KATIE ATKINSON, TREVOR BENCH-CAPON
and PETER McBURNEY

COMPUTATIONAL REPRESENTATION OF PRACTICAL ARGUMENT

ABSTRACT. In this paper we consider persuasion in the context of practical reasoning, and discuss the problems associated with construing reasoning about actions in a manner similar to reasoning about beliefs. We propose a perspective on practical reasoning as presumptive justification of a course of action, along with critical questions of this justification, building on the account of Walton. From this perspective, we articulate an interaction protocol, which we call *PARMA*, for dialogues over proposed actions based on this theory. We outline an axiomatic semantics for the *PARMA* Protocol, and discuss two implementations which use this protocol to mediate a discussion between humans. We then show how our proposal can be made computational within the framework of agents based on the Belief-Desire-Intention model, and illustrate this proposal with an example debate within a multi agent system.

1. INTRODUCTION

Practical reasoning is reasoning about what is best or most sensible to do. Despite the fact that such reasoning occurs on a common basis in the conduct of activities in the everyday life of most people, this type of reasoning has not been studied within computer science or philosophy nearly as extensively as reasoning about beliefs. In this paper, we seek to provide some foundations for practical reasoning by proposing an account of argument over proposed actions which is readily embodied in a protocol for a formal dialogue game. Although our model reflects aspects of human practical reasoning it is not intended to be prescriptive or descriptive for human reasoning. Our model will provide support for decision making in multi agent systems. Our account is based on the use of an argumentation scheme and associated critical questions, and builds on the approach of Walton (1996).

The paper is structured as follows. Section 2 discuss practical reasoning and highlights problems with the practical syllogism, a

Synthese (2006) 152: 157–206
Knowledge, Rationality & Action 191–240
DOI 10.1007/s11229-005-3488-2

traditional method of viewing reasoning about actions as akin to reasoning about beliefs. This section then goes on to examine and extend a particular argument scheme of Walton who gives an account of practical reasoning as presumptive justifications and critical questions. Section 3 articulates the extended schema and the additional critical questions of our framework in more detail and this forms the basis for a dialogue game protocol we call *PARMA*. We then present, in Section 4, a summary of the *PARMA* Protocol syntax and outline an axiomatic semantics for the protocol. Section 5 discusses two different implementations of the protocol designed to mediate a discussion between humans: one is a Java program representing a dialogue game in the tradition of the philosophy of argumentation, (e.g., MacKenzie 1979); the other is an online discussion forum realised using web technology. This is a step towards our ultimate goal of allowing persuasive dialogue between autonomous software agents. Section 6 presents an analysis of responses to the different types of critical questions and of the resolution of disputes. Section 7 sketches a proposal for generating our presumptive arguments and attacks from a BDI agent augmented with value functions and we illustrate this with an example. This section then shows how a value based argumentation framework (Bench-Capon 2003) can be used to filter options to decide on a course of action in the context of a multi-agent system. Section 8 offers some conclusions and possible extensions for future work. Finally, in the appendix we present an outline of an axiomatic semantics for the *PARMA* Protocol.

2. PRACTICAL REASONING

2.1. *The Importance of Practical Reasoning*

Much research in Artificial Intelligence (AI) has focused on mechanisms to enable artificial entities to reason about beliefs about the world. AI traditionally however, involves more than this. Since its earliest days it has also been concerned with artifacts capable of acting so as to modify their environment. Indeed, it could be argued that intelligence requires such an ability: that intelligence can only be manifested in behaviour. The recent growth of interest in software agent technologies (e.g., Wooldridge 2000), puts action at the centre of the stage. For software agents to have the capability of interacting intelligently with their environment they also need to be equipped with an ability to reason about what actions are the best to execute in given situations. In other words, intelligent agents need to be able to undertake practical reasoning. The most common response to this challenge has been to use some variant of the practical syllogism. In the next section we consider in detail the particular problems associated with the practical syllogism, first from the perspective of philosophy and then as seen in agent systems.

2.2. *Difficulties with the Practical Syllogism*

2.2.1. *In the Context of Philosophy*

Within philosophy, practical reasoning has been a topic of attention since at least the time of Aristotle. Recent discussions include collections of essays (Millgram 2001; Raz 1978) and a book by Searle (2001). Most of this work has taken as its starting point a version of the practical syllogism. Here is a typical example, taken from Kenny (1975):

K1 I'm to be in London at 4.15.
If I catch the 2.30, I'll be in London at 4.15
So, I'll catch the 2.30

Although Aristotle presented practical reasoning as a deduction, it has proved difficult to maintain that position (e.g., Anscombe's essays on the topic, Raz 1978) and this abductive form is now normally used. A problem is that it is possible to accept both of the premises yet deny the conclusion, based on at least three points of criticism:

C1 K1 represents a species of abduction, and so there may be alternative ways of achieving the goal.

C2 Performing one action typically excludes the performance of other actions, which might have other desirable results; these may be more desirable than the stated goal.

C3 Performing an action typically has a number of consequences in addition to the explicitly stated goal. If some of these are undesirable, they may be sufficiently bad to lead us to abandon the goal.

In order to act on the basis of an argument such as K1, therefore, we need to consider alternative actions, alternative goals and any additional consequences, and then choose the *best* of these alternative goals and actions. Note the element of choice here: we can choose which of our goals we will seek to realise, and which actions to undertake to realise these goals. We do not have such choice with regard to our beliefs. Given complete information, the world restricts us to a single rational choice of beliefs, but different people may rationally make different choices of goals and actions. We are not driven by our desires: we can resist them. And whereas the way the world is lies beyond our control, we can at least (to some extent) choose the way the world will be.

Given this element of choice therefore, practical argument is directed to a specific audience (in the sense of Perleman, 1969) at a specific time, to encourage them towards a particular choice of goals and/or actions; the objectivity that we can find in factual matters cannot in general be attained in practical reasoning. An attempt to modify K1, similar to one put forward by Searle (2001) (although not regarded by him as satisfactory) is:

S1 I want, all things considered, to achieve E
The best way, all things considered, to achieve E is to do M
So, I will do M.

The two different "all things considered" qualifications are supposed to deal with alternative desires and methods of achieving them. The "best" addresses the selection from the available options. However, this too presents problems: we cannot in general consider all things, because we have limited reasoning resources and imperfect information. Nor is it easy to say what is meant by "best" here. In computer science there are often attempts to define best using some kind of utility function but, as Searle points, out any preference ordering is more often the *product* of practical reasoning than an input to it. Coming to understand what we think is best is part of what we do in practical reasoning. This issue is discussed further in Section 6.5.

2.2.2. *Limitations in Computer Science and Agent Systems*

Searle's form of the practical syllogism given above can be applied to the reasoning mechanisms used in autonomous agents in order to equip them with the ability to reason about what it is best to do in a given situation. The standard view of the justification of an action in this context can be generally seen as:

PS1 Agent P wishes to realise goal G
If P performs action A, G will be realised
Therefore, P should perform A.

This view underlies well-known architectures, such as Belief-Desire-Intention (BDI) models (Wooldridge 2000), used in the reasoning mechanisms of autonomous agents. Because the BDI model has a number of proposed realisations, we will, when we need to be specific, take as our model the popular Procedural Reasoning System (PRS) (Georgeff and Lansky 1987).

The process of reasoning about action is described by Wooldridge (2000) as "the Deliberation Process" and this process is broken down into two phases: option generation and filtering. During the option generation phase the decision-making agent generates a set of possible alternative actions available for execution. These alternative options are generated by taking the agent's current beliefs and current intentions and applying the reasoning scheme of PRS to see which goals can now be pursued. These form the set of desires of the agent. Thus, the agent's desires correspond to the goals that it wishes to realise, though it may be the case that not all desires are achievable or not achievable together. In order to achieve these desires the agent must form a plan from the repertoire it holds in a pre-programmed plan library and check that the pre-conditions for executing this plan are satisfied by the agent's current beliefs about the world. This results in the agent building sets of actions (or plans) each of which would enable it to achieve one of its desires. The agent can now move on to the filtering phase where it simply chooses the "best" option to commit to from this set through the use of a filter function. The "best" option will typically be chosen through the application of some pre-existing utility function, and then added to the intentions of the agent.

