- CQ8:** Does doing the action have a side effect which demotes the value?
- CQ9:** Does doing the action have a side effect which demotes some other value?
- CQ10:** Does doing the action promote some other value?
- CQ11:** Does doing the action preclude some other action which would promote some other value?

As the focus of the dialogue is to agree to an action that achieves the *goal*, the incidental consequences (CQ5) and other potentially precluded actions (CQ11) are not relevant. The participating agents thus use CQ6–CQ10 to identify attacks on proposed arguments for action. Answers to these questions provide arguments for and against different actions to achieve a particular goal, where each argument is associated with a motivating value. These arguments and the attacks between them (determined by the critical questions) can be represented in a value-based argumentation framework [Bench-Capon, 2002] (see also Chapter 5 of this volume), an extension of the abstract argumentation frameworks of Dung [1995]. In a Dung argumentation framework, an argument is admissible with respect to a set of arguments S if its attackers are attacked by some argument in S , and no argument in S attacks an argument in S . In a value-based argumentation framework, an argument succeeds in defeating an argument it attacks only if its value is ranked higher than (if the attack is symmetric) or at least as high as (if the attack is asymmetric) the value of the argument attacked. A particular ordering of the values is characterised as an **audience**; this represents an individual agent's preferences over the values. Arguments in a value-based argumentation framework are admissible with respect to an audience A and a set of arguments S if they are admissible with respect to S in the Dung argumentation framework that results from removing all the attacks that are unsuccessful given the audience A . A maximal admissible set of a value-based argumentation framework is a **preferred extension**.

Value-based argumentation frameworks are commonly considered at an abstract level. The following definition gives a particular instantiation of value-based argumentation frameworks that captures the attack relation between

arguments that is generated from the critical questions. Condition (1) of the following attack relation allows for CQ8 and CQ9; condition (2) allows for CQ10¹⁷; condition (3) allows for CQ6 and CQ7. Note that attacks generated by condition (1) are not symmetrical, whilst those generated by conditions (2) and (3) are.

Definition 4.2 *An instantiated value-based argumentation framework (iVAF) is defined by a tuple $\langle Ar, att \rangle$ such that Ar is a finite set of arguments and $att \subset Ar \times Ar$ is the **attack relation**. A pair $(a, b) \in att$ is referred to as “ a attacks b ” or “ b is attacked by a ”. For two arguments $a = \langle act, p, v, s \rangle$, $a' = \langle act', p', v', s' \rangle \in Ar$, $(a, a') \in att$ iff $p = p'$ and either:*

- (1): $act = act'$, $s = -$ and $s' = +$; or
- (2): $act = act'$, $v \neq v'$ and $s = s' = +$; or
- (3): $act \neq act'$ and $s = s' = +$.

An **audience** for an agent x over the set of values V is a binary relation $\mathcal{R}^x \subset V \times V$ that defines a total order over V . An argument a is **preferred to** the argument b under the audience \mathcal{R}^x , denoted $a \succ_x b$, iff $(\text{Val}(a), \text{Val}(b)) \in \mathcal{R}^x$. If \mathcal{R}^x is an audience over the values V for the iVAF $\langle Ar, att \rangle$, then $\text{Vals}(Ar) \subseteq V$.

Given an iVAF and a particular agent’s audience, acceptability of an argument is determined as follows. Note that if an attack is symmetric, then an attack only succeeds in defeat if the attacker is more preferred than the argument being attacked; however, as in [Bench-Capon, 2002], if an attack is asymmetric, then an attack succeeds in defeat if the attacker is at least as preferred at the argument being attacked.

Definition 4.3 *Let \mathcal{R}^x be an audience and let $\langle Ar, att \rangle$ be an iVAF.*

*For $(a, b) \in att$ such that $(b, a) \notin att$, a **defeats** b under \mathcal{R}^x iff $b \not\succeq_x a$.*

*For $(a, b) \in att$ such that $(b, a) \in att$, a **defeats** b under \mathcal{R}^x iff $a \succ_x b$.*

*An argument $a \in Ar$ is **acceptable w.r.t** S under \mathcal{R}^x ($S \subseteq Ar$) iff: for every $b \in Ar$ that defeats a under \mathcal{R}^x , there is some $A_k \in S$ that defeats b under \mathcal{R}^x .*

*A subset S of Ar is **conflict-free** under \mathcal{R}^x iff: no argument $a \in S$ defeats another argument $b \in S$ under \mathcal{R}^x .*

*A subset S of Ar is **admissible** under \mathcal{R}^x iff: S is conflict-free under \mathcal{R}^x and every $a \in S$ is acceptable w.r.t S under \mathcal{R}^x .*

*A subset S of Ar is a **preferred extension** under \mathcal{R}^x iff it is a maximal admissible set under \mathcal{R}^x .*

*An argument A is **acceptable** in the iVAF $\langle Ar, att \rangle$ under audience \mathcal{R}^x iff there is some preferred extension containing it.*

¹⁷It may seem counter intuitive that CQ10 generates attacks between arguments for the same action. Although such arguments do not dispute the action that should be performed, they do dispute the reasons as to why. Where there are two arguments proposed for the same action but each is based upon different values, an agent may only accept the argument based on one of the values. Hence such arguments are seen to be in conflict.

| Move | Format |
|---------------|--|
| <i>open</i> | $\langle x, \text{open}, \gamma \rangle$ |
| <i>assert</i> | $\langle x, \text{assert}, a \rangle$ |
| <i>agree</i> | $\langle x, \text{agree}, act \rangle$ |
| <i>close</i> | $\langle x, \text{close}, p \rangle$ |

Table 4. Format for moves used in deliberation dialogues: p is a goal; act is an action; a is an argument; $x \in I$ is an agent identifier.

The mechanism defined in this section allows an agent to determine attacks between arguments for and against actions. It can then use its preference ordering over the motivating values (i.e., its audience) to determine the acceptability of those arguments.

4.2 Dialogue system

The system we present here [Black and Atkinson, 2011a; Black and Atkinson, 2011b] assumes exactly two participating agents, each with its own identifier taken from the set $I = \{1, 2\}$. Each participant takes it in turn to make a move to the other participant. Participants are referred to using the variables x and y such that: x is 1 if and only if y is 2; x is 2 if and only if y is 1.

The format for moves used in deliberation dialogues is shown in Table 4, and the set of all moves meeting the format defined in Table 4 is denoted M . Also, $\text{Sender} : M \mapsto I$ is a function such that $\text{Sender}(\langle Agent, Type, Content \rangle) = Agent$.

An *open* move $\langle x, \text{open}, \gamma \rangle$ opens a dialogue to agree on an action to achieve the goal γ ; an *assert* move $\langle x, \text{assert}, a \rangle$ asserts an argument a for or against an action to achieve a goal that is the topic of the dialogue; an *agree* move $\langle x, \text{agree}, act \rangle$ indicates that x agrees to performing action act to achieve the topic; a *close* move $\langle x, \text{close}, \gamma \rangle$ indicates that x wishes to end the dialogue. All dialogues start with an open move. In order to terminate a dialogue, either: two close moves must appear one immediately after the other in the sequence, in which case the dialogue is unsuccessful and no agreement is reached; or two moves agreeing to the same action must appear one immediately after the other in the sequence, in which case the dialogue is successful and the agents have found an action they can both agree on.

Definition 4.4 A **dialogue**, denoted d_t , is a sequence of moves $[m_1, \dots, m_t]$ where the following conditions hold:

- m_1 is a move of the form $\langle x, \text{open}, \gamma \rangle$;
- for all $1 \leq s \leq t$, $\text{Sender}(m_s) \in I$;
- for all $1 \leq s < t$, $\text{Sender}(m_s) \neq \text{Sender}(m_{s+1})$ for all $1 \leq s < t$;
- for all $1 < s \leq t$, if $m_{s-1} = \langle x, \text{close}, \gamma \rangle$ and $m_s = \langle y, \text{close}, \gamma \rangle$, then $s = t$ and d_t is **unsuccessfully terminated**;
- for all $1 < s \leq t$, if $m_{s-1} = \langle x, \text{agree}, act \rangle$ and $m_s = \langle y, \text{agree}, act \rangle$, then

$s = t$ and d_t is **successfully terminated with outcome** act .

The **topic** of the dialogue d_t is returned by $\text{Topic}(d_t) = \gamma$.

In order to determine the actions it finds agreeable, an agent considers both the private knowledge it has about actions (their preconditions, effects and the values they promote or demote) and the arguments put forward by the other agent during the dialogue. This is represented in what is called the agent's *dialogue iVAF*, which is the iVAF constructed from the union of the arguments the agent can construct from its own knowledge about actions and the arguments that have been asserted by the other agent¹⁸.

Definition 4.5 *A dialogue iVAF for an agent x participating in a dialogue d_t is denoted $\text{dVAF}(x, d_t)$. If $d_t = [m_1, \dots, m_t]$, then $\text{dVAF}(x, d_t)$ is the iVAF $\langle Ar, att \rangle$ where $Ar = \text{Args}_{\text{Topic}(d_t)}^x \cup \{a \mid \exists m_k = \langle y, \text{assert}, a \rangle (1 \leq k \leq t)\}$.*

An action is *agreeable* to an agent x if and only if there is some argument for that action that is acceptable in x 's dialogue iVAF under the audience that represents x 's preference over values. Note that the set of actions that are agreeable to an agent may change over the course of the dialogue, as new arguments are asserted by the other agent.

