
Autonomous Agents and Multi-Agent Systems, XX, XXX–XXX, 2005

A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action

KATIE ATKINSON k.m.atkinson@csc.liv.ac.uk
Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK

TREVOR BENCH-CAPON tbc@csc.liv.ac.uk
Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK

PETER MCBURNEY p.j.mcburney@csc.liv.ac.uk
Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK

Abstract. We present the syntax and semantics for a multi-agent dialogue game protocol which permits argument over proposals for action. The protocol, called the Persuasive Argument for Multiple Agents (PARMA) *Protocol*, embodies an earlier theory by the authors of persuasion over action which enables participants to rationally propose, attack, and defend, an action or course of actions (or inaction). We present an outline of both an axiomatic and a denotational semantics, and discuss implementation of the protocol, in the context of both human and artificial agents.

Keywords: argumentation, practical reasoning, protocols, BDI agents.

1. Introduction

Developers of real-world software agent systems typically desire either the system as a whole or the agents within it to effect changes in the state of the world external to the system. Whether the software agents represent human bidders in an online auction or the system collectively manages some resource, such as a utility network, the agents and/or the system usually need to initiate, maintain or terminate actions in the world [20]. Agent interaction protocols, therefore, must be concerned with argument over actions: even if agents in such systems are not concerned with sharing and reconciling one another's beliefs, these protocols will still assist in sharing and coordinating their actions.

Philosophers of argumentation, however, have mostly concentrated their attention on beliefs, and not on actions.[1] Computer scientists, also, have typically not distinguished between justifications for beliefs and for actions. Attempting to fill this gap, we have previously articulated a theory of persuasion over actions, in which a proponent of a proposed action can seek to persuade another party (a human or software agent) to endorse it [2,4,15]. By classifying all the possible attacks on a proposal for action, our theory permits dialogue participants to state, attack and defend a proposal for action in a systematic manner. Here we extend this work by presenting a novel dialogue game protocol, which we call the Persuasive ARgument for Multiple Agents (PARMA) *Action Persuasion Protocol*, in which proposals for action may be stated, and these attacks and defences may occur.

The paper is structured as follows: Section 2 reprises our general theory of persuasion over action, and indicates the possible attacks on a proposal for action and their resolution. Section 3 presents the syntax and an axiomatic semantics for the PARMA *Action Persuasion Protocol* while Section 4 outlines a denotational semantics for dialogues under the protocol. Section 5 then describes two implementations we have undertaken of the protocol and discusses the merits of each. Section 6 gives an outline of our current work which proposes how our theory can be used within the Belief-Desire-Intention (BDI) agent architecture. Section 7 offers come concluding remarks and indicates directions for possible future work.

It is important to note that dialogues under our protocol are Persuasion dialogues, in the influential terminology of Walton and Krabbe [34][2]. Both Negotiation dialogues (which concern the division of some scarce resource) and Deliberation dialogues (which concern what action to take in some circumstance) in this terminology also concern dialogues over action. A key difference between Negotiation and Deliberation dialogues, on the one hand, and Persuasion dialogues in our sense, on the other, is that Persuasion dialogues commence with at least one participant supporting the proposal for action under discussion (a proposal which may involve not acting). This is not necessarily the case with Negotiation dialogues or Deliberations, both of which may commence without any endorsement by a participant to any proposed action (or inaction), or, indeed, commence without any proposal for action before the participants.

2. A theory of Persuasion over action

Our focus is on rational interactions between agents engaged in joint practical reasoning, that is, seeking to agree on an action or course of action. We use the word *rational* in the sense of argumentation theory, where it is understood as the giving and receiving of reasons for beliefs or actions. In these interactions, we assume that one agent endorses a particular action, and seeks to have another agent do the same. This type of dialogue is a Persuasion dialogue, and our theory permits actions to be proposed, to be attacked, and to be defended by agents engaged in a Persuasion interaction. For such an interaction, we first define what it means to propose an action (Section 2.1), then consider rational attacks on it (Section 2.2), and counters to these attacks (Section 2.3). We then consider resolution, which depends on the nature of the attack (Sections 2.4 and 2.5).

2.1. Stating a Position

We give the following as the general argument schema (called AS1) for a rational position proposing an action:

Argument Schema (AS1):
In the current circumstances R
we should perform action A
to achieve new circumstances S

which will realise some goal G
which will promote some value V.

For current purposes, we need recognise no difference between resolving on a future action and justifying a past action. Moreover, an action may achieve multiple goals, and each goal may promote multiple values. For simplicity, we assume that the proponent of an action articulates an argument in the form of schema *AS1* for each goal realised and value promoted. We assume the existence of:

- A finite set of distinct actions, denoted *Acts*, with elements, A, B, C, etc.
- A finite set of propositions, denoted *Props*, with elements, p, q, r, etc.
- A finite set of states, denoted *States*, with elements, R, S, T, etc. Each element of *States* is an assignment of truth values $\{T, F\}$ to every element of *Props*.
- A finite set of propositional formulae called goals, denoted *Goals*, with elements G, H, etc.
- A finite set of values, denoted *Values*, with elements v, w, etc.
- A function *value* mapping each element of *Goals* to a pair $< v,\ sign >$, where $v \in$ *Values* and *sign* $\in \{+, =, -\}$.
- A ternary relation *apply* on *Acts* $\times$ *States* $\times$ *States*, with *apply(A, R, S)* to be read as: *"Performing action A in state R results in state S."*

The AS1 contains a number of problematic notions which are not readily formalised in classical logic. We can, however, see that there are four classical statements which must hold if the argument represented by schema AS1 is to be valid:

Statement 1: R is the case.
Statement 2: *apply(A, R, S)* $\in$ *apply*.
Statement 3: S $\models$ G. ("G is true in state S.")
Statement 4: *value(G)* $=< v, + >$.