Thus, an agent using the BDI model is able to address some of the difficulties associated with the practical syllogism highlighted above in the following ways:

– The agent has a repertoire of plans held in a finite plan library and this enables it to consider everything available to it that is relevant to the decision.
– The agent is able to define which action is the *best* one to take as it has a utility function, or some other filtering criterion, to enable it to compare potential outcomes of actions.
– Any undesirable side effects, brought about as a consequence of performing an action, cannot be considered as ruling out the plan as the agent's plan library should contain only plans approved by the designer, recognising these side effects.

While this approach provides a pragmatic resolution of the issues appropriate to some agent systems, it provides a less satisfactory solution to the general problems associated with practical reasoning. By its nature, the process of practical reasoning is open-ended and this in turn poses problems for its use in agent technology, as described above. Agents operate with a limited repertoire of plans and a fixed utility function and so the designer necessarily takes responsibility for pre-determining the options available to the agent. The agent can consider only the options it has been given, not "all things". Even with the autonomy afforded to agents, constraints are made upon the plans the agent will consider and find acceptable, as filtering of alternative plans is undertaken by means of a fixed utility function over goals supplied in advance by the designer. Because practical reasoning is intrinsically open-ended, unforeseen alternatives and consequences may arise, and revision of preferences may occur, at any time. This creates a challenge for agent design, which must, by its nature, make assumptions which circumscribe the considerations possible to the agent.

2.3. *Walton's Account of Practical Reasoning*

One way of addressing such problems with the practical syllogism is to regard practical reasoning as a species of presumptive argument. Given an argument like S1, we have a presumptive reason for performing the action. This presumption can, however, be challenged and withdrawn. Subjecting our argument to appropriate challenges is how we hope to identify and consider the alternatives that require consideration, and determine the best choice for us, in the particular context. Because the challenges are, in principle open ended, the process of justification does not end, and discussion can always be re-opened.

One account of presumptive reasoning is in terms of argument schemes and critical questions, as given by Walton (1996) and also discussed by Verheij (2003). The idea here is that an argument scheme gives a presumption in favour of its conclusion. Whether this presumption stands or falls depends on satisfactory answers being given to any critical questions associated with the scheme that are posed in the particular situation.

Walton (1996) gives two schemes for practical reasoning: the necessary condition scheme (called W1):[1]

W1 G is a goal for agent *a*
Doing action A is necessary for agent *a* to carry out goal G
Therefore agent *a* ought to do action A.

and the sufficient condition scheme (W2):

W2 G is a goal for agent *a*
Doing action A is sufficient for agent *a* to carry out goal G
Therefore agent *a* ought to do action A.

Walton associates with them four critical questions:

CQ1 Are there alternative ways of realising goal G?

CQ2 Is it possible to do action A?

CQ3 Does agent *a* have goals other than G which should be taken into account?

CQ4 Are there other consequences of doing action A which should be taken into account?

Here we will consider only W2: W1 is a special case in which CQ1 is answered in the negative. CQ1, CQ3 and CQ4 relate respectively to the criticisms C1, C2 and C3 identified above. We believe, however, that this argument scheme, and the critical questions need elaboration because the notion of a goal is ambiguous, potentially referring indifferently to any direct results of the action, the consequences of those results, and the reasons why those consequences are desired. We believe those distinctions to be important. Consider the following situation. I am in Liverpool. My friend X is currently in London (200 miles distant) and is about to go to Australia indefinitely. I am eager to say farewell to him. To catch him before he leaves London, it is necessary that I arrive in London before 4.30 pm. Note that practical reasoning is situated and it is therefore important to know the story behind the situation in order to be able to consider all the alternatives available in the particular context. So I may say:

AS1 I want to be in London before 4.30 pm.
The 1.30 pm train arrives in London at 4.15 pm.
So, I shall catch the 1.30 train.

Here I am justifying my action in terms of one of its consequences. Alternatively I may say:

AS2 I want to see person X before he leaves London.
The 1.30 pm train arrives in London at 4.15 pm.
So, I shall catch the 1.30 pm train.

Here the action is not justified by its direct consequences, but by something else that follows from them. I do not really desire to be in London at all, except in so far as it is a means to the end of seeing X before he departs for Australia. Alternatively there is a third justification:

AS3 Friendship requires that I see person X before he leaves London.
The 1.30 pm train arrives in London at 4.15 pm.
So, I shall catch the 1.30 pm train.

Now I justify my action not in terms of its direct consequences, nor in terms of a state of affairs which will result from the action, but in terms of the underlying social value which I hope to promote by the action.

Thus, we have taken Walton's notion of a goal and separated it into three distinct elements: states, goals and values. In our model we define *states* to be a set of propositions about the world to which we can assign a truth value, *goals* are propositional formulae on this set of propositions, and *values* are functions on goals. We make the distinction between states and goals to represent the important difference between effects of actions which the agent wishes attain, and the effects which follow from an action but are not necessarily desired by the agent. Looking to values, these in turn are different from goals as they provide the actual reasons for which an agent wishes to achieve a goal. Thus, we view values as being distinct from goals and not just sub or super goals. The importance of a distinct notion of 'value' has been discussed in previous work on practical reasoning with Perelman (1980) and Searle (2001) being two notable authors on the issue. Perelman and Searle both argue that values account for the fact that different people may rationally disagree upon an issue. Moreover, it is possible for two people (or agents) to agree upon a goal to be achieved, but the reasons for which they want to achieve it may be very different due to their contrasting value sets. Such use of argument based on values has also been featured in other computational work such as Bench-capon and Sartor (2003), Grasso et al. (2000), Gordon and Karacapilidis (1997) and Jarke et al. (1987). We believe that the distinction between the different aspects described above is an important factor in practical reasoning situations where the precise points of contention on an issue should be distinguishable and identifiable. This provides the motivation for our extension of Walton's scheme in this manner.

In general, instead of Walton's

W1a G is a goal for agent *a*

we may write

P1 Agent *a* wishes to achieve state S so as to bring about goal G which promotes a value V.

Note that the answers to CQ1 are different in the cases AS1-3:

- In the case of AS1, I must propose other ways of arriving in London on time, perhaps by driving;
- In the case of AS2 I need not go to London at all; for example I could drive to Heathrow Airport and say goodbye there;
- In the case of AS3 I need not meet with person X at all; perhaps a telephone call and an apology will be enough to promote friendship.

Given this more refined notion of a goal we can extend CQ1 to:

CQ1a Are there alternative ways of realising the same consequences?

CQ1b Are there alternative ways of realising the same goal?

CQ1c Are there alternative ways of promoting the same value?

We can also elaborate CQ3, in that it may be that doing action A realises some other goal which promotes some other value, or it may be that doing A prevents some other goal from being realised:

CQ3a Would doing action A promote some other value?

CQ3b Does doing action A preclude some other action which would promote some other value?

Also CQ4 has two aspects:

CQ4a Does doing action A have a side effect which demotes the value V?

CQ4b Does doing action A have a side effect which demotes some other value?

Secondly, apart from the possibility of the action, Walton does not consider other problems with soundness of W2, presupposing that the second premise is to be understood in terms of what an agent knows or reasonably believes. In Greenwood et al. (2003), we proposed an argument scheme which incorporates P1 and makes the factual context explicit:

G1 In the circumstances R
we should perform action A
to achieve new circumstances S
which will realise some goal G
which will promote some value V.

It could be that:

– Action A is not sufficient to bring about goal G; either because the current circumstances are not as presupposed, or because, although the beliefs about the current situation are correct, action A does not have the believed effects.
– Goal G is not a goal for agent *a*; either because there is some problem with the link between the circumstances brought about by doing action A with the value agent *a* assumes them to promote, or because goal G is not in fact a possible state of affairs.