Definition 4.6 *An action act is **agreeable** in the iVAF $\langle Ar, att \rangle$ under the audience \mathcal{R}^x iff $\exists a = \langle act, \gamma, v, + \rangle \in Ar$ such that a is acceptable in $\langle Ar, att \rangle$ under \mathcal{R}^x . The set of **all actions that are agreeable to agent x participating in a dialogue d_t** is $\text{AgActs}(x, d_t) = \{a \mid a \text{ is agreeable in } \text{dVAF}(x, d_t) \text{ under } \mathcal{R}^x\}$.*

As already discussed, a protocol specifies the moves that an agent is permitted to make at any point in the dialogue. In this dialogue, it is permissible to assert an argument for an action to achieve the goal that is the topic of the dialogue as long as that argument has not previously been asserted in the dialogue¹⁹. An agent can agree to an action that has been agreed to by the other agent in the preceding move. An agent can also agree to an action that has been proposed by the other participant, unless the agent has previously agreed to that same action and has not asserted any further arguments since then. This is to avoid the situation where an agent keeps repeatedly agreeing to an action that the other agent will not agree to: if an agent makes a move agreeing to an action and the other agent does not wish to also agree to that action, then the first agent must introduce some new argument that may convince the second agent to agree before being able to repeat its agree move. Agents may always make a close move. Note, it is possible to check conformance with the protocol as it only refers to public elements of the dialogue.

¹⁸While this system does not use explicit commitment stores, one could consider that an agent's dialogue iVAF is the iVAF constructed from the union of its own knowledge with the other agent's commitment store.

¹⁹A condition that is similar to those we have seen before.

Definition 4.7 *The deliberation protocol is a function $\pi : \mathcal{D} \mapsto 2^{\mathcal{M}}$. Let d_t be a dialogue ($1 \leq t$) such that $\text{Sender}(m_t) = y$, $x' \in \mathcal{I}$, and $\text{Topic}(d_t) = \gamma$.*

$$\pi(d_t) = P^{\text{ass}}(d_t) \cup P^{\text{ag}}(d_t) \cup \{\langle x, \text{close}, \gamma \rangle\}$$

such that

$$m \in P^{\text{ass}}(d_t) \text{ iff: } m = \langle x, \text{assert}, a \rangle, \text{Goal}(a) = \gamma \text{ and} \\ \neg \exists m_{t'} = \langle x', \text{assert}, a \rangle \text{ where } 1 < t' \leq t$$

$$m \in P^{\text{ag}}(d_t) \text{ iff: } m = \langle x, \text{agree}, act \rangle \text{ and either:} \\ m_t = \langle y, \text{agree}, act \rangle; \text{ otherwise} \\ \exists m_{t'} = \langle x', \text{assert}, a \rangle \text{ where } \text{Act}(a) = act, 1 < t' \leq t \text{ and} \\ \forall t'' \text{ such that } 1 \leq t'' < t, \text{ if } \exists m_{t''} = \langle x, \text{agree}, act \rangle, \\ \text{then } \exists m_{t'''} = \langle x, \text{assert}, a \rangle \text{ where } t'' < t''' \leq t.$$

Black and Atkinson [Black and Atkinson, 2011a] provide an ordered strategy that agents can use to determine which of the permissible moves (returned by the protocol) to make. According to this strategy: if it is permissible to make a move agreeing to an agreeable action, then make such an agree move; else, if it is permissible to assert an argument *for* an *agreeable* action, then assert some such argument; else, if it is permissible to assert an argument *against* an action that is *not agreeable*, then assert some such argument; else make a close move. When the strategy results in a choice of more than one agree or assert move, a Pick function is used to select a specific move; this is not specified but in its simplest form may return an arbitrary move from the input set. By specifying the Pick function to return exactly one move, we ensure that the strategy is decisive.

Definition 4.8 *The (ordered) strategy for an agent x is a function $\mathcal{S}_x : \mathcal{D} \mapsto \mathcal{M}$ as follows.*

$$\mathcal{S}_x(d_t) = \begin{cases} 1 : \text{Pick}(\mathcal{S}_x^{\text{ag}})(d_t) \\ 2 : \text{Pick}(\mathcal{S}_x^{\text{prop}})(d_t) \\ 3 : \text{Pick}(\mathcal{S}_x^{\text{att}})(d_t) \\ 4 : \langle x, \text{close}, \text{Topic}(d_t) \rangle \end{cases}$$

where the choices for the moves are given by the following subsidiary functions (where $\text{Topic}(d_t) = \gamma$):

$$\begin{aligned} \mathcal{S}_x^{\text{ag}}(d_t) &= \{\langle x, \text{agree}, act \rangle \in P^{\text{ag}}(d_t) \mid act \in \text{AgActs}(x, d_t)\} \\ \mathcal{S}_x^{\text{prop}}(d_t) &= \{\langle x, \text{assert}, a \rangle \in P^{\text{ass}}(d_t) \mid a \in \text{Args}_\gamma^x, \text{Act}(a) = act, \text{Sign}(a) = + \text{ and} \\ &\quad act \in \text{AgActs}(x, d_t)\} \\ \mathcal{S}_x^{\text{att}}(d_t) &= \{\langle x, \text{assert}, a \rangle \in P^{\text{ass}}(d_t) \mid a \in \text{Args}_\gamma^x, \text{Act}(a) = act, \text{Sign}(a) = -, \\ &\quad act \notin \text{AgActs}(x, d_t) \text{ and} \\ &\quad \exists m_{t'} = \langle x', \text{assert}, a' \rangle \text{ such that } x' \in \mathcal{I}, \\ &\quad 1 \leq t' \leq t, \text{Act}(a') = act \text{ and } \text{Sign}(a') = +\} \end{aligned}$$

The following section gives an example dialogue generated by two agents each using the strategy defined here.

4.3 Example dialogue

This example (first presented in [Black and Atkinson, 2011a]) involves two agents who share the goal to have dinner together (denoted *din*) and are considering two possible actions that will achieve this goal: go to an Italian restaurant (denoted *it*); go to a Chinese restaurant (denoted *ch*). The relevant values are: *d*, distance to travel; *e1*, agent 1's enjoyment; *e2*, agent 2's enjoyment; and *c*, cost. The agents' audiences (i.e., their preferences over the values) are:

$$\begin{aligned} d \succ_1 e1 \succ_1 c \succ_1 e2 \\ c \succ_2 e2 \succ_2 e1 \succ_2 d \end{aligned}$$

Agent 1 starts the dialogue.

$$m_1 = \langle 1, \text{open}, \text{din} \rangle$$

At this opening stage of the dialogue, the agents have only their private knowledge about actions, their preconditions, effects and the values they promote or demote. This knowledge is shown in Figures 1 and 2, where the nodes represent arguments and are labelled with the action that they are for (or the negation of the action that they are against) and the value that they are motivated by. If a node is labelled with the negation of an action, this denotes that performing that action will demote the value that labels the node; if a node is labelled with an action, this denotes that performing that action will promote the value that labels the node. The arcs represent the attack relation between arguments, and a double circle round a node means that the argument it represents is acceptable to that agent under its audience. For example, we can see from Figure 1 that agent 1 has two arguments it finds acceptable: $\langle \neg ch, din, e1, - \rangle$ says we should not go for a Chinese as this demotes value *e1*; $\langle it, din, d, + \rangle$ says we should go for an Italian as this promotes value *d*.

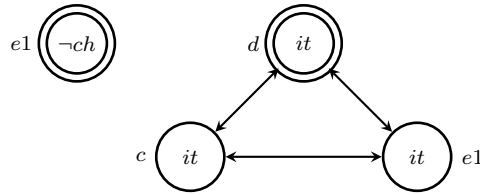
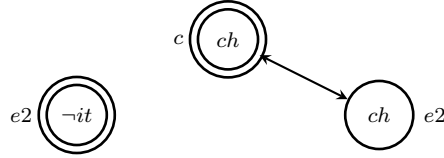


Figure 1. Agent 1's dialogue iVAF at $t = 1$, $dVAF(1, d_1)$.

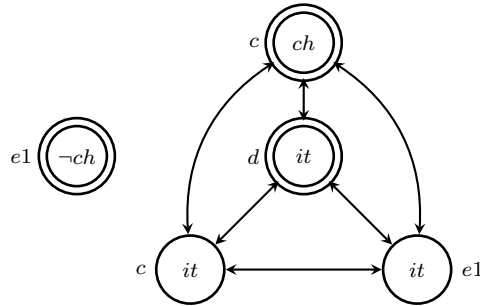
It is now agent 2's turn to make a move. Recall that, according to the strategy, if the protocol permits the agent to agree to an action that it finds acceptable, then it will do so. However, the agent is not permitted (according to the protocol) to make an agree move at this stage, since the previous move was not an agree move and no assert moves have yet been made. If no agree moves can be made, the strategy states that (if the protocol allows it and some such


 Figure 2. Agent 2's dialogue iVAF at $t = 1$, $dVAF(2, d_1)$.

argument exists) the agent should make a proposing assert move and assert a positive argument for an action that it finds agreeable. At this point in the dialogue, there is only one argument *for* an action that is acceptable to agent 2 ($\langle ch, din, c, + \rangle$, see Figure 2), hence ch is the only action that is agreeable to agent 2. Agent 2 must therefore, if it can, assert an argument for going to the Chinese restaurant. There are two such arguments that the Pick function could select: $\langle ch, din, c, + \rangle$, $\langle ch, din, e2, + \rangle$. Let us assume that $\langle ch, din, c, + \rangle$ is selected.

$$m_2 = \langle 2, \text{assert}, \langle ch, din, c, + \rangle \rangle$$

This new argument is added to agent 1's dialogue iVAF, to give $dVAF(1, d_2)$ (Figure 3). Note that there is no change to the arguments that agent 1 finds acceptable, since it does not prefer the value that motivates this new argument to those that motivate the arguments that were already present in its dialogue iVAF.