We can represent a position expressed according to AS1 in the following diagrammatic form, where the notation is, we hope, obvious and which will be clarified by the denotational semantics in Section 4:

$$R \xrightarrow{A} S \models G \uparrow v.$$

The possible attacks on a position presented in the next sub-section may be viewed as attacking one or more elements of this representation, or the connections between them.

2.2. *Attacking a Position*

A position proposing an action may be attacked in a number of ways, and we have identified what we believe is an exhaustive list of rational attacks. In Table 1 we summarise these attacks, and indicate the number of variants for each. The fourth column of this table indicates the basis for resolution of any disagreement, which we

Table 1. Attacks on a proposal for action.

Attack	Variants	Description	Basis of resolution
1	2	Disagree with the description of the current situation	What is true
2	7	Disagree with the consequences of the proposed action	What is true
3	6	Disagree that the desired features are part of the consequences	Representation
4	4	Disagree that these features promote the desired value	What is true
5	1	Believe the consequences can be realised by some alternative action	What is best
6	1	Believe the desired features can be realised through some alternative action	What is best
7	2	Believe that the desired value can be realised in an alternative way	What is best
8	1	Believe the action has undesirable side effects which demote the desired value	What is best
9	1	Believe the action has undesirable side effects which demote some other value	What is best
10	1	Agree that the action should be performed, but for different reasons	What is best
11	3	Believe that the action will preclude some more desirable action	What is best
12	1	Believe that the action is impossible	What is true
13	1	Believe that the circumstances as described are not possible	Represenation
14	1	Believe that the consequences as described are not possible	Represenation
15	1	Believe that the desired features cannot be realised	Representation
16	1	Disagree that the desired value is worth promoting	Representation

discuss in Section 2.3. Some attacks (Attacks 1–4) deny the truth or validity of elements of a position, such as the validity of the inference that $S \models G$, for a state S and goal G. A second group of attacks (Attacks 5–7) argue that the same effects can be achieved in an alternative way. A third group (Attacks 8,9, 11) argue against the action proposed because of its undesirable side effects or because of interference with other, preferred, actions. Attack 10 agrees with the action proposed, but offers different reasons from those stated in the position. Such an attack may be important in domains, such as legal reasoning, where the reasons given for actions act as precedents for future decisions. Finally, the last group of attacks (Attacks 12–16) argue that elements of the stated position are invalid or impossible, as, for example, when the attacker disagrees that the proposed action is possible.

The variants on these attacks follow a pattern. An attacker may simply express disagreement with some aspect of a position, as when an attacker denies that R is the current state of the world. Beyond this minimalist attack, an attacker may also state an alternative position to that proposed, for example, expressing not only that R is not the current state of the world, but also that T is the current state. A full list and description of the attacks and their variants are given in Refs. 2 and 15.

2.3. *Responding to an attack and resolution*

How a proponent of a proposal for action responds to an attack depends upon the nature of the attack. For those attacks which explicitly state an alternative position, the original proponent is able to counter-attack with some subset of the attacks listed in Table 1. For example, if a proponent argues for an action on the grounds that this will promote some value *v*, and an attacker argues in response that the proposed action will also demote some other value *w*, then the proponent may respond to this attack by arguing that the action does not have this effect on *w* (Attack 4), or that an alternative action can promote *w*, or that *w* is not worth promoting (Attack 16), etc.

Whether or not two participants may ultimately reach agreement on a proposed action will depend on the participants and on the precise nature of the disagreement. A basis for any resolution between participants for each type of attack is shown in the fourth column of Table 1. Here we identify four different categories of attack which we will now go on to examine individually and discuss how each is used in resolving a dispute.

2.4. *Sources of disagreement*

2.4.1. Factual disagreements. If the disagreement concerns the nature of the current world-state (Attacks 1 and 13), i.e. a dispute about ''What is true'', then some process of agreed empirical investigation may resolve this difference between the participants. The same process would also apply to the resolution of disputes regarding causal relations (Attacks 2 and 4). This may involve the participants entering a sub-dialogue involving a third party outside their own dialogical exchange in order to resolve the dispute through the elicitation of the authoritative knowledge of the third party. Alternatively one of the participants may have a role in the dialogue which entitles the opinion of that party to be authoritative (cf. [23, 28]).

2.4.2. Different preferences. Disputes about ''What is best'' relate to the preferences of the individual participants. Often such disputes arise from participants ranking their preferences differently. Thus, even where there is no dispute for example, as to the possibility of the performance of the action in question, a dispute can still arise if one party believes there to be a better action to perform in the given situation. There may be a number of reasons as to why a participant does not endorse their opponent's action. There may be alternative possibilities which have the same effect of producing the desired results where this alternative is more preferable to a participant (Attacks 5–7). Conversely, an action may have previously unconsidered detrimental side effects, with respect to the goals it achieves and the values promoted by these goals. Thus a participant may propose an alternative action which will not bring about such undesired side effects (Attacks 8–10) and they may prefer the goals and values endorsed by this alternative action. Finally, a participant may deem an action as undesirable if it interferes with other actions in question, with respect to the promotion of another value, previously not considered (Attack 11). The relation of dialectical status to ranking of values can be exploited by representing such arguments in a Value-Based

Argumentation Framework (VAF) [8]. Having identified the differences in values which led to the acceptance or rejection of the argument, disputes must be decided by determining the party whose wishes are to be represented, or by some form of negotiation.