We can therefore add the critical questions:

CQ5 Are the circumstances such that doing action A will bring about goal G?

CQ6 Does goal G promote value V?

CQ7 Is goal G possible?

Note that an answer to CQ5 needs to address four issues:

(a) Whether the believed circumstances R are possible.
(b) Whether the believed circumstances R are true.

(c) Assuming both of these, whether the action A has the stated consequences S.
(d) Assuming all of these, whether the action A will bring about the desired goal G.

Similarly, if we take the more articulated view of G expressed as P1, CQ6 needs to address both:

(a) Whether goal G does realise the value intended; and
(b) Whether the value proposed is indeed a legitimate value.

Also, taking G in terms of P1, CQ7 needs to address both:

(a) Whether the situation S believed by agent *a* to result from doing action A is a possible state of affairs,
(b) Whether the particular aspects of situation S represented by G are possible.

We thus have an elaborated set of critical questions: four variants of CQ5; three variants of CQ1; two variants of each of CQ3, CQ4, CQ6 and CQ7; and CQ2, making sixteen questions in all. We will use these sixteen questions as the basis for the development of our general theory of persuasion over action to be presented in Section 4.

3. MAKING THE CRITICAL QUESTIONS PRECISE

In this section we will revisit our argument scheme and attempt to make the sixteen critical questions identified in Section 2 more precise, by giving relatively formal definitions of them. We will then give formal semantics in Section 4. The specific situation that we consider is where one agent is attempting to persuade another to adopt a course of action, and that other agent is arguing against this. Because we see this situation as one of conflict, we will refer to the various critical questions as “attacks”. Persuasion is intended to be rational, and so reasons are advanced, and attacked, by each side. Moreover, persuasion is intended to lead to action, so the debates are examples of practical reasoning.

We will also consider a number of variants on the basic attacks. When an element of a position is disputed, the attacker may simply disagree, or may additionally offer extra information which indicates the source of the disagreement or makes the disagreement more concrete. Thus, for example, if there is a disagreement as to what is in fact the current situation, an opponent may simply deny what the proponent has said, or may also add what he or she thinks is really the case.

The theory proposed here forms the basis for a dialogue game protocol named the *PARMA Action Persuasion Protocol*.

3.1. *Stating a Position*

In Section 2 we gave the following as the general schema for a position motivating an action (Schema G1):

G1 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.

We need recognise no difference between deciding 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 G1 for each goal realised and value promoted. We may then formalise the Schema as follows. 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 a truth value from the set $\{T, F\}$ to every element of *Props*.
- A finite set of propositional formulae, *Goals*, called goals, with elements G, H, etc.
- A finite set of values *Values*, with elements v, w, etc.
- A function *value* mapping each element of *Goals* to a pair $\langle v, sign \rangle$, 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 argument schema G1 contains reference to actions and deontic modalities which are not readily formalised in classical logic. We can, however, see that there are four statements of classical logic which must hold if the argument represented by schema G1 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: $value(G) = \langle v, + \rangle$.

3.2. *Attacking a Position*

In this subsection we will describe the attacks corresponding to the critical questions of Section 2.3 in terms of the elements identified in Section 3.1. We will group them in a slightly different manner, in order to emphasize different connections between the attacks. This will also show the relationships between the attacks and each of the four elements of the statement of a position G1 in Section 3.1; in each case, we will also identify the source critical question.

3.2.1. *Denial of Premises*

A proposal for a particular action A can first be attacked by denying one of the four statements which must obtain for the proposal to be valid. Three of these premises relate to the action realising the goal, and so relate to Critical Question CQ5, whereas as one concerns the realisation of the claimed value and so relates to CQ6.

Attack 1 (CQ5b): R is not the case.

Attack 2 (CQ5c): It is not the case that $apply(A, R, S) \in apply$.

Attack 3 (CQ5d): It is not the case that $S \models G$.

Attack 4 (CQ6a): It is not the case that $value(G) = \langle v, + \rangle$.

Each of these attacks may be executed with differing degrees of force, depending on whether positive information accompanies the attack, and the severity of the consequences of disagreement, and so we are able to distinguish variants of the main attack. Consideration of later elements presupposes agreement on earlier elements of a position for a proposal for action. For example, unless there is agreement on the current circumstances, the effects of an action will not be considered. This is also effected in the implemented dialogue game, as discussed later in Section 5.

We can identify two variant attacks for **Attack 1**.

Attack 1a: R is not the case.

Attack 1b: R is not the case, and there is a circumstance $Q \in \text{States}$, where $R \neq Q$, such that Q is the case.

We can likewise identify seven variant attacks for **Attack 2**.

Attack 2a: It is not the case that $apply(A, R, S) \in apply$.

Attack 2b: It is not the case that $apply(A, R, S) \in apply$, and it is the case that $apply(A, R, T) \in apply$, where $T \neq S$.

Attack 2c: It is not the case that $apply(A, R, S) \in apply$, and it is the case that $apply(A, R, T) \in apply$, where $T \neq S$, but it is not the case that $T \models G$.

Attack 2d: It is not the case that $apply(A, R, S) \in apply$, and it is the case that $apply(A, R, T) \in apply$, where $T \neq S$, and it is the case that $T \models G$, but it is not the case that $value(G) = \langle v, + \rangle$.

Attack 2e: It is not the case that $apply(A, R, S) \in apply$, and it is the case that $apply(A, R, T) \in apply$, where $T \neq S$, and it is the case that $T \models G$, but $value(G) = \langle v, - \rangle$.

Attack 2f: It is not the case that $apply(A, R, S) \in apply$, and it is the case that $apply(A, R, T) \in apply$, where $T \neq S$, and it is the case that $T \models G$, but $value(G) = \langle w, + \rangle$, where $w \neq v$.

Attack 2g: It is not the case that $apply(A, R, S) \in apply$, and it is the case that $apply(A, R, T) \in apply$, where $T \neq S$, and it is the case that $T \models G$, but $value(G) = \langle w, - \rangle$, where $w \neq v$.

Similarly, we may distinguish six variants of **Attack 3**:

Attack 3a: It is not the case that $S \models G$.

Attack 3b: It is not the case that $S \models G$ and there is a goal $H \in Goals$, $H \neq G$, such that $S \models H$.

Attack 3c: It is not the case that $S \models G$ and there is a goal $H \in Goals$, $H \neq G$, such that $S \models H$ and with $value(H) \neq \langle v, + \rangle$.

Attack 3d: It is not the case that $S \models G$ and there is a goal $H \in Goals$, $H \neq G$, such that $S \models H$ and with $value(H) = \langle v, - \rangle$.

Attack 3e: It is not the case that $S \models G$ and there is a goal $H \in Goals$, $H \neq G$, and a value $w \in Values$, $w \neq v$, such that $S \models H$ and with $value(H) = \langle w, + \rangle$.

Attack 3f: It is not the case that $S \models G$ and there is a goal $H \in Goals$, $H \neq G$, and a value $w \in Values$, $w \neq v$, such that $S \models H$ and with $value(H) = \langle w, - \rangle$.

Likewise, we may distinguish four variants of **Attack 4**:

Attack 4a: It is not the case that $value(G) = \langle v, + \rangle$.

Attack 4b: It is not the case that $value(G) = \langle v, + \rangle$ and $value(G) = \langle v, - \rangle$.

Attack 4c: It is not the case that $value(G) = \langle v, + \rangle$ and there is a value $w \in Values$, $w \neq v$, such that $value(G) = \langle w, + \rangle$.

Attack 4d: It is not the case that $value(G) = \langle v, + \rangle$ and there is a value $w \in Values$, $w \neq v$, such that $value(G) = \langle w, - \rangle$.

3.2.2. *Alternative Ways to Satisfy the Same Value*

These four attacks all relate to Critical Question CQ1, in that they each propose an alternative way of achieving the same desired value. Note that Attack 7b does not of itself dispute that the action should be performed, nor that the value will be promoted. Its significance comes when the discussion concerns the justification of a past action which is taken as a precedent for some future action. This becomes important in, for example, legal applications, as discussed in Greenwood et al. (2003).