 Figure 3. Agent 1's dialogue iVAF at $t = 2$, $dVAF(1, d_2)$.

Although agent 2 has proposed going to the Chinese restaurant, this action is not agreeable to agent 1 at this point in the dialogue (as there is no argument for this action that is acceptable in Figure 3). There is, however, an argument for the action it ($\langle it, din, d, + \rangle$) that is acceptable in agent 1's dialogue iVAF (Figure 3), and so going to the Italian restaurant is agreeable to agent 1. Hence, agent 1 must make an assert move proposing an argument for the action it ,

and there are three such arguments that the Pick function can select from: $\langle it, din, d, + \rangle$, $\langle it, din, c, + \rangle$, $\langle it, din, e1, + \rangle$. Let us assume that $\langle it, din, c, + \rangle$ is selected.

$$m_3 = \langle 1, \text{assert}, \langle it, din, c, + \rangle \rangle$$

This new argument is added to 2's dialogue iVAF, to give $dVAF(2, d_3)$ (Figure 4).

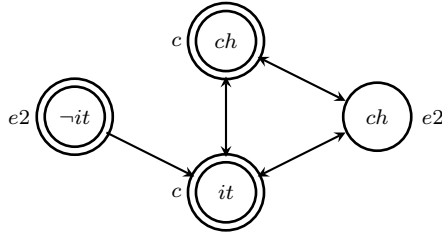


Figure 4. Agent 2's dialogue iVAF at $t = 3$, $dVAF(2, d_3)$.

Going to the Italian restaurant is now agreeable to agent 2 since the new argument introduced promotes the value ranked most highly for agent 2, i.e. cost, and so this argument is acceptable. So, agent 2 agrees to this action.

$$m_4 = \langle 2, \text{agree}, it \rangle$$

Going to the Italian restaurant is also agreeable to agent 1 (as the argument $\langle it, din, d, + \rangle$ is acceptable in its dialogue iVAF, which is still the same as that shown in Fig. 3 as agent 2 has not asserted any new arguments), hence agent 1 also agrees to this action.

$$m_5 = \langle 1, \text{agree}, it \rangle$$

The dialogue has thus terminated successfully and the agents are each happy to agree to go to the Italian restaurant. Note, however, that this action is agreeable to each agent for a different reason. Agent 1 is happy to go to the Italian restaurant as it promotes the value of distance to travel (the Italian restaurant is close by), whereas agent 2 is happy to go to the Italian restaurant as it will promote the value of cost (as it is a cheap restaurant). The agents need not be aware of one another's audience in order to come to an agreement that they are both happy with²⁰.

²⁰Note that the use of values — and the agents' differing preferences over values — adds an extra dimension to the dialogue in comparison with the example given in Section 2, where it is assumed that the agents agree on the defeat relation between arguments. In this case, the agents could also consider a dialogue where they try to reach agreement on a particular audience over the values.

4.4 Evaluation of the dialogue system

Black and Atkinson [2011a] show through theoretical analysis that in their dialogue system:

- all dialogues terminate;
- if a dialogue terminates successfully, then the outcome is agreeable to both participants;
- if a dialogue terminates and there is some action that is agreeable to both participants, then the dialogue will have a successful outcome.

However, they also show the following undesirable property.

- In the case where, if we take the union of the agents' knowledge about actions, there is some action that is agreeable to both of the agents, it is possible that the dialogue will not reach an agreement.

Hence, even if the arguments exist that will enable the agents to reach an agreement, the particular arguments selected for assertion by the Pick function may not allow agreement to be reached. Black and Atkinson [2011b] later explore how an agent may use a model of what it believes to be the other participating agent's preferences over values as a parameter of the Pick function, and how this can lead to better dialogue outcomes.

The behaviour of this dialogue system has also been explored experimentally. [Black and Bentley, 2012] compares the performance of the dialogue system with a simple consensus forming approach, where the agents do not share any knowledge and only agree if there is an action that each finds agreeable given their private knowledge about actions. In this work, random example scenarios (that initialise the agents' private knowledge) are generated, varying the number of arguments, actions and values available to the agents, and the outcome of the dialogue is compared with the outcome of consensus forming. The results show that:

- dialogues are significantly more likely to be successful than consensus forming;
- successful dialogues are more likely with higher numbers of actions and values;
- dialogues produce better quality outcomes than consensus forming (the quality of an outcome is determined by examining whether the action agreed to would be agreeable to one, both or neither of the agents given the union of the knowledge available to each); and
- dialogue length grows exponentially with the number of arguments available.

-
1. Agents report their individual view on the issue to the central authority, which then assigns (privately) each agent to PRO or CON.
 2. The first round starts with the issue on the gameboard and the turn given to CON.
 3. Until a group of agents cannot move:
 - (a) agents independently propose moves to the central authority;
 - (b) the central authority picks the first (or at random) relevant move from the group of agents whose turn is active, update the gameboard, and passes the turn to the other group
-

Figure 5. A multiparty persuasion protocol

The first of these results fits with the prediction in [Loui and Moore, 1997] and the results in [Rahwan *et al.*, 2009] — dialogues can expand the space of possible agreements by making participants aware of tradeoffs.

5 Key challenges

In this section, we discuss what we believe to be some of the key challenges that must be addressed if we are to realise the full potential of argumentation dialogue systems.

5.1 Multiparty dialogues

Relaxing the assumption that there are only two agents in the dialogue poses several challenges. [Dignum and Vreeswijk, 2004] listed a number of issues which are still relevant: the system may be open or closed, the variety of roles that agents may take is much larger, the dialogue may be mediated or not, the turn-taking policy becomes more difficult to define, and termination conditions also can be defined in different ways (simply think of what consensus means when more than two agents are involved). There is thus much less work addressing multiparty dialogues in the literature, the exception being negotiation, where it is not uncommon to have settings involving a large number of agents.

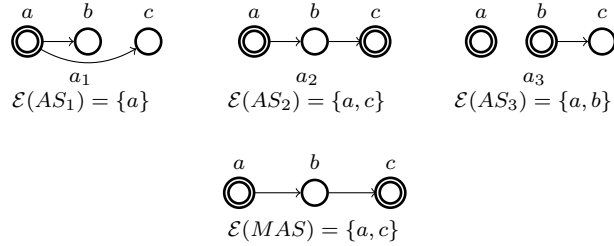
In persuasion this is less common. In what follows, we describe a proposal due to [Bonzon and Maudet, 2011]. We assume that agents share the same set of arguments but may have different opinions regarding attack relations among them (as in Section 4, this may result from an underlying value-based argumentation systems, with agents holding different preference ordering over values). Agents may make claims regarding attacks among arguments.

The protocol is *mediated*, focused on a single *issue*, based on *roles* that agents endorse at the beginning of the dialogue (depending on their stance regarding the issue), and it builds upon the relevance-based protocol idea put forward in

[Prakken, 2005]. It is described in Figure 5.

We see that the turn-taking issue is solved here by letting agents of two different “groups” alternate, with moves being picked by a mediator. Termination occurs when a group is left without any legal move to make. Example 5.1 illustrates, for a given initial situation, the different possible executions of the protocol. It is interesting to observe that the outcome may be different depending on the sequence of moves, and that it may differ from the outcome which we would be obtained from merging the agents’ argumentation systems (recall that a similar issue occurred in Section 4).

Example 5.1 *Let three agents with their argumentation systems, and the following merged argumentation framework:*



The issue of the dialogue is the argument c . We have $CON = \{a_1, a_3\}$, $PRO = \{a_2\}$. We store in RP_x^t the moves played by agent x until turn t . At the beginning, we have $RP_1^0 = RP_2^0 = RP_3^0 = \emptyset$, $AS^0(GB) = \langle \{c\}, \emptyset \rangle$ and $\mathcal{E}(AS^0(GB)) = \{c\}$.

All the possible sequences of moves allowed by the protocol are represented on tree depicted in Figure 6. As it can be seen on Figure 6, any sequence of the protocol stops with a stable gameboard where $\mathcal{E}(AS(GB)) = \{a, b\}$. Note also that some sequences take more moves than others.

5.2 Dialogue strategies

As discussed earlier in this chapter, a dialogue protocol identifies the space of permissible dialogues. At any point in a dialogue, the protocol will typically identify multiple permissible moves that the agent may make. The choice of which permissible move to make is determined by the agent’s *strategy*. In order to improve the likelihood of achieving their dialogue goal, an agent needs to be able to select an effective strategy to apply. Note this section is not intended as an exhaustive review of dialogue strategies, but rather intends to introduce the reader to some of the different approaches that have been taken in the literature and to highlight what we consider to be some of the key challenges for the area. For a different perspective on this, see [Thimm, 2014].

Identifying effective strategies is a challenging problem²¹, as an agent must

²¹Chapter 14 of this volume discusses in depth the complexity of *strategic argumentation problems*, a sub-class of persuasion dialogues where the agents are only able to assert arguments to one another.