2.4.3. Representation. Disputes which relate to representation issues are concerned with the language being used and the logic being deployed in the argument (Attacks 3, 13–16). Language is intrinsically connected with meaning and understanding; thus, if both parties involved in the dialogue speak the same language and are competent users of an agreed logic, then the resolution of a dispute over representation should be straightforward. One way of ensuring that computer agents share the same language is through access to the same ontologies, such as those used in [9,29,30], to establish the common language of the topic in question.

Our model assumes that such matters of meaning and context are agreed upon by the participants of a dialogue and therefore such attacks concerning representation should not occur frequently in dialogue exchanges. However, these attacks remain possible, especially in systems which permit encounters with unfamiliar or unpredictable agents, and should not be overlooked. There is also evidence that not all human societies use the same rules of logical inference [12,25].

2.4.4. Clarification of a Position. In everyday conversations disputes often arise due to participants making ill-informed assumptions about each other's positions. As conversations progress the players' positions become clearer and more explicit and earlier ill-informed assumptions may be dissolved. However, players may not be aware of their opponents' full position about an issue. If the position is not fully explicit then the players may try to elucidate their opponent's position through questioning, in order to be able to make an attack on it.[3]

2.5. Resolution

Successful resolution of a dispute partially depends upon which of the above types of dispute is encountered. Disputes over facts should be easily resolved if some process of empirical investigation is agreed upon between the participants. Issues of representation should also be easily resolved by agreeing on language and context before the dialogue starts, and by aligning participant's ontologies to ensure a shared understanding of the concepts in the given topic of conversation. Both disagreements about representation and disagreements about facts should be resolved before disagreements about choice can be addressed. Disagreements turning on values are explained using VAFs [8], as can be seen in the examples in [6,7].

Resolution of disputes about what is best typically depends on the context in which the dialogue is taking place. It may be the case that one party is an authority on the matter in question and so this will facilitate resolution. For example, in government issues it is usual for government advisors to find out the facts of the situation, and then ministers make the choices between actions on the basis of these facts.

Naturally, resolution will also occur if one party allows themself to be persuaded that their preference ordering is wrong or they concede to the ordering of their

opponent's preferences. For such dynamic ranking of values see [11]. If agents are able to agree on preferences over actions and over values then they should be able to agree overall. However, if the participants disagree over which value should be promoted in the circumstances (Attacks 9 or 16), then resolution will require agreement between them on a preference ordering over values. Such resolution may require other types of dialogue, and some of these interactions have received considerable attention from philosophers, for example [16,24,26]. A formalism to represent disagreement involving arguments which rely on values is proposed in Ref. 8, and is discussed below.

When there is no authority on the matter to whom an appeal can be made, then we must consider *how* the question of what is best is decided. In considering this question however, we should not overlook the fact that it is always possible for rational disagreement to occur in practice and so we must make allowance for this in our model. It is simply not the case the everyone need make the same choices. Not only may different agents have different desires, but they also may legitimately take different views on what is best. A discussion of rational disagreement is given by Searle [27] and we give a fuller account of how this relates to our model in Ref. 4 .

Due to the fact that we need to account for such differing views and preferences, we therefore need to employ some method for choosing between alternatives. So, after disputes relating to representation and fact have been addressed, we are left with a number of competing arguments to the effect that an action should or should not be performed, each of them deriving their strength from the value they promote or demote. The set of competing arguments suggests that we can use an argumentation framework such as that developed by Dung [10] to resolve factual disagreements. To accommodate the strength of arguments in terms of values, we can use the extension of this framework to accommodate values developed by Bench-Capon [8]. In both Refs. 8 and 10, the use of preferred semantics gives rise to the possibility of different but defensible choices, thus accommodating the possibility of rational disagreement. As we will discuss in Section 6, we have used our theory of persuasion over action to devise a formalism to allow BDI agents augmented to incorporate the use of value functions to reason about proposals for action, based upon our model presented here.

To summarise, successful resolution of a dispute depends upon a number of issues including: the type of dispute encountered; the relationship between the participants; and, their individual preference orderings. But we must also note that our model should and does allow for the possibility of rational disagreement; it is often a difficult task to persuade others to change their ranking of personal values, and thus it is always possible that such arguments will terminate in conflict. Resolution of conflicts may also be achieved by an agreed procedure, such as voting, or the agents may agree to disagree. In summary, where there is no 'right' answer we must always model the possibility of different, but acceptable solutions.

3. The PARMA protocol

In this section we present the syntax of the PARMA *Action Persuasion Protocol* together with an outline of an axiomatic semantics for the protocol. We assume, as in

recent work in agent communications languages [19], that the language syntax comprises two layers: an inner layer in which the topics of conversation are represented formally, and an outer, wrapper, layer comprising locutions which express the illocutionary force of the inner content.

The locutions of the PARMA Protocol are shown in the left-most columns of Tables 2–6. These tables also present the pre-conditions necessary for the legal utterance of each locution under the Protocol, and any post-conditions arising from their legal utterance. Thus, Tables 2–6 present an outline of an axiomatic semantics [31] for the PARMA Protocol, and imply the rules governing the combination of locutions under the protocol [21]. We further assume, following [17] and in accordance with recent work in agent communications, that a *Commitment Store* is associated with each participant, which stores, in a manner which all participants may read, the commitments made by that participant in the course of a dialogue. The post-conditions of utterances shown in Tables 2–6 include any commitments incurred by the speaker of each utterance while the pre-conditions indicate any prior commitments required before an utterance can be legally made. Commitments in this protocol are dialogical – i.e., statements which an agent must defend if attacked, and may not be a true expression of the agent's real beliefs or intentions [17].