Attack 5 (CQ1a): There exists an action $B \in Acts$, with $B \neq A$, and $apply(B, R, S) \in apply$.

Attack 6 (CQ1b): There exists an action $B \in Acts$, with $B \neq A$, and $apply(B, R, T) \in apply$, with $T \models G$.

Attack 7a (CQ1c): There exists an action $B \in Acts$, with $B \neq A$, and $apply(B, R, T) \in apply$, with $T \models H$, and $value(H) = \langle v, + \rangle$.

Attack 7b (CQ1c): There is a goal $H \in Goals$, with $H \neq G$, such that $apply(A, R, S) \in apply$ with $S \models H$, and with $value(H) = \langle v, + \rangle$.

3.2.3. *Side Effects of the Action*

Two of these attacks relate to unconsidered consequences of the action, raised by Critical Question CQ4. The third offers a different justification for the action, and so relates to other goals that need to be considered, as in Critical Question CQ3.

Attack 8 (CQ4a): There is a goal $H \in Goals$, with $H \neq G$, such that $apply(A, R, S) \in apply$ with $S \models H$, and with $value(H) = \langle v, - \rangle$.

Attack 9 (CQ4b): There is a goal $H \in Goals$, with $H \neq G$, and there is a value $w \in values$, with $w \neq v$, such that $apply(A, R, S) \in apply$ with $S \models H$, and with $value(H) = \langle w, - \rangle$.

Attack 10 (CQ3a): There is a goal $H \in Goals$, with $H \neq G$, and there is a value $w \in values$, with $w \neq v$, such that $apply(A, R, S) \in apply$ with $S \models H$, and with $value(H) = \langle w, + \rangle$.

3.2.4. *Interference with Other Actions*

This group of attacks all relate to the promotion of some other value, and so derive from Critical Question CQ3b. The three variants arise respectively from: consideration of the compatibility of the proposed action with some other action; whether the proposed action realises a state of affairs incompatible with the goal of another action; or whether the state of affairs realised is incompatible with *all* ways of promoting some other value.

Attack 11a: It is the case that $apply(A, R, S) \in apply$. There is a value $w \in values$ with $w \neq v$. There is an action $B \in Acts$ with $B \neq A$, such that $apply(B, R, T) \in apply$, with $T \models H$, and $value(H) = \langle w, + \rangle$. However, there is no state $X \in States$ such that $apply(A\&B, R, X)^2 \in apply$.

Attack 11b: It is the case that $apply(A, R, S) \in apply$. There is a value $w \in values$ with $w \neq v$. There is a goal $H \in Goals$, such that $value(H) = \langle w, + \rangle$. However, $S \models \neg H$.

Attack 11c: It is the case that $apply(A, R, S) \in apply$. There is a value $w \in values$ with $w \neq v$. However, if there is a goal $J \in Goals$, with $value(J) = \langle w, + \rangle$, then $S \models \neg J$.

3.2.5. *Disagreements Relating to Impossibility*

The final group of attacks all relate to whether an element of the position is possible or not. In the critical questions we considered possibility together with the other questions relating to the element under dispute. Therefore these attacks relate to a number of different critical questions, as indicated below.

Attack 12 (CQ2): It is not the case that $A \in Acts$.

Attack 13 (CQ5a): It is not the case that $R \in States$.

Attack 14 (CQ7a): It is not the case that $S \in States$.

Attack 15 (CQ7b): It is not the case that $G \in Goals$.

Attack 16 (CQ6b): It is not the case that $v \in Values$.

We can summarise our attacks and their relation to the critical questions in Table I. The last column will be discussed in Section 6.

TABLE I

Nature of Conflict for Critical Questions

Attack	Characterisation	Critical question	Dispute to be resolved
1	Disagree with the description of the current situation	CQ5b	What is true
2	Disagree with the consequences of the proposed action	CQ5c	What is true
3	Disagree that the desired features are part of the consequences	CQ5d	Representation
4	Disagree that these features promote the desired value	CQ6a	What is true
5	Believe that the consequences can be realised by some alternative action	CQ1a	What is best
6	Believe that the desired features can be realised through some alternative action	CQ1b	What is best
7	Believe that the desired value can be realised in an alternative way	CQ1c	What is best
8	Believe that the action has undesirable side effects which demote the desired value	CQ4a	What is best
9	Believe that the action has undesirable side effects which demote some other desired value	CQ4b	What is best
10	Agree that the action should be performed, but for different reasons	CQ3a	What is best
11	Believe that the action will preclude some more desirable action	CQ3b	What is best
12	Believe that the action is impossible	CQ2	What is true
13	Believe that the circumstances as described are not possible	CQ5a	Representation
14	Believe that the consequences as described are not possible	CQ7a	Representation
15	Believe that the desired features cannot be realised	CQ7b	Representation
16	Disagree that the desired value is worth promoting	CQ6b	Representation

4. THE PARMA PROTOCOL

In Section 3 we laid the foundations for a multi-agent dialogue game protocol, called *PARMA*, which enables persuasive argument over proposed courses of action to be undertaken by two or more participants. A proponent of an action may state and justify his or her proposal for action in the form of Schema G1, and opponents may attack this position according to the 16 different attacks presented in Section 3. In Section 4, we now outline the syntax of the *PARMA* protocol, along with an outline of an axiomatic semantics for it.

4.1. *Syntax of PARMA*

In this section we present the syntax of the *PARMA* protocol. We assume, as in recent work in agent communications languages (Labrou et al. 1999), that the language syntax comprises two layers: an inner layer in which the topics of conversation are represented formally, and an outer, or wrapper, layer comprising locutions which express the illocutionary force of the inner content. In our presentation of the axiomatic semantics we assume classical propositional logic as the formal representation of the inner layer, but this restriction is for simplicity of presentation only.

We present the syntax of *PARMA* by listing the twenty-five legal locutions in Tables II and III, grouped into five classes. Fifteen locutions are shown in Table II, grouped into three classes (columns): locutions to control the dialogue; locutions to state a position for the justification of an action; and locutions to ask about an agent's position.

Table III contains another ten locutions, grouped into two classes (columns): locutions to attack elements of a position; and locutions to attack the validity of elements of a position.

TABLE II

Locutions to control the dialogue, ask about a position and state a position

'Control' Locutions	'Ask' Locutions	'State' Locutions
Enter dialogue	Ask circumstances(R)	State circumstances(R)
Leave dialogue	Ask action(A)	State action(A)
Turn finished	Ask consequences(A, R, S)	State consequences(A, R, S)
Accept denial	Ask logical consequences(S, G)	State logical consequences(S, G)
Reject denial	Ask purpose(G, V, D)	State purpose(G, V, D)

TABLE III

Locutions to attack a position and attack the validity of elements

‘Deny’ Locutions	‘Deny Existence’ Locutions
Deny circumstances(R)	Deny initial circumstances exist(R)
Deny action(A)	Deny action exists(A)
Deny consequences(A, R, S)	Deny resultant state exists(S)
Deny logical consequences(S, G)	Deny goal exists(G)
Deny purpose(G, V, D)	Deny value exists(V)

The detailed pre-conditions for the performance of all these locutions and the post-conditions which occur upon their utterance can be found in Appendix A.

Locutions for Attacks

The attacks listed in Section 3.2 can be realised in the *PARMA* protocol by means of one or more of the 25 primitive locutions listed in Tables II and III. If more than one locution is required to be uttered for an attack to be realised, the order of utterance is not important. As illustration, we show which locutions are required for four attacks:

- Location for Attack 1a: Deny circumstances(R)
- Locutions for Attack 1b: State circumstances(Q) AND Deny circumstances(R)
- Location for Attack 2a: Deny consequences(A, R, S)
- Locutions for Attack 2b: State consequences(A, R, T) AND Deny consequences(A, R, S).