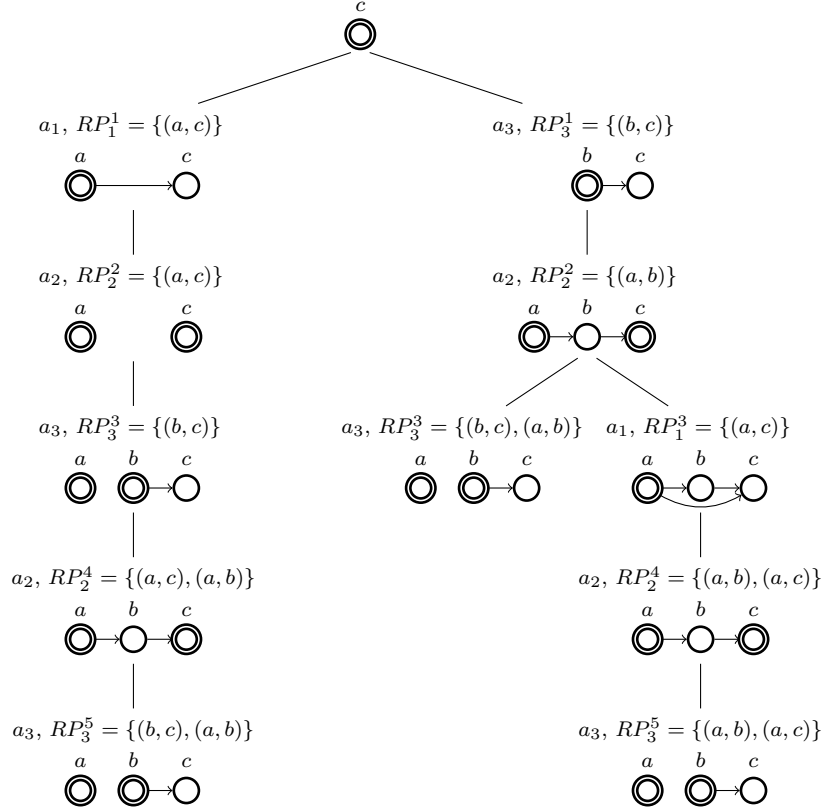


Figure 6. The different sequences of moves for Example 5.1

consider not only its own moves but also the possible responses of the other agent(s) participating in the dialogue, often with only limited or uncertain knowledge of the other agents. Further, in adversarial dialogues, one must take care not to divulge information that the other participant(s) may use to gain an advantage. Consider for example the following dialogue, where a child (C) is trying to persuade their parent (P) that they have done their homework (which requires internet access).

P: Have you finished all your homework?

C: Yes, I finished it this afternoon.

P: Are you sure? I thought I heard you chatting online with your friend all afternoon.

C: I wasn't chatting online! The internet hasn't been working all afternoon.

P: If the internet hasn't been working then you can't have finished your homework!

By divulging that the internet has not been working, the child provides new information to their parent, which their parent then uses to construct an argument that counters their child's claim.

Although challenging, identification of effective strategies is crucial if we are to realise the potential of argument dialogue systems, and there has been significant progress towards this aim in recent years. We can categorise approaches to dialogue strategies into two classes: those where the strategising agent uses a model of some aspect of its interlocutor's (or interlocutors') private state to determine a strategy that is optimised for that particular interlocutor, and those that do not use any such model. We will refer to these two classes as *model-based* and *model-free* respectively.

5.2.1 Model-free approaches

Model-free approaches define a strategy — usually for a particular type of dialogue — that can be applied directly by the strategising agent. Typically, these are *ordered strategies* (to use the terminology from Section 2). For example, the dialogue system [Black and Atkinson, 2011a; Black and Atkinson, 2011b] presented in Section 4 provides a strategy for deliberation dialogues as follows.

- If it is permissible to **agree** to an action that the agent finds to be agreeable, then make some such **agree** move; else
- if it is permissible to **assert** a positive argument *for* an action that the agent finds agreeable, then **assert** some such argument; else
- if it is permissible to **assert** a negative argument *against* an action that the agent does not find agreeable, then **assert** some such argument; else
- make a **pass** move.

Model-free strategies can be straightforward to apply and do not require any knowledge of the other agents participating in the dialogue. While they are not specific to a particular interlocutor, they can be made context-dependent so that priorities may vary depending on the current circumstances of the dialogue, as for instance the topic under discussion (see [Kakas *et al.*, 2005] for a proposal along these lines). A disadvantage of this approach is that it is difficult to know how well the strategy will perform, as the complexity of argument dialogue systems makes it hard to theoretically analyse their general behaviour. Table 5 gives some examples of model-free strategies, noting the type of dialogue for which the strategy is intended, the principle embodied by the strategy, and key properties of the strategy.

5.2.2 Model-based approaches

In contrast to model-free approaches, model-based approaches require a strategy to be computed, taking into account the beliefs the strategist has about

| Citation | Type of dialogue | Strategic principle | Properties |
|--|---|---|--|
| [Amgoud and Parsons, 2001] | Persuasion | Ask questions where possible | Proven that if both agents use this strategy then dialogues won't terminate, but if only one does then termination can be achieved. |
| [Atkinson <i>et al.</i> , 2013] | Persuasion over preferences | Prioritise inquiring about attributes that are (subjectively) more important. | Proven to produce dialogues that terminate in fewest number of steps. |
| [Black and Hunter, 2009] | Inquiry | Exhaustively assert all relevant knowledge. | Proven to be sound and complete in relation to reasoning with union of agents' knowledge. |
| [Black and Atkinson, 2011a; Black and Bentley, 2012] | Deliberation | Agree if possible. Prioritise positive arguments for an action to negative arguments against an action. | If at termination there is some action agreeable to each agent, successful outcome guaranteed. Experimental evaluation shows benefits over consensus forming with no share of information. |
| [Kontarinis <i>et al.</i> , 2014] | Persuasion | Computation of target sets | Prioritises moves which will affect the issue of the dialogue. |
| [Medellin-Gasque <i>et al.</i> , 2013] | Joint planning | Specifies the order in which the elements of the plan are questioned. | Experimental evaluation shows dialogues produced shorter than those produced by random strategy. |
| [Murphy <i>et al.</i> , 2016] | Uni-directional persuasion | Prioritise supporting arguments that are closer to the topic (as defined by the attack relation). | Experimental evaluation shows success more likely than with random strategy. |
| [Oren <i>et al.</i> , 2006] | Persuasion | Prioritise arguments that reveal less information. | Maintains focus of dialogue. Reasoning is not complete. |
| [Parsons <i>et al.</i> , 2003] | Information-seeking, inquiry and persuasion | Agree if possible. This is not stated explicitly, but is implicit in the protocol. | Proven that dialogues terminate. |

Table 5. Examples of model-free strategies from the literature

the other specific dialogue participant(s). These beliefs are often referred to as an *opponent model*, and we will also use this term. Model-based approaches typically involve mapping the problem of finding a strategy that will be effective against one's opponent(s) to an optimisation problem, so that existing techniques can be exploited to identify such a strategy. Techniques employed include automated planning [Black *et al.*, 2014; Black *et al.*, 2017], decision trees [Hadoux and Hunter, 2017], mixed observability Markov decision processes [Hadoux *et al.*, 2015], evolutionary search [Murphy *et al.*, 2018], partially observable Markov decision processes [Rosenfeld and Kraus, 2016b], game theory [Kacpraz *et al.*, 2014], and Markov games [Rach *et al.*, 2018]. To date, these optimising model-based approaches have primarily focused on a restricted class of persuasion dialogues (hence the term *opponent model*) that only involve the exchange of arguments through assertions (in contrast to dialogues that allow a range of moves such as *accept*, *challenge*, *question*, etc.); this speaks to the challenge of efficiently mapping more flexible dialogue systems to a tractable optimisation problem.

Representing and updating an opponent model. Model-based approaches make different assumptions about the knowledge available to the strategist in its opponent model. For example: [Black *et al.*, 2017] assumes that the strategising agent has an uncertain model of the knowledge available to its opponent, captured as a probability distribution over sets of knowledge, but no information about the strategy the opponent employs (seeking strategies that will perform well no matter which moves the opponent plays); [Dimopoulos *et al.*, 2019] builds on their framework of control argumentation systems to distinguish those arguments that are certainly known by the opponent and those which the strategist does not know whether the opponent knows; [Hadoux and Hunter, 2017] requires a prediction to be made about the decision rule used by the opponent to determine its strategy, and explore how errors in identifying this decision rule affect performance of the system; [Rosenfeld and Kraus, 2016b] and [Hadoux *et al.*, 2015] each assumes the strategising agent has a stochastic model of the opponent's expected dialogue behaviour; while game theoretic approaches such as that of [Kacpraz *et al.*, 2014] require knowledge of the opponent's preferences over dialogue outcomes and assume that the opponent plays optimally to maximise those preferences.