4. A denotational semantics

We now outline a denotational semantics for the PARMA protocol, that is a semantics which maps statements in the syntax to mathematical entities [31]. Our approach draws on a branch of category theory, namely topos theory. Our reason for using this, rather than (say) a Kripkean possible worlds framework or a labelled transition system, is that topos theory enables a natural representation of logical consequence ($S \models G$) in the same formalism as mappings between spaces ($R \xrightarrow{A} S$ and $G \uparrow v$). To our knowledge, no other non-categorical denotational semantics currently proposed for action formalisms permits this.

We begin by representing proposals for action. We assume, as in Section 2.1, finite sets of Acts, Propositions, States, Goals and Values, and various mappings. For simplicity, we assume there are n propositions. Each State may be considered as being equivalent to the set of propositions which are true in that State, and so there

Table 2. Locutions to control the dialogue.

Locution	Pre-conditions	Post-conditions
Enter dialogue	Speaker has not already uttered enter dialogue	Speaker has entered dialogue
Leave dialogue	Speaker has uttered enter dialogue	Speaker has left dialogue
Turn finished	Speaker has finished making their move	Speaker and hearer switch roles so new speaker can now make a move
Accept denial	Hearer has made an attack on an element of speaker's position	Speaker committed to the negation of the element that was denied by the hearer
Reject denial	Hearer has made an attack on an element of speaker's position	Disagreement reached

Table 3. Locutions to propose an action.

Locution	Pre-conditions	Post-conditions
State circumstances(R)	Speaker uttered enter dialogue	Speaker committed to R Speaker committed to R ∈ States
State action (A)	Speaker uttered enter dialogue Speaker committed to R Speaker committed to R ∈ States	Speaker committed to A Speaker committed to A ∈ Acts
State consequences(A,R,S)	Speaker uttered enter dialogue Speaker committed to R Speaker committed to R ∈ States Speaker committed to A Speaker committed to A ∈ Acts	Speaker committed to apply (A,R,S) ∈ apply Speaker committed to S ∈ States
State consequences(S,G)	Speaker uttered enter dialogue Speaker committed to R Speaker committed to R ∈ States Speaker committed to A Speaker committed to A ∈ Acts Speaker committed to apply (A,R,S) ∈ apply Speaker committed to S ∈ States	Speaker committed to S ⊨ G Speaker committed to G ∈ Goals
State purpose(G,V,D)	Speaker uttered enter dialogue Speaker committed to R Speaker committed to R ∈ States Speaker committed to A Speaker committed to A ∈ Acts Speaker committed to apply (A,R,S) ∈ apply Speaker committed to S ∈ States Speaker committed to S ⊨ G Speaker committed to G ∈ Goals	Speaker committed to (G,V,D) Speaker committed to V ∈ Values

are 2^n States. We consider the space $\mathscr{C}$ of these States, with some additional structure to enable the representation of actions and truth-values. We consider elements of values to be mappings from Goals to some space of evaluations, called $\mathscr{S}$. This need not be the three-valued set $Sign = \{+, =, -\}$ that we assumed in Section 2.1, although we assume that $\mathscr{S}$ admits at least one partial order. The structures we assume on $\mathscr{C}$, on $\mathscr{S}$ and between them are intended to enable us to demonstrate that these are categorical entities [13]. We begin by listing the mathematical entities, along with informal definitions.

Table 4. Locutions to ask about an agent's position.

Locution	Pre-conditions	Post-conditions
Ask circumstances(R)	Hearer uttered enter dialogue Speaker uttered enter dialogue Speaker not committed to circumstances(R) about topic in question	Hearer must reply with state circumstances(R) or don't know(R)
Ask action(A)	Hearer uttered enter dialogue Speaker uttered enter dialogue Speaker not committed to action(A) about topic in question	Hearer must reply with state action(A) or don't know(A).
Ask consequences(A,R,S)	Hearer uttered enter dialogue Speaker uttered enter dialogue Speaker not committed to consequences(A,R,S) about topic in question	Hearer must reply with state consequences(A,R,S) or don't know(A,R,S)
Ask logical consequences(S,G)	Hearer uttered enter dialogue Speaker uttered enter dialogue Speaker not committed to consequences(S,G) about topic in question	Hearer must reply with state logical logical consequences(S,G) or don't know(S,G)
Ask purpose(G,V,D)	Hearer uttered enter dialogue Speaker uttered enter dialogue Speaker not committed to purpose(G,V,D) about topic in question	Hearer must reply with state purpose(G,V,D) or don't know(G,V,D)

- The space $\mathscr{C}$ comprises a finite collection $\mathscr{C}_0$ of objects and a finite collection $\mathscr{C}_1$ of arrows between objects.
- $\mathscr{C}_0$ includes 2^n objects, each of which may be considered as representing a State. We denote these objects by the lower-case Greek letters, $\alpha, \beta, \gamma, \ldots$, and refer to them collectively as *state objects* or *states*. We may consider each state to be equivalent (in some sense) to the set of propositions which are true in the state.
- $\mathscr{C}_1$ includes arrows between state objects, denoted by lower-case Roman letters, $f, g, h, \ldots$ If f is an arrow from object α to object β, we also write $f : \alpha \rightarrow \beta$. Some arrows between the state objects may be considered as representing actions leading from one state to another, while other arrows are causal processes (not actions of the dialogue participants) which take the world from one state to another. There may be any number of arrows between the same two objects: zero, one or more than one.
- Associated with every object $\alpha \in \mathscr{C}_0$, there is an arrow $1_\alpha \in \mathscr{C}_1$ from α to α, called the identity at α. In the case where α is a state object, this arrow may be considered as that action (or possibly inaction) which preserves the status quo at a state α.
- If $f : \alpha \rightarrow \beta$ and $g : \beta \rightarrow \gamma$ are both arrows in $\mathscr{C}_1$, then we assume there is an arrow $h : \alpha \rightarrow \gamma$. We denote this arrow h by $g \circ f$ (*"g composed with f"*). In other words, actions and causal processes may be concatenated.
- We assume that $\mathscr{C}_0$ includes a special object *Prop*, which represents the finite set of all propositions. We further assume that for every object $\alpha \in \mathscr{C}_0$ there is a monic arrow $f_\alpha : \alpha \rightarrow Prop$. Essentially, a monic arrow is an injective (one-to-one) mapping.