The full list of locutions used to realise each of the attacks of Section 3.2 can be found in Appendix B.

4.2. *Axiomatic Semantics for* PARMA

An axiomatic semantics presents the pre-conditions necessary for the legal utterance of each locution under the protocol, and any post-conditions arising from their legal utterance. We assume, following Hamblin (1970) and in accordance with most work on dialogue games (e.g., MacKenzie (1979) and recent work in agent communications, that a *Commitment Store* is associated with each participant. These stores hold, in a manner which all participants may read, the commitments made by that participant in the course of a dialogue. The pre-conditions of legal utterances indicate any prior commitments required before the utterance can be legally made, and the post-conditions of utterances include any commitments incurred by the speaker upon that utterance. Commitments in this protocol are dialogical – i.e., statements which an agent must defend in the dialogue if attacked – and may bear no relation to the agent's real beliefs or intentions (Hamblin 1970). We thus make no assumptions about the private mental states of the agents involved in the dialogue. In addition to the presentation of the axiomatic semantics given in the appendix, they are also discussed further in Atkinson et al. (2005b).

In the dialogue game built on these semantics (discussed in the next section and in further detail in Atkinson et al. 2005b), once a move has legally been executed by a player, the turn can be passed, where the next player then has a set of moves from which the choice of the next utterance may be made. These next available moves are entirely defined by the pre-conditions of the locutions. This means that checking the pre-conditions for the legality of moves ensures that the dialogue is sensibly structured and that irrelevant or inappropriate utterances cannot be made legally during the course of the dialogue. In related work we have supplemented the axiomatic semantics for PARMA with an outline of a denotational semantics for the protocol, the details of which can be found in Atkinson et al. (2005b).

5. IMPLEMENTATION

We have realised the *PARMA* protocol in two entirely different implementations; one in the form of a conventional dialogue game designed to mediate an exchange between human participants implemented in the Java programming language, the other in the form of an online discussion forum named *PARMENIDES*, implemented using MySQL and PHP scripts. We now discuss each of these implementations in turn.

5.1. *General Dialogue Game*

Our first implementation of the *PARMA Action Persuasion Protocol* is in the form of a Java program intended to provide a proof of concept. The program implements a version of the protocol so that dialogues between two human participants can be undertaken under the protocol mediated by the system, with each participant taking turns to propose and attack positions uttering the locutions specified in Section 4.1 above. A further discussion of this implementation can be found in Atkinson et al. (2005b) though we summarise the main aspects of the program below:

- The program verifies that the dialogue always complies with the protocol by checking the legality of the participants' chosen moves. This verification ensures that all pre-conditions, expressed in terms of the commitment stores and dialogue history, for the move hold.
- Attempts by participants to make illegal moves are identified and the program provides them with the opportunity to choose an alternative move.
- The program updates the commitment stores of participants, once a legal move has been made, to contain any new commitments incurred by the utterance.
- The program maintains a history which records all attempted moves, whether legal or illegal, made by each participant.
- After a legal move has been made, the commitment store of the player who successfully executed 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. Similarly, such display 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 the source of disagreement about a position is identified. This occurs when a participant states an element of a position which is consequently attacked by the other participant, and the first participant refuses to accept the reasons for the attack. Resolution of such disagreements will require entry into a different dialogue game e.g. inquiry. Dialogues may also reach

a natural end with agreement between the two participants on a course of action. (Note that resolution of disagreement is outside of the scope of this initial system, which provides only passive mediation. Resolution of disagreement is discussed in Section 6.)
– 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 dialogues conducted under the protocol.

Implementing the dialogue game has proved to be a very useful way of evaluating the protocol, as it meets our goal of providing a proof of concept by showing that our general theory of persuasion can be conducted via computer mediated dialogues of this form. This implementation has however also raised a number of interesting issues in relation to our underlying argumentation scheme. Below we summarise the three main insights which have arisen through our evaluation of the implemented dialogue game protocol:

1. The system, acting as referee, cannot use pre-conditions based on mental states of the participants: it infers these from the moves the players make. This means that the pre-conditions to allow a move may be different from those required to sincerely make a move.
2. Natural dialogue is very flexible. Giving support to interactions modelled on natural language utterances requires constraints, and what constraints are appropriate depends on context and purpose. The protocol may impose too few constraints to allow scope for useful computer support.
3. Goodwill and some co-operation is required to make sensible progress and this is again due to the fact that natural language dialogue is so flexible. Thus, uncooperative players can abuse the protocol to stultify the interaction.

5.2. *Dialogue Tailored to Particular Context of Use*

After reflecting on the issues raised in the previous section regarding the Java implementation of *PARMA*, we concluded that the implementation poses many problems for casual users of the system. In order to correctly follow the protocol, users require prior knowledge of the underlying theory of persuasion. Without such knowledge they will be unable to recognise which locutions need to be chosen in order to realise the correct attack, in a given situation. Users must also be familiar with the names and meanings of the locutions used to represent the statement and denial of a position. As well as these usability problems, as we mentioned in the previous section, the dialogue game does rely somewhat on the goodwill of the players to use the protocol sensibly, as legal moves may still be unhelpful and unconstructive.

Some of these problems have arisen due to the amount of freedom of expression afforded by the program and this leaves users with a great variety of options to select between. All these problems concerning usability of the program are obviously undesirable. Therefore, we have addressed these issues by going on to implement our theory of persuasion in an entirely different format and using a particular domain – eDemocracy.

The particular system we have built complements recent research on the application of information technologies to support democratic participation and debate. Systems such as ZENO (Gordon and Karacapilidis 1997) and DEMOS (Lührs et al. 2003) aim to assist citizens to communicate with one another and with public officials over matters of community concern and to do so in a dialogue possibly involving multiple simultaneous parties.

We have implemented a second version of the protocol in the form of an online discussion forum, named *PARMENIDES* (for Persuasive ArguMENt In DEmocracieS), which allows a much simpler form of interaction to take place. Users are led through a particular path appropriate to the specific application in order to elicit their views on a particular topic, in accordance with our theory. The user interaction occurs through a simple web based interface which guides them in a structured fashion through a justification of an action, giving opportunities to disagree at selected points. Each of these disagreements represents one of the attacks from our theory of persuasion, so the exact nature of the disagreement can be unambiguously identified. By constraining the choice of the user in such a way, we eliminate the need for them to understand the underlying argumentation scheme and to select the correct moves. The responses of the users are written to a database so we are able to gather and analyse the information in order to identify what elements of an argument are more strongly supported than others. An extended motivation is given in Atkinson et al. (2004).

This system has been successfully implemented. Given a particular situation of intended use, we are satisfied that it is an improved alternative implementation to the Java program, as it overcomes many of the usability problems highlighted in Section 5.1. We have also extended the program to provide for the construction of positive alternative arguments. We intend also to consider how this approach might be adapted to different use situations, including a different selection of attacks. Details of the *PARMENIDES* online discussion forum and a particular application can be found in Atkinson et al. (2004).[3]

6. RESPONDING TO ATTACKS

Now that the statement of a position and the criticism of the elements of such a position have been defined we examine the ways in which the recipient of an attack can respond to their opponent's criticism. Note that this may require leaving the persuasion dialogue to enter a dialogue of a different type, described as 'nesting' in Walton and Krabbe (1995).

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 I in Section 3. 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 (Attack 7), 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 relationship between 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 I. We will now examine each individual basis for resolution, discussing the precise nature of the dispute and how resolution of the dispute could be reached.

6.1. *Factual Disagreements*

If the disagreement concerns the nature of the current world-state (Attacks 1 and 12), 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, perhaps 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. Sierra et al. 1998).

6.2. *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, there is no dispute as to the possibility of the performance of, for example, the action in question, but a dispute can arise due to one party believing the action not to be the best one to perform in the given situation. As mentioned in Section 2, there may be a number of reasons as to why a participant does not endorse their opponent's action. There may be alternative possible actions which have the same effect of producing the desired results and this alternative action may be more preferable to a participant (Attacks 5, 6 and 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 (Attacks 8, 9 and 10). 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). In such cases, disputes must be decided by determining the party whose wishes are to be represented, by constructing a preference order (Doutre et al. 2005) or by some form of negotiation.