Of course, the effectiveness of model-based strategies depends on how accurate the underlying model's representation of the opponent is. However, the important question of how to develop an accurate opponent model is, as yet, somewhat under-explored in the computational argumentation community. Some works consider how one can develop an opponent model from historical data. For example, [Hadjinikolis *et al.*, 2013] explores how an agent can use its knowledge of past dialogue interactions to predict what is likely to be believed by a new opponent, while [Rosenfeld and Kraus, 2016b] and [Hunter and Polberg, 2017] each show how machine learning techniques can be applied to argumentative data collected from humans in order to make predictions about

an opponent ([Rosenfeld and Kraus, 2016b]) seeks to predict a stochastic model of an opponent’s dialogue behaviour; [Hunter and Polberg, 2017] aims to predict the believability, convincingness and appeal of specific arguments to an opponent). Other works explore how one can update an opponent model either during or after a dialogue with that opponent, based on their dialogue behaviour. For example, [Rach *et al.*, 2019] shows how emotion recognition can be used to determine a model of a human opponent’s preferences over arguments, by examining their emotional responses to presented arguments, and [Hunter *et al.*, 2018] explores how an epistemic graph (which uses probabilities to express the degree to which an argument is believed or disbelieved by an agent) can be updated based on the opponent’s dialogue behaviour (see Chapter 9 of this volume for more details).

Rienstra *et al.* [2013] and Black and Hunter [2016] each look at how to update an uncertain opponent model in the case the opponent makes a move that is inconsistent with some of the possibilities represented by the model. In these works, the opponent model can — in its simplest form — be represented as a set $OM = \{S_1, \dots, S_n\} \subset 2^{Args}$, where it is believed that the arguments available to the opponent are those in some $S_i \in OM$. If the opponent asserts an argument a that is not part of one of these possible sets, Black and Hunter [2016] — who assume an accurate opponent model in the sense that the opponent’s actual arguments are represented by one of the possible sets S_i — remove from the model each S_i that does not contain a . In contrast, Rienstra *et al.* [2013] add the newly asserted argument to every possible set from the model (and so $S_i := S_i \cup \{a\}$ for each i).

Scalability of model-based approaches. Another key challenge for model-based dialogue strategies is scalability. These approaches typically assume an uncertain model of the opponent. Thus, to evaluate a potential strategy, one may need to consider all the different sets of arguments that could be available to the opponent, and all the possible ways the opponent may behave during the dialogue given each set of arguments that may be available to them. In the worst case, where one has complete uncertainty over which arguments are available to the opponent, there are 2^n — where n is the number of arguments in the problem domain — sets of arguments that may be available to the opponent. For each of these sets, the number of ways the opponent may behave in the dialogue grows factorially with the number of arguments (since, in the general case, agents can choose to assert any number of those arguments, in any order). Similarly, the number of potential strategies to consider for the strategising agent grows factorially with the number of arguments in the problem domain. This scalability challenge is compounded if one is considering multiple opponents. It can be possible to exploit the structured nature of argumentation in order to prune the problem instance and improve scalability. For example, in persuasion dialogues one may be able to remove dominated arguments (arguments that are known to be attacked by an argument that is itself not attacked by anything) from the problem [Hadoux *et al.*, 2015].

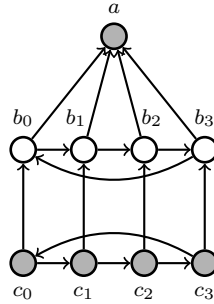


Figure 7. Example of a problem that is particularly challenging for a persuader

Another approach to improving scalability can be to restrict the potential strategies for consideration. The automated planning approach of Black *et al.* [2017] considers only what they refer to as *simple* strategies (a sequence of moves to be made by the agent regardless of how the opponent behaves in the dialogue, in contrast to a policy where the move to be made by the strategising agent depends on the opponent’s moves) returning a simple strategy that maximises — according to the opponent model — the probability of guaranteed success no matter which strategy the opponent employs. One may also consider approaches that, rather than guaranteeing a strategy that is optimal given the opponent model (such as [Hadoux *et al.*, 2015; Hadoux and Hunter, 2017]²²) aim to return a near optimal strategy. For example, Murphy *et al.* [2018] show how the problem of determining a strategy for one-to-many persuasion can be encoded using techniques from search-based model engineering, such that evolutionary search can be applied to find a near optimal strategy. This also allows for optimisation of multiple objectives, such as maximising the number of opponents who are persuaded while also minimising the number of arguments shared.

The structure of the underlying argumentation framework — representing the domain knowledge — can also have a significant effect on the time taken to find an effective strategy. Black *et al.* [2017] consider different structures of frameworks, representing the arguments potentially available to a persuader and its opponent. They show that it is significantly faster to find strategies for bipartite graphs, where the persuader does not need to worry about undermining its own arguments, than it is for graphs that contain cycles, where whether asserting a particular argument is helpful or harmful for the persuader depends on the arguments available to the opponent. Consider for example the argumentation framework shown in Figure 7 [Black *et al.*, 2017], where the greyed out nodes are the arguments available to the persuader, the white nodes are the

²²[Black *et al.*, 2017] guarantee an optimal *simple* strategy, but such strategies may be outperformed by an optimal *policy*.

arguments that may be known by the opponent (and so we are assuming that the arguments known to the opponent are some element of $2^{\{b_0, b_1, b_2, b_3\}}$), and the edges represent attacks between arguments. The persuader aims to convince its opponent of the acceptability of argument a . In this case, asserting the argument c_0 could be beneficial to the proponent if the opponent knows the argument b_0 (as c_0 attacks b_0) but if the opponent knows b_1 then asserting c_0 could be detrimental to the proponent’s goal (since c_0 attacks c_1 , which is the only argument available that attacks b_1). The proponent must therefore take care to consider all the possibilities captured by its opponent model in order to determine whether an argument is likely to be detrimental to achieving its goal.

5.2.3 Benchmarks for evaluating dialogue strategies

The discussion above about how the structure of the underlying argumentation framework can affect the difficulty of finding effective strategies highlights another challenge for argument dialogues: the lack of benchmark problems for evaluation. The International Competition on Computational Models of Argumentation (ICCMA) provides benchmark problems for the evaluation of abstract argument solvers (see for example [Gaggl *et al.*, 2020]). However, while there are a standard set of decision problems addressed by argument solvers (for example: is a particular argument credulously accepted under the preferred semantics; or return the grounded extension) identifying meaningful benchmarks for argument dialogue systems is more challenging because: these systems are designed with different goals in mind (and even dialogue systems for the same type of dialogue typically employ different protocols); there is no standardised way of representing dialogue systems (see Section 3); there is no agreement on the underlying argumentation formalism used for reasoning; and different assumptions are made about the knowledge available to the strategising agent.

5.3 Using enthymemes in dialogues

Enthymemes are incomplete arguments, where some of the premises and/or the claim of the intended complete argument are omitted. Arguments presented by humans are normally enthymemes [Walton, 1989], and so if we want to support human-agent communication we need to be able to handle enthymemes in dialogues.²³ Humans normally find it easy to understand the intended meaning of an enthymeme. Consider the following example (adapted from [Sperber and Wilson, 1986]):

A: Would you like a coffee?

B: Coffee will keep me awake.

Here, B presents an enthymeme that is missing its claim (since it does not explicitly answer A’s question) and also some premises. If A knows that B

²³Even if only considering agent-agent communication, we may want to handle enthymemes in order to improve efficiency of the dialogue.

wants to go to sleep (perhaps B is in their pyjamas ready for bed) then A can infer that B’s intended argument was “Coffee will keep me awake, I want to go to sleep, so I would not like a coffee”. If, however, A knows that B needs to stay awake (because they have a paper deadline to meet) then A can infer that B’s intended argument was “Coffee will keep me awake, I need to stay awake, so I would like a coffee”. This illustrates how common knowledge between the participants (specifically, whether B wants to go to sleep or needs to stay awake) can be used to reconstruct the intended complete argument from an enthymeme.

The context can also be important for correctly understanding enthymemes. Consider the following example (also adapted from [Sperber and Wilson, 1986]):

A: Sarita bought The Times.

It is common knowledge²⁴ that The Times can refer to a copy of the British newspaper The Times, or to the company that publishes The Times, and so there are two possible intended arguments to consider: “Sarita bought The Times, The Times refers to a copy of The Times newspaper, so Sarita bought a copy of The Times newspaper” and “Sarita bought The Times, The Times refers to the company that publishes The Times newspaper, so Sarita bought the company that publishes The Times newspaper”. If the statement “Sarita bought the Times” was made during a conversation about how successful a businesswoman Sarita is, one can infer that the most relevant intended argument is the one that concludes that Sarita bought the company that published the The Times. If it was made in response to the statement “I must get a newspaper today”, one can infer that the most relevant intended argument is the one claiming that Sarita bought a copy of the newspaper The Times.

Despite their ubiquity in human communication, there are few formal proposals of how enthymemes can be used in computational argumentation dialogues. Key questions to be addressed include:

- How can an agent determine an appropriate enthymeme to put forward, so that the receiver of the enthymeme can accurately reconstruct the intended complete argument?
- How can an agent receiving an enthymeme accurately identify the intended complete argument?

We have seen in the examples above that common knowledge is important for understanding enthymemes. In [Hunter, 2007] a proposal is made for how the common knowledge can be used both to determine an appropriate enthymeme to send and to identify the intended complete argument. This work,

²⁴For members of some groups — in New York and much of the rest of the United States most people would take “The Times” to refer to the New York Times, and would call the similarly named British newspaper “The London Times”, even though that is not its name. For those people, the reference in the example changes to the New York Times, but is otherwise unchanged since the company that publishes it is also known as “The Times”.

which considers only enthymemes with missing premises (and so cannot handle enthymemes with missing claims), assumes that both the proponent and the receiver of an enthymeme have a model of what they believe is common knowledge between the two. This model assigns a value between 0 and 1 to each element of the domain language, representing the degree to which the agent believes the element can be used as common knowledge, and so anything with a value over a particular threshold can be used as common knowledge. Determining an enthymeme to present is straightforward, one simply removes from the premises of the intended argument anything that one believes can be used as common knowledge. Reconstructing the intended argument involves adding to the enthymeme elements from what is perceived to be the common knowledge in order to build up a complete argument. However, it is not guaranteed that the reconstruction will result in the correct intended argument: the agents' perceptions of the common knowledge may differ, or there may be multiple possible intended arguments to select from.