Table 5. Locutions to attack elements of a position.

Locution	Pre-conditions	Post-conditions
Deny circumstances(R)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States	Speaker committed to deny circumstances(R)
Deny consequences(A,R,S)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A Hearer committed to A ∈ Acts Hearer committed to apply(A,R,S) ∈ apply Hearer committed to S ∈ States	Speaker committed to deny consequences(A,R,S) ∈ apply
Deny logical consequences(S,G)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A Hearer committed to A ∈ Acts Hearer committed to apply(A,R,S) ∈ apply Hearer committed to S ∈ States Hearer committed to S⊨G Hearer committed to G ∈ Goals	Speaker committed to deny logical consequences(S,G) S⊨G
Deny purpose(G,V,D)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A Hearer committed to A ∈ Acts Hearer committed to apply(A,R,S) ∈ apply Hearer committed to S ∈ States Hearer committed to S⊨G Hearer committed to G ∈ Goals Hearer committed to (G,V,D) Hearer committed to V ∈ Values	Speaker committed to deny purpose(G,V,D)

- We assume that $\mathscr{C}_0$ has a terminal object, **1**, ie, an object such that for every object $\alpha \in \mathscr{C}_0$, there is precisely one arrow $\alpha \rightarrow \mathbf{1}$.
- We assume that $\mathscr{C}$ has a special object Ω, and an arrow $true : \mathbf{1} \rightarrow \Omega$, called a *sub-object classifier*. The object Ω may be understood as the set comprising $\{True, False\}$.
- We assume that $\mathscr{S}$ is a space of objects over which there is a partial order $<_i$ corresponding to each participant in the dialogue. Such a space may be viewed as a category, with an arrow between two objects α and β whenever $\alpha <_i \beta$. For each participant, we further assume the existence of one or more mappings v between $\mathscr{C}$ and $\mathscr{S}$, which takes objects to objects, and arrows to arrows. We denote the collection of all these mappings by $\mathscr{V}$.

Table 6. Locutions to attack validity of elements.

Locution	Pre-conditions	Post-conditions
Deny initial circumstances exist(R)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R ∈ States	Speaker committed to deny initial circumstances exist(R)
Deny action exists(A)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A ∈ Acts	Speaker committed to deny action exists(A)
Deny resultant state exists(S)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A ∈ Acts Hearer committed to S ∈ States	Speaker committed to deny resultant state exists(S)
Deny goal exists(G)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A ∈ Acts Hearer committed to S ∈ States Hearer committed to G ∈ Goals	Speaker committed to deny goal exists(G)
Deny value exists(V)	Speaker uttered enter dialogue Hearer uttered enter dialogue Hearer committed to R Hearer committed to R ∈ States Hearer committed to A ∈ Acts Hearer committed to S ∈ States Hearer committed to G ∈ Goals Hearer committed to V ∈ Values	Speaker committed to deny value exists(V)

The assumptions we have made here enable us to show that $\mathscr{C}$ is a category [13], and we can thus represent the statement $R \xrightarrow{A} S$, for states R and S, and action A. Moreover, the presence of a sub-object classifier structure enables us to represent statements of the form $S \models G$, for state S and goal G, inside the same category $\mathscr{C}$. This structure we have defined for $\mathscr{C}$ creates some of the properties needed for $\mathscr{C}$ to be a topos [13]. Finally, each space $\mathscr{S}$ with partial order $<_i$ is also a category, and the mappings v are functors (structure-preserving mappings) between $\mathscr{C}$ and $\mathscr{S}$. This then permits us to represent statements of the form $G \uparrow v$, for goal G and value v.

We define a denotational semantics for the PARMA Protocol by associating dialogues conducted according to the Protocol with mathematical structures of the type defined above. Thus, the statement of a proposal for action by a participant in a dialogue

$$R \xrightarrow{A} S \models G \uparrow v$$

is understood semantically as the assertion of the existence of objects representing R and S in $\mathscr{C}$, the existence of an arrow representing A between them, the existence of an

arrow with certain properties[4] between *Prop* and Ω, and the existence of a functor $v \in \mathscr{V}$ from $\mathscr{C}$ to $\mathscr{S}$. Attacks on this position then may be understood semantically as denials of the existence of one or more of these elements, and possibly also, if the attack is sufficiently strong, the assertion of the existence of other objects, arrows or functors.

Thus, our denotational semantics for a dialogue conducted according to the PARMA Protocol is defined as a countable sequence of triples,

$$\langle \mathscr{C}_1, \mathscr{S}_1, \mathscr{V} \rangle, \langle \mathscr{C}, \mathscr{S}, \mathscr{V} \rangle, \langle \mathscr{C}, \mathscr{S}, \mathscr{V} \rangle, \ldots,$$

where the kth triple is created from the kth utterance in the dialogue according to the representation rules just described. Then, our denotational semantics for the PARMA Protocol itself is defined as the collection of all such countable sequences of triples for valid dialogues conducted under PARMA. This approach views the semantics of the protocol as a space of mathematical objects, which are created incrementally and jointly by the participants in the course of their dialogue together. The approach derives from the constructive view of human language semantics of Discourse Representation Theory [18], and is similar in spirit to the denotational semantics, called a *trace semantics*, defined for deliberation dialogues in Ref. 22, and the *dialectical graph* recording the statements of the participants in the Pleadings Game of Gordon [14]. We are currently engaged in specifying formally this denotational semantics in accordance with the outline presented here.