6.3. *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, 14, 15 and 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 and concepts is through commitment to the same ontologies, to establish the common language of the topic in question. Ontological differences and their resolution are discussed in Beun and van Eijk (2004), Tamma and Bench-Capon (2001) and Tamma et al. (2004).

Our model assumes that such matters of meaning and context are agreed upon by the participants of a dialogue beforehand 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.

6.4. *Clarification of a Position*

A common cause of disputes in everyday conversations is that participants make 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 recognise that that they are not aware of their opponents' full position about an issue. If the position is not fully explicit then the players may have to elucidate their opponent's position through questioning in order to be able to make an attack on it.

6.5. *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 participants' 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 are addressed.

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 thus this will facilitate resolution. For example, in government issues it is usual for government advisors to find out the facts of the situation and for ministers to make the choices between possible actions on the basis of these facts in the light of the ministers' values. The advisors are then authorities as to facts, as the ministers are authorities as to values. Similarly in a court case, juries are authoritative as to facts, while the role of the judge is to choose legal interpretations.

Naturally, resolution will also occur if one party allows himself to be persuaded that his preference ordering is wrong or he concedes to the ordering of his opponent's preferences. 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 by the action, then resolution may 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, Habermas 1996; Perelman and Olbrechts-Tyteca 1969 and Richardson 1994).

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. Two phenomena need to be respected: the possibility of rational disagreement, and value preferences emerging from the reasoning. As to rational disagreement, 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. As Searle puts it:

> Assume universally valid and accepted standards of rationality, assume perfectly rational agents operating with perfect information, and you will find that rational disagreement will still occur; because, for example, the rational agents are likely to have different and inconsistent values and interests, each of which may be rationally acceptable. [Searle 2001, p. xv]

With regard to emerging values, although many current agent systems use a general utility function, Searle also observes:

> This answer, [*that an audience can provide a ranking for goals*] though acceptable as far as it goes, mistakenly implies that the preferences are given prior to practical reasoning, whereas, it seems to me, they are typically the product of practical

> reasoning. And since ordered preferences are typically products of practical reason, they cannot be treated as its universal presupposition. [Searle 2001, p. 253]

If Searle is right, and intuitively it seems more plausible than arguing that all people make their selections according to pre-existing utility functions, this too needs to be accounted for. Therefore, we 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 could use an argumentation framework such as that developed by Dung (1995) 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 (2003). How this may be achieved is discussed in the next section; note that the resolution of disputes about choice can be resolved using Value-Based Argumentation Frameworks, as shown in Dunne and Bench-Capon (2004). In both, Dung (1995) and Bench-Capon (2003), the use of preferred semantics gives rise to the possibility of different but defensible choices, thus accommodating the possibility of rational disagreement. These issues have also been addressed in systems designed to mediate human to human dialogues, such as Brewka and Gordon (1994), Gordon and Karacapilidis (1997), Jarke et al. (1987) and Lührs et al. (2003). Doutre et al. (2005) address the second issue and define a dialogue which allows value preferences to emerge from the dialogue.

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 such arguments could terminate in conflict. The way in which attacks are resolved is highly context dependant. A detailed example of how this is applied in the domain of law is given in Atkinson et al. (2005a). This paper gives pre-conditions for instantiations of the argument scheme, posing the critical questions, arranging them into an argumentation framework and evaluating their dialectical status.

7. VALUES AND THE BDI ARCHITECTURE

As stated earlier in the paper, one of the main motivations of this work is to build a computational system based upon the preceding model we have presented. In this section we will sketch how our approach to practical reasoning can be used by autonomous agents within the framework of the Belief-Desire-Intention model. Here the agents will themselves generate arguments and critical questions. Current BDI architectures do not use the notion of values, and so we extend the architecture to include values which provide justifications for the agent's choice of intentions, based upon its beliefs and desires. 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 the critical questions/attacks which can be used against these arguments. Note that these critical questions can themselves be couched in the form of arguments. To perform the filtering we form these arguments into a Value Based Argumentation Framework in the manner of Bench-Capon (2003) and determine the preferred extension for our agent, using the ordering of values chosen by that agent as required. This preferred extension (as shown in Bench-Capon (2003), there is always a unique, non-empty, preferred extension, given an ordering on values) will form the set of intentions of the agent.

Each agent will have two belief predicates, BelA(Formula, Beliefs) and BelS(Formula, Beliefs). BelA(Formula, Beliefs) will be true if Formula is consistent with the agent's Beliefs; that is, it can be made true by making assumptions which would make appropriate assignments to the elements which the agent neither believes true nor believes false. Such assumptions could be unrestricted, or might need to pass some kind of plausibility test: this is local to the implementation of the agent. BelS(Formula, Beliefs) is true only if the beliefs of the agent are such that Formula is true without assumptions.

The agent will also have a set of actions, and beliefs about the pre-conditions of each action and the consequences of performing that action.

Now consider an agent j with beliefs B_j. Suppose that this agent has available an action A for which it believes the pre-conditions to be R_{ja}, and that after performing the action its beliefs will be S_{ja}. Further suppose that its desires include D_{jg}, which is satisfied if G is true, and that V_j is included in its value functions.

Now if $\text{BelA}(R_{ja},B_j)\ \&\ \text{BelA}(G,S_{ja})\ \&\ V_j(D_{jg}) > 0$ holds the agent will have a presumptive argument for performing A, which could be expressed in the form of G1:

G1 In the current circumstances R_{ja}
we should perform action A
to achieve new circumstances S_{ja}
which will realise some goal G
which will promote some value V_j.

We may express our various attacks in a similar fashion. Those attacks which we will use in the example later in the section are given here. The detailed definitions for stating a position and attacking it can be found in Atkinson et al. (2005a). Attack 1b may be made by agent k if not $\text{BelA}(R_{ja},B_k)$ (since some element of the preconditions for A is believed by agent k not to hold). The attack may be expressed as an argument of the form, *"The following assumptions are false: F"*, where F is the set of elements of R believed by agent k to be false. Similarly, attack 9 can be made by agent k if it can produce the appropriate presumptive argument. That is: if

$$\text{BelA}(P_{ja},B_k)\ \&\ \text{BelA}(H,S_{jb})\ \&\ W_k(D_{kh}) < 0$$

where H is a desire distinct from the original G, and W a value distinct from the original V. To make attack 11a, agent k must be able to construct a presumptive argument for an action B, distinct from A, and also show that B is incompatible with A, for example by showing that the post-conditions for A entail that the pre-conditions for B are unsatisfied and vice versa.

7.1. *Example: Treatment of Heart Disease*

We now present a detailed illustration of this approach, adapted from an example of Sanjay Modgil on treatment of heart disease (Modgil and Fox 2004). The action to be chosen by the decision-making agent concerns the appropriate treatment for a patient threatened by blood clotting. The choice of actions is between:

– *Administer-aspirin*: pre-conditions are that the patient has high platelet adhesion. The post-conditions are that platelet adhesion is low and that expense is medium.
– *Administer-chlopidogrel*: pre-conditions are that the patient has high platelet adhesion. The post-conditions are that platelet adhesion is low and that expense is large.
– *Do Nothing*: There are no pre-conditions, but the post conditions are that platelet adhesion is high and that expense is small.

Desires are: (a) that the patient has reduced blood clotting, which is satisfied if blood clotting is less than high, which promotes the value of safety; and (b) that expense is as small as possible, promoting the value of cost. Note that expense is both a direct result of action, and a desire. We also consider dangerous acidity levels, which will result if there is a history of gastritis and no proton pump inhibitor is available, as a desire related to safety, even though it is a negative desire.