In the “The Times” example above, where there are multiple possible intended arguments that can be reconstructed from the enthymeme put forward, we see how contextual information about the relevance of arguments can be useful when selecting the appropriate intended argument. Black and Hunter [2012b] extend the proposal discussed above to also support enthymemes with missing claims, taking into account not only the common knowledge but also the relevance of arguments. In addition to modelling what each agent believes can be used as common knowledge, a public agenda is assumed for each agent, which represents a ranking of the agent's information requirements (i.e., propositions such that the agent would like to know if there are reasons to believe they do or do not hold). By harnessing some of the principles of relevance theory [Sperber and Wilson, 2004] (essentially that “relevance of an utterance depends on maximising cognitive effect while minimising cognitive effort” [Black and Hunter, 2012b, p. 56]) Black and Hunter show how one can use both the common knowledge and the agenda to determine enthymemes that are relevant to their recipient, and to overcome some of the ambiguities that arise in selecting the most appropriate intended argument when receiving an enthymeme. Argumentation schemes [Walton and Krabbe, 1996; Macagno *et al.*, 2018] may also be helpful in providing more contextual information that can be used to help determine the appropriate intended argument [Walton and Reed, 2005; Panisson and Bordini, 2017].

Although not the focus of Black and Hunter's work [2012b], they briefly discuss how the information requirements might be obtained. A straightforward case is where an agent asks a specific question, and so explicitly states an information requirement (as in the coffee example above). One might also assume that an agent will typically always be concerned with aspects such as its own welfare and safety, and so will have information requirements to reflect this. It may also be possible to derive information requirements from the context of the dialogue: if an agent is discussing Sarita's success as a

businesswoman, one might infer information requirements relating to Sarita’s business achievements.

There is also the question of how an agent can develop a view of what can be used as common knowledge. Again, there are some seemingly straightforward cases. If an agent has uttered a proposition, or been present in a dialogue where the proposition was uttered, one may consider that the proposition can be regarded as common knowledge with that agent. However, what if the agent uttered the proposition last week, or last year; can we assume it still remembers, or believes in, the proposition? [Hosseini *et al.*, 2014] considers how an agent x might develop a model of the common knowledge with an agent y taking into account both the direct and indirect information gained about y ’s knowledge (for example, if y uses some proposition α in a dialogue with x , or some other agent z informs x that y knows α), as well as considering whether there are things that one would typically expect y to know (for example, if y works at a university, one might assume they are familiar with the regulations of that university). [Black and Hunter, 2008] also looks at how an agent x might update its view of the common knowledge with agent y based on their dialogue behaviour. Neither of these works considers the temporal persistence of common knowledge.

So far, our discussion of enthymemes has focused on how an agent can determine an appropriate enthymeme to send such that the receiver is likely to be able to accurately reconstruct the intended argument, and how an agent who receives an enthymeme can identify the argument that was intended; we have not yet said anything about how these enthymemes might actually be used in a dialogue system. The use of enthymemes in dialogues implies the need for locutions that allow agents to recover from misunderstandings: while we hope our processes – for identifying an enthymeme to send and for reconstructing the intended argument – will perform well, misunderstandings will inevitably occur (for example, if there are errors in the common knowledge, or the agents have different perceptions of the relevance of an argument).

There are very few proposals for dialogue systems that support the use of enthymemes. Black and Hunter [2008] present a system for inquiry dialogues that handles enthymemes with missing premises, which includes a **quiz** locution that agents can use to ask for clarification when they cannot reconstruct any complete arguments from a received enthymeme. [Dupin de Saint-Cyr, 2011] and [Xydis *et al.*, 2020] each present a dialogue system that supports enthymemes which may miss premises or their claim, and show that different locutions are needed to handle the misunderstandings that can occur from these different cases. The use of enthymemes in dialogue raises the question of whether, and how, they can be used to give a strategic advantage. When an enthymeme e is moved against an argument a , by omitting some elements of e ’s intended argument one makes it harder for the receiver of e to identify counter arguments, and to identify whether e is indeed a counter argument for a . Consider the following example from Schopenhauer [1831] (also discussed in

[Dupin de Saint-Cyr, 2011]):

I asserted that the English were supreme in drama. My opponent attempted to give an instance to the contrary, and replied that it was a well-known fact that in music, and consequently in opera, they could do nothing at all. I repelled the attack by reminding him that music was not included in dramatic art, which covered tragedy and comedy alone. This he knew very well. What he had done was to try to generalise my proposition, so that it would apply to all theatrical representations, and, consequently, to opera and then to music, in order to make certain of defeating me.

Schopenhauer’s interpretation here is that his opponent presents an enthymeme in the full knowledge that the enthymeme’s intended argument does not in fact counter Schopenhauer’s claim, hoping this will go unnoticed by Schopenhauer. As far as we are aware, there are no existing works that explore how enthymemes can be used to give such a strategic advantage in computational argumentation-based dialogues.

6 Conclusions

This chapter has given an introduction to work on argumentation-based dialogue, focusing on what we believe are the key aspects that have been studied so far. We started with an overview of the basic elements of argumentation-based dialogues, the speech acts from which dialogues are formed, the protocols that govern them, the strategies that can be employed within them, and ways that dialogues can be evaluated. We illustrated these ideas with a simple Hamblin-style dialogue game. This was followed by a discussion of approaches to representing dialogues, something that is necessary if we are going to be able to compare different approaches formally, and if we are going to be able to examine the properties of dialogues at a suitably abstract level. We then gave a detailed example of a dialogue system — the value-based deliberation system of [Black and Atkinson, 2011a; Black and Atkinson, 2011b] — before discussing what we believe are the major current challenges in argumentation-based dialogue. These challenges are: multiparty dialogues, dialogue strategy, and handling enthymemes.

There are also a number of other topics that we think are important for achieving the full potential of argumentation-based dialogue systems, but which we have not had the space to discuss here in detail. One thing we have not considered here is where an agent gets its arguments from. For some applications, it may be feasible for a domain expert to manually provide the knowledge that agents can use to construct arguments. However, this is not always practical and in some cases we may need agents that are capable of discovering arguments for themselves, for example using *argument mining* techniques to identify arguments from text on the web (see, for example, [Budzynska and Villata, 2018; Lawrence and Reed, 2020]). We briefly discussed earlier in this chapter the need

for benchmark problems for evaluation of argumentation-based dialogue systems, and argument mining techniques may also help us to identify real-world examples for this. Also on the subject of evaluation, if we aim for dialogue systems that can support human-agent interaction then we need to perform experiments with humans in order to ensure, for example, that our formal models of reasoning align with human behaviour (see Chapter 18 of this volume for a survey of work in this area) and that the dialogue systems we develop do indeed benefit human users in the ways we envisage. We need also to consider the broader ethical issues associated with developing argumentation-based dialogue systems that can, potentially, affect human users' beliefs or decision making, for example: is it ethical for an agent to withhold some relevant arguments from a human — or even to lie²⁵ — in order to reach what might be viewed as a better dialogue outcome; is it responsible to develop agents that are capable of persuading a human to change their behaviour; are the dialogue systems we develop likely to disadvantage some subgroups of the population?

As artificial intelligence systems become increasingly ubiquitous, we believe that argumentation-based dialogues have great potential for allowing such systems to engage with each other and with human users in joint reasoning and decision making. An example of the former is [Kodeswaran *et al.*, 2010], and the latter is [Sklar and Azhar, 2015]. In both cases a key advantage is the ease with which such systems can be understood, indeed in [Sklar and Azhar, 2015; Azhar and Sklar, 2017] the system explicitly uses argumentation to explain their reasoning to human users. This ability to support explainable decisions is crucial if we are to be able to trust that artificial intelligence systems are acting in our best interests.

Acknowledgments

This work was partially supported by EPSRC grants EP/P010105/1, EP/R033722/1, and EP/M01892X/1. The opinions expressed in this chapter are those of the authors and do not necessarily reflect the opinions of the funders.

BIBLIOGRAPHY

- [Amgoud and Parsons, 2001] Leila Amgoud and Simon Parsons. Agent dialogues with conflicting preferences. In *Proceedings of the 8th International Workshop on Agent Theories, Architectures and Languages*, pages 1–15, 2001.
- [Amgoud *et al.*, 2000a] Leila Amgoud, Nicolas Maudet, and Simon Parsons. Modelling dialogues using argumentation. In *Proceedings of the 4th International Conference on Multi-Agent Systems*, pages 31–38, 2000.
- [Amgoud *et al.*, 2000b] Leila Amgoud, Simon Parsons, and Nicolas Maudet. Arguments, dialogue, and negotiation. In *Proceedings of the 14th European Conference on Artificial Intelligence*, pages 338–342, 2000.
- [Atkinson and Bench-Capon, 2007] Katie Atkinson and Trevor J.M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10–15):855–874, 2007.