5. Implementation of the dialogue game

We have implemented the PARMA *Action Persuasion Protocol* in the form of a Java program. The program implements the protocol so that dialogues between two human participants can be undertaken under the protocol, with each participant taking turns to propose and attack positions uttering the locutions specified above. The program checks the legality of the participants' chosen moves by verifying that all pre-conditions for the move hold. Thus, the participants are able to state and attack each other's positions with the program verifying that the dialogue always complies with the protocol. If a participant attempts to make an illegal move then they are informed of this and given the opportunity to chose an alternative move. After a move has been legally uttered, the commitment store of the participant who made the move is updated to contain any new commitments created by the utterance. All moves, whether legal or illegal, are entered into the history, which records which moves were made by which participant and the legality of the move chosen. After a move has been legally made, the commitment store of the player who made the move is printed to the screen to show all previous commitments and any new ones that have consequently been added. By publicly displaying the commitment stores in this way each participant is able to see their own and each other's commitments. Thus, participants can determine which of their commitments overlap with those of the other participant, and thereby identify points of agreement. Conversely, this also allows each participant to identify any commitments of the other participant in conflict with their own, and thus which commitments are susceptible to an attack.

Dialogues undertaken via the program can terminate in a number of ways. A participant can decide to leave the game by exiting at any time, thereby terminating the dialogue. A dialogue can also terminate if disagreement about a position is reached. This occurs when a participant states an element of a position which is consequently attacked by the other participant, and the first participant disputes the validity of the attack. If the first participant refuses to accept the reasons for the attack then disagreement has been identified and the dialogue terminates. Dialogues may also reach a natural end with agreement between the two participants on a course of action. If this occurs, both players may choose to exit the dialogue.

When a dialogue terminates, whether in agreement or disagreement, the history and commitment stores of both players are printed on screen and also to a file. The dialogue may then be analysed, for example, to see which attacks occurred, or how often or how successful they were. Such analysis may be useful for a study of appropriate strategies for dialogue conducted under the protocol. Further details of the implementation can be found in Ref. 3.

5.1. Issues raised by the implementation

We are satisfied that this implementation meets our objective of allowing the reconstruction of a wide range of natural arguments concerning a number of topics using our protocol. Thus, our model suffices to give an account of practical argumentation. Our overall aim, however, was to provide computer support to improve the quality of such argumentation. The implementation identified three major obstacles to such improvement, which will become more severe if the intention is to allow autonomous agents to participate in such dialogues:

- Successful conduct of an argument requires considerable goodwill on the part of the participants. The relevance of contributions and the avoidance of fruitless lines of arguments is ensured only by the cooperation of the participants.
- The rules rely on syntactic elements, but such dialogues often turn on semantic and pragmatic features of the utterances.
- The large number of moves available and the fact that they can be deployed at many different stages of the dialogue, means that it is hard to enlist the support of the computer in guiding the moves of the participants. Playing the game is no easier than conducting an argument verbally: thus the problems of quality are not addressed.

These points are also found in other empirical work, such as [36], where often participants were mystified by the effects of the protocol and artefacts of the protocol could be exploited to win the dispute.

Indeed it is this flexibility that presents problems in natural dialogue. Correctly interpreting the force of particular utterances and deciding how best to respond can lead to misunderstandings, arguments at cross purposes, and inefficiencies both in natural dialogue and in its computational representation.

Given these problems, we now believe there is a better way to support the construction of arguments about action than by modelling natural dialogue. Instead, the

insights drawn from a consideration of natural dialogue – the moves that are required and typical patterns of natural dialogues in particular contexts – can be used to provide a tool which instead of attempting to mimic natural dialogue provides a well defined and productive route through a dialogue capable of addressing a specific situation. In this way we hope that misunderstanding of the justification can be minimised, and the most pertinent attacks made.

We have taken this approach in the Persuasive ARguMENt In DEmocracieS (PARMENIDES) system [5]. The specific situation we have chosen is that one where a democratic Government wishes to solicit views on some particular policy. The key features of this situation are:

- It is essential that the initial statement of the particular policy be fully explicit, and unambiguous. This is so that the Government cannot fudge the issues, and so that criticisms are really directed against the policy as it is understood by the Government, rather than some possibly inaccurate interpretation of it,
- Critics must be allowed to make a sufficient range of attacks,
- It must be clear for any criticism, exactly what element of the justification is being objected to.

Parmenides uses a simple web interface to solicit criticisms of a particular policy argument.[5] First the justification is stated in full. This is to give the critic an overview of the justification. Then a succession of screens solicit objections to: the values pursued and alternative values that might be considered; the connections between goals and values; the connections between the consequences of the actions and the goals; the claimed consequences of the actions; any alternative actions claimed to lead to the same ends; and the description of the current fact situation. Each of these points of disagreement represent an attack from our theory, so in this way the critic has the opportunity to attack any part of the justification that they do not agree with in a systematic manner. Attacks relating to the possibility of states of affairs are not supported: it is assumed that the original position will be correct in these respects. Moreover, since there is a database at the backend, the Government could collect a number of such responses and see which parts of its argument found favour and which did not, and the extent to which they did so. However, once the user has submitted their opinions as to why they disagree with the justification presented (if indeed they do disagree with it) they are then given the opportunity to construct their own position on the matter in question. This is again done through the navigation of a succession of screens which ask the user to enter the facts, action(s), consequences, goals and values they believe hold for the issue and this forms a new position to represent their views.