The value functions, shared by all agents, are:

$$\text{cost(expense(large))} = -1$$
$$\text{cost(expense(medium))} = 0$$
$$\text{cost(expense(small))} = 1$$
$$\text{safety(reduced blood clotting)} = 1$$
$$\text{safety(dangerous acidity levels)} = -1$$

There are three agents with beliefs as follows:

Agent	platelet-adhesion (high)	platelet-adhesion (medium)	platelet-adhesion (low)	history of gastritis	proton pump inhibitor available
Jay	True	False	False	Unknown	Unknown
Kay	Unknown	Unknown	Unknown	True	Unknown
El	Unknown	Unknown	Unknown	Unknown	True

For ease of reference, we call our agents, “Jay,” “Kay” and “El.” Agent Kay may now begin the discussion by proposing that nothing is done. She instantiates the argument scheme G1 by proposing the following argument:

Argument A1

R1: Assuming that platelet-adhesion is low,
A1: we should do nothing,
S1: which will leave the situation unchanged,
G1: and this has small expense,
V1: which promotes the value of cost.

Agent Jay will attack this with attacks 1b, 9 and 11a. The latter two instantiate the argument scheme as shown in A3 and A4. 1b is a factual disagreement and does not follow this scheme as it simply states what is disputed and any alternatives believed, as shown in A2:

Argument A2

Your assumption that platelet adhesion is low, is false. Platelet adhesion is high.

Argument A3

R3: As platelet adhesion is high,
A3: we should not do nothing,
S3: since this will result in high platelet adhesion,
G3: so that blood clotting is not reduced,
V3: which will demote the value of safety.

Argument A4

R4: As platelet adhesion is high,
A3: we should administer aspirin,
S4: since this will result in low platelet adhesion,
G3: so that blood clotting is reduced,
V3: which will promote the value of safety.

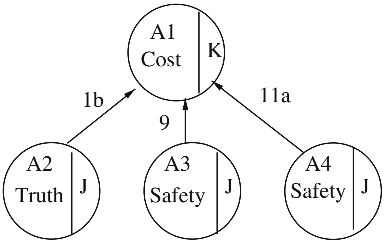

Figure 1.

This situation can be depicted as a Value-Based Argumentation Framework (Bench-Capon 2003) as shown in Figure 1. In this diagram and the ones which follow, nodes represent arguments and the arrows between them represent attacks; the head of the arrow indicates the argument being attacked, and the tail indicates the attacking argument. Nodes are labelled with the name of the argument they represent (A1, A2, etc), along with the value which that argument promotes (*Cost, Safety*, etc.) and a letter on the right hand side representing the agent introducing the argument. Arcs are labelled with the attack that is being made by the argument. Note that argument A2 is a factual argument, and thus promotes the value *Truth*. This representation follows that of Bench-Capon (2003), where the value *truth* is ranked as the most important value by all audiences.

A2 will be preferred to A1 as truth is always the most highly ranked value. The preferred extension is thus {A2, A3, A4}, suggesting that aspirin should be administered. Agent Kay may now, however, make attacks of her own on the presumptive arguments put forward by Jay. A4 may be attacked using attack 9 to instantiate the argument scheme.

Argument A5

R5: Since there is a history of gastritis and assuming no proton pump inhibitor is available,

A5: we should not administer aspirin,

S5: as this will result in dangerous acidity levels,

G5: which would risk ulceration,

V5: which will demote the value of safety.

The resulting argumentation framework is shown in Figure 2.

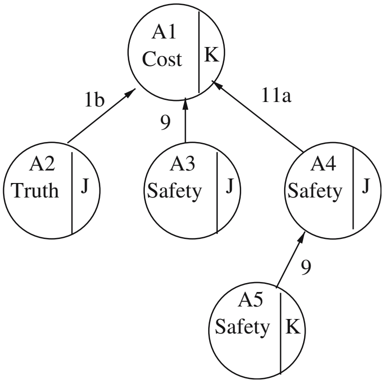

Figure 2.

Now the preferred extension contains (A2, A3, and A5). Unfortunately this leaves us unable to do nothing, but also unable to administer aspirin. Jay can now attack A4 also using attack 5, to suggest an alternative to aspirin.

Argument A6

R6: As platelet adhesion is high,

A6: we should administer chlopidogrel,

S6: since this will result in low platelet adhesion,

G6: so that blood clotting is reduced,

V6: which will promote the value of safety.

A6 proposes an alternative, but incompatible, course of action to that proposed in A4. Such attacks will always be mutual, in that A4 can equally be seen as an attack of type 5 on A6. Represented directly this would give rise to a two-cycle in the same value, but in Bench-Capon (2003), cycles in the same value are excluded. Recall, however, from Section 6.2 that such attacks are to be resolved through preferences over the actions proposed. We represent this by directing the attack by the less preferred argument through an argument stating the preference between the two actions, with the value *choice*, giving rise to a three cycle, $A6 \Rightarrow (A4 > A6) \Rightarrow A4 \Rightarrow A6$. As proven in Bench-Capon (2003), when we have a three cycle in which two values are common and one is distinct, the preferred extension will be the argument with the distinct value and one of the others, whatever the relative ranking of the values. If we rank choice as the lowest value for all audiences (as seems consistent with our treatment of truth), this will mean that if none of the arguments are defeated from outside the cycle, we will accept the preference and the preferred argument, but if the preferred argument is defeated from outside, we will only have the other argument, which will itself defeat the preference. In our current framework attacks on preferences do not arise. Suppose El is considered the authority with respect to choice of actions. Now El may give argument A6a. Again, this is not an instantiation of the scheme, choices are merely stated.

Argument A6a

Aspirin is preferable to chlopidogrel.

As chlopidogrel is also an expensive drug, however, attack 9 can be directed at it giving argument 7, which all agents can make:

Argument 7

R7: The cost of chlopidogrel is high.
A7: we should not administer chlopidogrel,
S7: as this will result in large expense,
G7: which will exceed our budget,
V7: which will demote the value of cost.

This produces the argumentation framework as shown in Figure 3.

Now we find that A2, A3 and A5 are all preferred, but we have a choice between A6 and A7 depending on whether safety or cost is to be preferred. Note that if we choose safety A6a is defeated by A6. Fortunately, agent El is able to resolve this difficult choice between values by attacking argument 5 with attack 1b, stating that a proton pump inhibitor is available thus, administering aspirin will not produce dangerous acidity levels. Again, this is a factual disagreement and so does not instantiate the usual argument scheme.

Argument 8

Your assumption, that there is no proton pump inhibitor available, is false. A proton pump inhibitor is available.

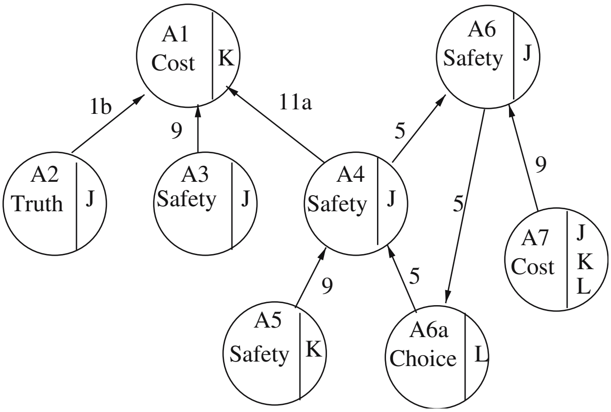

Figure 3.

This gives the final framework as shown in Figure 4.

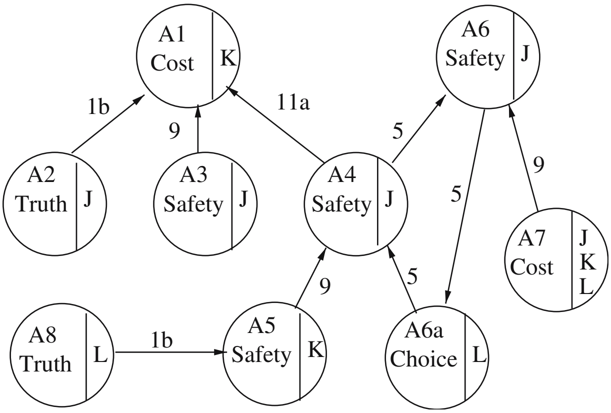

Figure 4.