²⁵The area of mechanism design can support the design of dialogue systems in which there is no incentive to lie, see for example [Rahwan and Larson, 2009].

- [Atkinson *et al.*, 2013] Katie Atkinson, Priscilla Bench-Capon, and Trevor J.M. Bench-Capon. A strategy for efficient persuasion dialogues. In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence*, pages 332–247, 2013.
- [Austin, 1975] John Langshaw Austin. *How to do things with words*. Oxford University Press, 1975.
- [Azhar and Sklar, 2017] M Q Azhar and Elizabeth I Sklar. A study measuring the impact of shared decision making in a human-robot team. *International Journal of Robotics Research (IJRR)*, 36:461–482, 2017.
- [Bench-Capon, 2002] Trevor J.M. Bench-Capon. Agreeing to differ: Modelling persuasive dialogue between parties without a consensus about values. *Informal Logic*, 22(3):231–245, 2002.
- [Black and Atkinson, 2009] Elizabeth Black and Katie Atkinson. Dialogues that account for different perspectives in collaborative argumentation. In *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 867–874, 2009.
- [Black and Atkinson, 2011a] Elizabeth Black and Katie Atkinson. Agreeing what to do. In *Proceedings of the 7th International Workshop on Argumentation in Multi-Agent Systems*, pages 12–30, 2011.
- [Black and Atkinson, 2011b] Elizabeth Black and Katie Atkinson. Choosing persuasive arguments for action. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, pages 849–856, 2011.
- [Black and Bentley, 2012] Elizabeth Black and Katie Bentley. An empirical study of a deliberation dialogue system. In *Proceedings of the 1st International Workshop on the Theory and Applications of Formal Argumentation*, pages 132–146, 2012.
- [Black and Hunter, 2008] Elizabeth Black and Anthony Hunter. Using enthymemes in an inquiry dialogue system. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 437 – 444, 2008.
- [Black and Hunter, 2009] Elizabeth Black and Anthony Hunter. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, 19(2):173–209, 2009.
- [Black and Hunter, 2012a] Elizabeth Black and Anthony Hunter. Executable logic for dialogical argumentation. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pages 15–20, 2012.
- [Black and Hunter, 2012b] Elizabeth Black and Anthony Hunter. A relevance-theoretic framework for constructing and deconstructing enthymemes. *Journal of Logic and Computation*, 22(1):55 – 78, 2012.
- [Black and Hunter, 2016] Elizabeth Black and Anthony Hunter. Reasons and options for updating an opponent model in persuasion dialogues. In *Proceedings of the 3rd International Workshop on the Theory and Applications of Formal Argumentation*, pages 21–39, 2016.
- [Black and Sklar, 2016] Elizabeth Black and Elizabeth I. Sklar. Computational argumentation to support multi-party human-robot interaction: challenges and advantages. In *Proceedings of the Groups in Human-Robot Interaction Workshop: A workshop at the IEEE International Symposium on Robot and Human Interactive Communication*, 2016.
- [Black *et al.*, 2014] Elizabeth Black, Amanda J Coles, and Sara Bernardini. Automated planning of simple persuasion dialogues. In *Proceedings of the 15th International Workshop on Computational Logic in Multi-Agent Systems*, pages 87–104, 2014.
- [Black *et al.*, 2017] Elizabeth Black, Amanda J Coles, and Christopher Hampson. Planning for persuasion. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 933–942, 2017.
- [Bonzon and Maudet, 2011] Elise Bonzon and Nicolas Maudet. On the outcomes of multiparty persuasion. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS’11)*, pages 47–54, May 2011.
- [Budzynska and Villata, 2018] Katarzyna Budzynska and Serena Villata. Processing natural language argumentation. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation, Volume 1*, pages 577 – 627. College Publications, 2018.
- [Dignum and Vreeswijk, 2004] Frank P.M. Dignum and Gerard A.W. Vreeswijk. Towards a testbed for multi-party dialogues. In *Advances in Agent Communication*, volume 2922 of *Lecture Notes in Computer Science*, pages 1955–1955. Springer, 2004.

- [Dimopoulos *et al.*, 2019] Yannis Dimopoulos, Jean-Guy Mailly, and Pavlos Moraitis. Argumentation-based negotiation with incomplete opponent profiles. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1252–1260, 2019.
- [Dung, 1995] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n -person games. *Artificial Intelligence*, 77:321–357, 1995.
- [Dupin de Saint-Cyr, 2011] Florence Dupin de Saint-Cyr. Handling enthymemes in time-limited persuasion dialogs. In *Proceedings of the 5th International Conference on Scalable Uncertainty Management*, pages 149–162, 2011.
- [Endriss *et al.*, 2004] Ulle Endriss, Nicolas Maudet, Fariba Sadri, and Francesca Toni. Logic-based agent communication protocols. In F. Dignum, editor, *Advances in Agent Communication*, volume 2922 of *LNAI*, pages 91–107. Springer-Verlag, 2004. Invited contribution.
- [Fan and Toni, 2012] Xiuyi Fan and Francesca Toni. Agent strategies for aba-based information-seeking and inquiry dialogues. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pages 324–329, 2012.
- [Fernández and Endriss, 2007] Raquel Fernández and Ulle Endriss. Abstract models for dialogue protocols. *Journal of Logic, Language and Information*, 16(2):121–140, 2007.
- [Gabbay and Woods, 2001a] Dov M. Gabbay and John Woods. More on non-cooperation in Dialogue Logic. *Logic Journal of the IGPL*, 9(2):321–339, 2001.
- [Gabbay and Woods, 2001b] Dov M. Gabbay and John Woods. Non-cooperation in Dialogue Logic. *Synthese*, 127(1-2):161–186, 2001.
- [Gaggl *et al.*, 2020] Sarah A. Gaggl, Thomas Linsbichler, Marco Maratea, and Stefan Woltran. Design and results of the second international competition on computational models of argumentation. *Artificial Intelligence*, 279:103193, 2020.
- [Gordon, 1994] Thomas F. Gordon. The Pleadings Game: An exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.
- [Hadjinikolis *et al.*, 2013] Christos Hadjinikolis, Yiannis Siantos, Sanjay Modgil, Elizabeth Black, and Peter McBurney. Opponent modelling in persuasion dialogues. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 164–170, 2013.
- [Hadoux and Hunter, 2017] Emmanuel Hadoux and Anthony Hunter. Strategic sequences of arguments for persuasion using decision trees. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 1128–1134, 2017.
- [Hadoux *et al.*, 2015] Emmanuel Hadoux, Aurélie Beynier, Nicolas Maudet, Paul Weng, and Anthony Hunter. Optimization of probabilistic argumentation with Markov decision models. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 2004–2010, 2015.
- [Hamblin, 1970] Charles L. Hamblin. *Fallacies*. Methuen and Co Ltd, London, UK, 1970.
- [Hamblin, 1971] Charles L. Hamblin. Mathematical models of dialogue. *Theoria*, 37:130–155, 1971.
- [Hitchcock, 1991] David Hitchcock. Some principles of rational mutual inquiry. In *Proceedings of the 2nd International Conference on Argumentation*, pages 236–243, 1991.
- [Hosseini *et al.*, 2014] Seyed Ali Hosseini, Sanjay Modgil, and Odinaldo Rodrigues. Enthymeme construction in dialogues using shared knowledge. In *Proceedings of the 5th International Conference on Computational Models of Argument*, pages 325 – 332, 2014.
- [Hulstijn, 2000] J. Hulstijn. *Dialogue Models for Inquiry and Transaction*. PhD thesis, Universiteit Twente, Enschede, The Netherlands, 2000.
- [Hunter and Polberg, 2017] Anthony Hunter and Sylwia Polberg. Empirical methods for modelling persuadees in dialogical argumentation. In *Proceedings of the 29th IEEE International Conference on Tools with Artificial Intelligence*, pages 382–389, 2017.
- [Hunter *et al.*, 2018] Anthony Hunter, Sylwia Polberg, and Nico Potyka. Updating belief in arguments in epistemic graphs. In *Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning*, pages 138–147, 2018.
- [Hunter, 2007] Anthony Hunter. Real arguments are approximate arguments. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*, pages 66–71, 2007.
- [Kacpraz *et al.*, 2014] Magdalena Kacpraz, Marcin Dziubinski, and Katarzyna Budzyska. Strategies in dialogues: A game-theoretic approach. In *Proceedings of the 5th International Conference on Computational Models of Argument*, pages 333 – 344, 2014.