Given a particular situation of intended use, we are satisfied that PARMENIDES is an improved alternative implementation to the Java program, as it overcomes many of the usability problems highlighted earlier in this section whilst being driven by the same model of argument scheme and attacks. In future work we intend to consider how this approach might be adapted to different use situations, including a different selection of attacks.

Despite the shortcomings we have highlighted with the Java implementation of the dialogue game, we believe that implementing it has proved very useful as we have

shown that our general theory of persuasion can be conducted via computer mediated dialogues of this form. This implementation has also raised a number of interesting issues in relation to our underlying argumentation schema, which we have addressed in our current work that enables BDI agents to use our model of persuasion over action and this work is summarised briefly in the next section.

6. Extension of PARMA for Use in BDI Agents

As stated above, in Ref.1 we have gone on to show how our theory can be made computational within the framework of an agent based on the BDI model. Current BDI architectures [35] do not use the notion of values, and so we have extended the architecture to include values which provide justifications for the agent's choice of intentions, based upon its beliefs and desires. Here we assume that the agent has a set of beliefs and a set of desires, in the standard way for a BDI agent. We add to this a set of *value functions*, one for each value recognised by the agent, which takes a desire as argument and returns a real number x such that $-1 \leq x \leq 1$. Positive values of x indicate a degree of promotion of the value represented by the satisfaction of the desire and negative values of x represent the degree of demotion of the value represented by the satisfaction of the desire. Thus desires include both states of affairs which are desired to be true and states of affairs which are desired to be false. It is the value function that distinguishes them.

The normal BDI intention-selection process is that the agent first generates a set of options given its beliefs and desires, and then filters this set of candidates to select its intentions. In our model corresponding to the generation of options we generate a set of presumptive arguments for actions, and attacks which can be used against these arguments. Note that these attacks can themselves be couched in the form of arguments. To perform the filtering we form these arguments into a VAF in the manner of [8] and determine the preferred extension for our agent, using the ordering of values chosen by that agent as required. This preferred extension will form the set of intentions of the agent.

In Ref. 1 we give the definitions for how an agent can construct a position based upon its beliefs about the world, the set of actions available for performance, the agent's desires, and its values. The agent can then construct a justification of its position about an action it is proposing. We then go on to specify a full set of preconditions for the execution of the attacks in our theory. The pre-conditions for each individual attack must be met in order for an opposing agent to make the attack on the first agent's position. In Refs. 6 and 7, we provide example applications in the domains of medicine and law to show how this formalism can be successfully used by BDI agents augmented with value functions to reason about proposals for action, in accordance with our theory. These applications show both the general applicability of our approach and how it must be realised differently in different domains. Such differences are important: they help to explain why the implementation of the unadorned Java program described above seems so unsupportive in concrete cases of use.

This computational use of our model in BDI agents will be the focus of our future work in which we hope to further develop and demonstrate the usefulness of this

approach in enabling agents to reason and argue rationally about proposals for action.

7. Concluding remarks

This paper has presented the syntax and semantics for a novel agent dialogue game protocol for argument over proposals for action. The protocol, called the PARMA *Action Persuasion Protocol*, implements our theory of persuasion over action developed in Refs. 2, 4 and 15, which presents a general argument schema for the advocacy and justification of actions, and so supports rational discourse over proposed courses of actions. We have described two implementations based on the underlying theory and have discussed the merits of each. We have also given a brief summary of how our current work extends the theory to enable it to be deployed in a BDI agent system. The further development and application of this will be the focus of our future work.

Acknowledgements

Katie Atkinson is grateful for support from the EPSRC. Trevor Bench-Capon and Peter McBurney acknowledge partial support received from the European Commission, through Project ASPIC (IST-FP6-002307). This paper is a revised and extended version of a paper presented at the First International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2004), held at AAMAS 2004 in New York City, NY, USA in July 2004. We thank the anonymous referees and the participants at that workshop for their comments.

Notes

1. Stephen Toulmin's book entitled *"Knowing and Acting"* [32], for example, has 18 chapters on beliefs, and one on actions. Walton, whose work we build on in this paper, is a notable exception.
2. Although Deliberation rather than Persuasion dialogues in the revised typology of [33].
3. This is discussed by Walton and Krabbe in [34] as dark-side commitments.
4. This arrow is the characteristic function for the object representing G, and the properties are that a certain diagram commutes in $\mathscr{C}$.
5. A prototype implementation of an example debate can be seen at `http://www.csc.liv.ac.uk/~katie/Parmenides.html`

References

1. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. "Attacks on a presumptive argument scheme in multi-agent systems: pre-conditions in terms of beliefs and desires", Technical Report ULCS-04-015, Department of Computer Science, University of Liverpool, UK, 2004.
2. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney, "Computational representation of persuasive argument", Technical Report ULCS-04-006, Department of Computer Science, University of Liverpool, UK, 2004.

3. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney, "Implementation of a dialogue game for persuasion over action", Technical Report ULCS-04-005, Department of Computer Science, University of Liverpool, UK, 2004.
4. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. "Justifying practical reasoning". In F. Grasso, C. Reed, and G. Carenini, (eds.), *Proceedings of the Fourth International Workshop on Computational Models of Natural Argument (CMNA 2004)*, Valencia, Spain, pp. 87–90, 2004.
5. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney, "Parmenides: Facilitating democratic debate", In R. Traunmüller, (ed.), *Electronic Government 2004*, Lecture Notes in Computer Science 3183, Springer, Berlin, pp. 313–316, 2004.
6. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney, "Arguing about cases as practical reasoning", In *Proceedings of the 10th International Conference on Artificial Intelligence and Law (ICAIL 2005). ACM Press*, New York, USA. In Press.
7. K. M. Atkinson, T. J. M. Bench-Capon, and S. Modgil, "Value added: Processing information with argumentation", Technical Report ULCS-05-004, Department of Computer Science, University of Liverpool, UK, 2005.
8. T. J. M. Bench-Capon, "Persuasion in practical argument using value based argumentation frameworks", *J. Logic Comput*, vol.13, no.3, pp. 429–48, 2003.
9. R-J. Beun and R. M. van Eijk, "A co-operative dialogue game for resolving ontological discrepancies", In F. Dignum, (ed.), *Advances in Agent Communication*, Lecture Notes in Artificial Intelligence 2922, Springer, Berlin, Germany, pp. 349–363, 2004.
10. P. M. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games", *Artif Intelligence*, vol.77 pp. 321–357, 1995.
11. P. E. Dunne and T. J. M. Bench-Capon, "Identifying audience preferences in legal and social domains", In F. Galindo, M. Takizawa, and R. Traunmüller, (eds.), *Proceedings of the 16th International Conference on Database and Expert Systems Applications (DEXA 2004)*, Lecture Notes in Computer Science, 3180, Springer Verlag, Berlin, Germany, pp. 518–527, 2004.
12. P. Gárdenfors, "The role of expectations in reasoning", In M. Masuch and L. Pólos, (eds.), *Knowledge Representation and Reasoning under Uncertainty: Logic at Work*, Lecture Notes in Artificial Intelligence 808, Springer, Berlin, Germany, pp. 1–16, 1994.
13. R. Goldblatt,"*Topoi: The Categorial Analysis of Logic*. North-Holland", Amsterdam, The Netherlands, 1979.
14. T. F. Gordon, "The pleadings game: An exercise in computational dialectics", *Artif. Intelligence and Law*, 2 vol. pp. 239–292, 1994.
15. K.M. Greenwood, T.J.M. Bench-Capon, and P.M. McBurney. "Towards a computational account of persuasion in law", In *Proceedings of the 9th International Conference on AI and Law (ICAIL-2003)*, New York, NY, USA, 2003. ACM Press, pp. 22–31, 2003.
16. J. Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press: Cambridge, MA, USA, 1996. (Translation by W. Rehg).
17. C. L. Hamblin, *Fallacies*. Methuen: London, UK, 1970.
18. H. Kamp and U. Reyle. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, vol. 2 Kluwer Academic: Dordrecht, The Netherlands, 1993.
19. Y. Labrou, T. Finin, and Y. Peng, "Agent communication languages: The current landscape". *IEEE Intelligent Syst*, vol. 14, no.2, pp. 45–52, 1999.
20. M. Luck, P. McBurney, and C. Preist, *Agent Technology: Enabling Next Generation Computing. A Roadmap for Agent Based Computing*. AgentLink II, Southampton, UK, 2003.
21. P. McBurney and S. Parsons, "Games that agents play: A formal framework for dialogues between autonomous agents," *J. Logic, Lang Inf*.,vol. 11, no.3, pp. 315–334, 2002.
22. P. McBurney and S. Parsons. "A denotational semantics for deliberation dialogues", In N. R. Jennings, C. Sierra, E. Sonenberg, and M. Tambe, (eds.), *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, pp. 86–93, 2004.
23. N. Oren, T. Norman, A. Preece, and S. Chalmers, "Policing virtual organisations. In C. Ghidini, P. Giorgini, and W. van der Hoek, (eds.), *Proceedings of the 2nd European Conference on Multi Agent Systems (EUMAS 2004)*, Barcelona, Spain, pp. 499–508, 2004.

24. C. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation*, University of Notre Dame Press: Notre Dame, IN, USA, 1969.
25. D. Raven, ''The enculturation of logical practice'', *Configurations*, vol.3, pp. 381–425, 1996.
26. H. S. Richardson, *Practical Reasoning about Final Ends*. Cambridge University Press: Cambridge, UK, 1994.
27. J. R. Searle, *Rationality in Action*. MIT Press: Cambridge, MA, USA, 2001.
28. C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons, ''A framework for argument based negotiation'', In A. Rao M. Singh and M. Wooldridge, (eds.), *Intelligent Agents IV*, Lecture Notes in Artificial Inteligence 1365, Springer, Berlin, Germany, pp. 177–192, 1998.
29. V. Tamma, I. Blacoe, B. Lithgow-Smith, and M. Wooldridge. ''Serse: Searching for semantic web content'', In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-04)*, Valencia, Spain, 2004.
30. V. A. M. Tamma and T. J. M. Bench-Capon. ''A conceptual model to facilitate knowledge sharing in multi-agent systems. In *Proceedings of Autonomous Agents 2001 Workshop on Ontologies in Agent Systems (OAS 2001)*, Montreal, Canada, pp. 69–76, 2001.
31. R. D. Tennent, *Semantics of Programming Languages*, Prentice-Hall: Hemel Hempstead, UK, 1991.
32. S. E. Toulmin. *Knowing and Acting: An Invitation to Philosophy*, Macmillan: New York, NY, USA, 1976.
33. D. N. Walton, *The New Dialectic: Conversational Contexts of Argument*, University of Toronto Press: Toronto, Ontario, Canada, 1998.
34. D. N. Walton and E. C. W. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press: Albany, NY, USA, 1995.
35. M. J. Wooldridge, *Reasoning about Rational Agents*. MIT Press: Cambridge, MA, USA, 2000.
36. T. Yuan, *Human Computer Debate, a Computational Dialectics Approach*. PhD thesis, Leeds Metropolitan University, Leeds, UK, 2004.