- [Kakas *et al.*, 2005] Antonis C. Kakas, Nicolas Maudet, and Pavlos Moraitis. Modular representation of agent interaction rules through argumentation. *Autonomous Agents Multi-Agent Systems*, 11(2):189–206, 2005.
- [Kodeswaran *et al.*, 2010] Palanivel Kodeswaran, Wenjia Li, Anupam Joshi, Tim Finin, and Filip Perich. Enforcing secure and robust routing with declarative policies. In *Military Communications Conference (MILCOM)*, pages 44–49. IEEE, 2010.
- [Kontarinis *et al.*, 2014] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, and Pavlos Moraitis. Empirical evaluation of strategies for multiparty argumentative debates. In *15th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA'14)*, pages 105–122, 2014.
- [Lawrence and Reed, 2020] John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2020.
- [Loui and Moore, 1997] R. P. Loui and D. M. Moore. Dialogue and deliberation. Report WUCS-97-11, Computer Science and Engineering, Washington University in St. Louis, 1997.
- [Macagno *et al.*, 2018] Fabrizio Macagno, Douglas Walton, and Chris Reed. Argumentation schemes. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation, Volume 1*, pages 519 – 576. College Publications, 2018.
- [MacKenzie, 1979a] Jim D. MacKenzie. How to stop talking to tortoises. *Notre Dame Journal of Formal Logic*, 20 (4):705–717, 1979.
- [MacKenzie, 1979b] Jim D. Mackenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–113, 1979.
- [MacKenzie, 1990] Jim D. MacKenzie. Four dialogue systems. *Studia Logica*, 49 (4):567–583, 1990.
- [McBurney and Parsons, 2002] Peter McBurney and Simon Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language, and Information*, 11(3):315–334, 2002.
- [McBurney *et al.*, 2007] Peter McBurney, David Hitchcock, and Simon Parsons. The eight-fold way of deliberation dialogue. *International Journal of Intelligent Systems*, 22(1):95–132, 2007.
- [Medellin-Gasque *et al.*, 2013] Rolando Medellin-Gasque, Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Strategies for question selection in argumentative dialogues about plans. *Argument and Computation*, 4(2):151–179, 2013.
- [Mercier and Sperber, 2011] Hugo Mercier and Dan Sperber. Why do humans reason? arguments for an argumentative theory. *Behavioral and brain sciences*, 34(2):57–74, 2011.
- [Modgil and Caminada, 2009] Sanjay Modgil and Martin Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.
- [Modgil and Prakken, 2013] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.
- [Murphy *et al.*, 2016] Josh Murphy, Elizabeth Black, and Michael Luck. A heuristic strategy for persuasion dialogues. In *Proceedings of the 6th International Conference on Computational Models of Argument*, pages 411 – 418, 2016.
- [Murphy *et al.*, 2018] Josh Murphy, Alexandru Burdusel, Michael Luck, Steffen Zschaler, and Elizabeth Black. Deriving persuasion strategies using search-based model engineering. In *Proceedings of the 7th International Conference on Computational Models of Argument*, pages 221–232, 2018.
- [Oren *et al.*, 2006] Nir Oren, Timothy J. Norman, and Alun Preece. Loose lips sink ships: A heuristic for argumentation. In *Proceedings of the 3rd International Workshop on Argumentation in Multi-Agent Systems*, pages 121 – 134, 2006.
- [Panisson and Bordini, 2017] Alison R. Panisson and Rafael H. Bordini. Uttering Only What is Needed: Enthymemes in Multi-Agent Systems. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems*, pages 1670–1672, 2017.
- [Parsons *et al.*, 2003] Simon Parsons, Michael Wooldridge, and Leila Amgoud. Properties and complexity of formal inter-agent dialogues. *Journal of Logic and Computation*, 13(3):347–376, 2003.

- [Parsons *et al.*, 2011] Simon Parsons, Yuqing Tang, Elizabeth I. Sklar, Peter McBurney, and Kai Cai. Argumentation-based reasoning in agents with varying degrees of trust. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems*, 2011.
- [Plotkin, 1981] Gordon D. Plotkin. A structural approach to operational semantics. Technical Report DAIMI FN-19, Computer Science Department, Aarhus University, 1981.
- [Prakken and Sartor, 1998] Henry Prakken and Giovanni Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6:231–287, 1998.
- [Prakken, 2000] Henry Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In *Proceedings of the 7th European Workshop on Logic in Artificial Intelligence*, pages 224–238, 2000.
- [Prakken, 2001a] Henry Prakken. Modelling reasoning about evidence in legal procedure. In *Proceedings of the 8th International Conference on Artificial Intelligence and Law*, pages 119–128, 2001.
- [Prakken, 2001b] Henry Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.
- [Prakken, 2005] Henry Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 2005.
- [Prakken, 2006] Henry Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 21(2):163–188, 2006.
- [Rach *et al.*, 2018] Nicklas Rach, Wolfgang Minker, and Stefan Ultes. Markov games for persuasive dialogue. In *Proceedings of the 7th International Conference on Computational Models of Argument*, pages 213 – 220, 2018.
- [Rach *et al.*, 2019] Niklas Rach, Klaus Weber, Annalena Aicher, Florian Lingenfeller, Elisabeth Andre, and Wolfgang Minker. Emotion recognition based preference modelling in argumentative dialogue systems. In *Proceedings of the 1st International Workshop on Pervasive Computing and Spoken Dialogue Systems Technology*, pages 838–843, 2019.
- [Rahwan and Larson, 2009] Iyad Rahwan and Kate Larson. Argumentation and game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009.
- [Rahwan *et al.*, 2003] Iyad Rahwan, Sarvapalid D. Ramchurn, Nicholas R. Jennings, Peter McBurney, Simon Parsons, and Liz Sonenberg. Argumentation-based negotiation. *Knowledge Engineering Review*, 18(4):343–375, 2003.
- [Rahwan *et al.*, 2009] Iyad Rahwan, Philippe Pasquier, Liz Sonenberg, and Frank Dignum. A formal analysis of interest-based negotiation. *Annals of Mathematics and Artificial Intelligence*, 55(3-4):253, 2009.
- [Rienstra *et al.*, 2013] Tjitze Rienstra, Matthias Thimm, and Nir Oren. Opponent models with uncertainty for strategic argumentation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence International Joint Conference on Artificial Intelligence*, pages 332–338, 2013.
- [Rosenfeld and Kraus, 2016a] Ariel Rosenfeld and Sarit Kraus. Providing arguments in discussions on the basis of the prediction of human argumentative behavior. *ACM Trans. Interact. Intell. Syst.*, 6(4):30:1–30:33, 2016.
- [Rosenfeld and Kraus, 2016b] Ariel Rosenfeld and Sarit Kraus. Strategical argumentative agent for human persuasion. In *Proceedings of the 22nd European Conference on Artificial Intelligence*, volume 285, pages 320–328. IOS Press, 2016.
- [Russell and Norvig, 2020] Stuart Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Pearson Education Limited, 4th edition edition, 2020.
- [Sadri *et al.*, 2001] Fariba Sadri, Francesca Toni, and Paolo Torroni. Dialogues for negotiation: Agent varieties and dialogue sequences. In John-Jules Ch. Meyer and Milind Tambe, editors, *Intelligent Agents VIII, 8th International Workshop, ATAL 2001 Seattle, WA, USA, August 1-3, 2001, Revised Papers*, volume 2333 of *Lecture Notes in Computer Science*, pages 405–421. Springer, 2001.
- [Schopenhauer, 1831] Arthur Schopenhauer. The art of always being right: 38 ways to win an argument. https://en.wikisource.org/wiki/The_Art_of_Being_Right, 1831. Original title: Die Kunst, Recht zu behalten (Translated by Thomas Saunders in 1896).
- [Searle, 1969] John R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, 1969.

- [Singh, 2000] Mumindar P. Singh. A social semantics for agent communication languages. In Frank Dignum and Mark Greaves, editors, *Issues in agent communication*, volume 1916 of *Lecture Notes in Computer Science*, pages 31–45. Springer, Berlin, Heidelberg, 2000.
- [Sklar and Azhar, 2015] E. I. Sklar and M. Q. Azhar. Argumentation-based dialogue games for shared control in human-robot systems. *Journal of Human-Robot Interaction*, 4(3):120–148, 2015.
- [Sperber and Wilson, 1986] Dan Sperber and Deidre Wilson. *Relevance: Communication and Cognition*. Blackwell Publishing, 1986.
- [Sperber and Wilson, 2004] Dan Sperber and Deidre Wilson. Relevance theory. In Laurence R. Horn and Gregory Ward, editors, *The Handbook of Pragmatics*, pages 607 – 632. Blackwell Publishing, 2004.
- [Thielscher, 2010] Michael Thielscher. A general game description language for incomplete information games. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 994–999, 2010.
- [Thimm, 2014] Matthias Thimm. Strategic argumentation in multi-agent systems. *KI*, 28(3):159–168, 2014.
- [Torrioni, 2002] Paolo Torrioni. A study on the termination of negotiation dialogues. In *Proceedings of The 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1223–1230, 2002.
- [Walton and Krabbe, 1996] Douglas N. Walton and Erik C. W. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*. Suny Press, 1996.
- [Walton and Reed, 2005] Douglas N. Walton and Chris Reed. Argumentation schemes and enthymemes. *Synthese*, 145:339–370, 2005.
- [Walton, 1989] Douglas N. Walton. *Informal Logic: A Handbook for Critical Argumentation*. Cambridge University Press, 1989.
- [Wells and Reed, 2012] Simon Wells and Chris Reed. A domain specific language for describing diverse systems of dialogue. *Journal of Applied Logic*, 10(4):309 – 329, 2012.
- [Wooldridge, 2009] Michael Wooldridge. *An Introduction to Multiagent Systems*. Wiley, 2nd edition, 2009.
- [Xydis *et al.*, 2020] Andreas Xydis, Christopher Hampson, Sanjay Modgil, and Elizabeth Black. Enthymemes in dialogue. In *Proceedings of the 8th International Conference on Computational Models of Argument*, 2020. (In press).

Elizabeth Black
 King’s College London, UK
 Email: elizabeth.black@kcl.ac.uk

Nicolas Maudet
 Sorbonne University, France
 Email: nicolas.maudet@lip6.fr

Simon Parsons
 University of Lincoln, UK
 Email: sparsons@lincoln.ac.uk