Now, A4 is reinstated because A5 is defeated. This gives {A2, A3, A4, A6a, A7, A8} as the only the preferred extension whether we prefer safety to cost or not, given that we respect El's choice and prefer aspirin. The action administer-aspirin will thus emerge as the intention of our agents, now that the history of gastritis and the availability of the proton pump inhibitor are known.

In this section we have sketched, through an example, how we may generate presumptive arguments and attacks for a BDI agent augmented with value functions. We have also shown how a Value-based Argumentation Framework can be used to filter options to produce an intention in the context of a multi-agent system, using the principles for resolving attacks described in Section 6.

8. CONCLUSIONS

In this paper we have considered practical reasoning and the challenges associated with this reasoning. We examined the problems of the practical syllogism in the context of philosophy and also considered the limitations of incorporating the practical syllogism into computational systems, and in particular, BDI agents. Following Walton (1996), we have proposed a perspective on practical reasoning as presumptive justification and critical questions. We have extended and made formal one of Walton's argument schemes to give us a richer model. In particular, we have unpacked Walton's notion of a goal into its objective and subjective components.

We have gone on to present a general theory of persuasion in practical reasoning, and used this to articulate a protocol, *PARMA*, for a multi-agent dialogue game based on this theory. We have outlined an axiomatic semantics for *PARMA* as well as two implementations based on the protocol. This work has also drawn our attention to the importance of the context in which such a protocol is useful. One line of future work will be to explore the *PARMA* Protocol in different specific contexts. We also note that formalisms for representing actions and their effects have received a great deal of attention in AI, for example, the situation calculus (McCarthy and Hayes 1969). We hope to explore the connections between these formalisms and our approach in future work. Also, we have thus far excluded from our schema any consideration of: time and temporal factors; uncertainty of consequences; or obligations and moral arguments. We hope to consider these various issues in future development of the *PARMA* protocol.

We believe that the account of practical reasoning which we have given here offers a solid justification of a method by which reasoning about actions is enabled and that this can be applied to models of reasoning used in certain agent architectures, such as the BDI model. This led us to provide a protocol and two different implementations of the model proposed, which provide valuable insights into the way in which practical reasoning can be used by autonomous agents. In our approach, practical reasoning has two aspects: forming critical questions and resolving them. Supporting the formation of critical questions is often enough for mediated debate, as in the main form of *PARMA*. However, resolution is necessary for multi-agent systems. We have sketched here an approach to resolution and shown how our model can be used in BDI agents; this was done by extending the BDI architecture to include the notion of values. This has enabled us to make use of Value-based Argumentation Frameworks in the filtering process of a BDI agent in order to form the agent's intentions. The immediate focus of our future work will be to fully formalise our model of attacks for use in BDI agents.

ACKNOWLEDGEMENTS

Katie Atkinson is grateful for support from the UK Engineering and Physical Sciences Research Council (EPSRC). Trevor Bench-Capon and Peter McBurney acknowledge financial support received from the European Commission's Information Society Technologies (IST) programme, through Project ASPIC (IST-FP6-002307). We are also grateful to Paul Dunne for fruitful discussions and comments on Value Based Argumentation Frameworks.

APPENDIX

A. AXIOMATIC SEMANTICS FOR THE PARMA PROTOCOL

In this appendix we present an outline of an axiomatic semantics for the *PARMA Action Persuasion Protocol*. We assume, as in recent work in agent communications languages (Labrou et al. 1999), 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 IV–VIII. 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 IV–VIII present an outline of an axiomatic semantics (Tennent 1991) for the *PARMA* Protocol, and imply the rules governing the combination of locutions under the protocol (McBurney and Parsons 2002). We further assume, following Hamblin (1970) and in

TABLE IV

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

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 IV–VIII 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 – ie, statements which an agent must defend if attacked, and may not be a true expression of the agent's real beliefs or intentions (Hamblin 1970).

TABLE V

Locutions to Propose an Action

Locution	Pre-conditions	Post-conditions
State circumstances(R)	Speaker uttered enter dialogue.	Speaker committed to R. Speaker committed to $R \in States$.
State action(A)	Speaker uttered enter dialogue. Speaker committed to R. Speaker committed to $R \in States$.	Speaker committed to A. Speaker committed to $A \in Acts$.
State consequences(A, R, S)	Speaker uttered enter dialogue. Speaker committed to R. Speaker committed to $R \in States$. Speaker committed to A. Speaker committed to $A \in Acts$.	Speaker committed to $apply(A, R, S) \in apply$. Speaker committed to $S \in States$.
State logical consequences(S, G)	Speaker uttered enter dialogue. Speaker committed to R. Speaker committed to $R \in States$. Speaker committed to A. Speaker committed to $A \in Acts$. Speaker committed to $apply(A, R, S) \in apply$. Speaker committed to $S \in States$.	Speaker committed to $S \models G$. Speaker committed to $G \in Goals$.
State purpose(G, V, D)	Speaker uttered enter dialogue. Speaker committed to R. Speaker committed to $R \in States$. Speaker committed to A. Speaker committed to $A \in Acts$. Speaker committed to $apply(A, R, S) \in apply$. Speaker committed to $S \in States$. Speaker committed to $S \models G$. Speaker committed to $G \in Goals$.	Speaker committed to (G, V, D). Speaker committed to $V \in Values$.

TABLE VI

Locutions to Ask About an Agent's Position

Location	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 logical consequences(S, G) about topic in question.	Hearer must reply with state 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).

TABLE VII

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.

TABLE VII

Continued

Deny purpose(G, V, D)	Speaker uttered enter dialogue. Hearer uttered enter dialogue. Hearer committed to R. Hearer committed to $R \in States$. Hearer committed to A. Hearer committed to $A \in Acts$. Hearer committed to $apply(A, R, S) \in apply$. Hearer committed to $S \in States$. Hearer committed to $S \models G$. Hearer committed to $G \in Goals$. Hearer committed to (G, V, D). Hearer committed to $V \in Values$.	Speaker committed to deny purpose(G, V, D).

TABLE VIII

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).

B. LOCUTIONS FOR ATTACKS IN THE *PARMA PROTOCOL*

The full set of locutions used to realise each of the attacks from the *PARMA Protocol* in Section 4 is given below.

Attack 1a: deny circumstances(R).

Attack 1b: deny circumstances(R) AND state circumstances(Q).

Attack 2a: deny consequences(A, R, S).

Attack 2b: state consequences(A, R, T) AND deny consequences (A, R, S).

Attack 2c: state consequences(A, R, T) AND deny consequences (A, R, S) AND deny logical consequences(T, G).

Attack 2d: state consequences(A, R, T) AND state logical consequences(T, G) AND deny purpose(G, V, D+) AND deny consequences(A, R, S).

Attack 2e: state consequences(A, R, T) AND state logical consequences(T, G) AND state purpose(G, V, D−) AND deny consequences(A, R, S).

Attack 2f: state consequences(A, R, T) AND state logical consequences(T, G) AND state purpose(G, W, D+) AND deny consequences(A, R, S).

Attack 2g: state consequences(A, R, T) AND state logical consequences(T, G) AND state purpose(G, W, D−) AND deny consequences(A, R, S).

Attack 3a: deny logical consequences(S, G).

Attack 3b: state logical consequences(S, H) AND deny logical consequences(S, G).

Attack 3c: state logical consequences(S, H) AND state purpose (H, V, D+) AND deny logical consequences(S, G).

Attack 3d: state logical consequences(S, H) AND state purpose (H, V, D−) AND deny logical consequences(S, G).

Attack 3e: state logical consequences(S, H) AND state purpose (H, W, D+) AND deny logical consequences(S, G).

Attack 3f: state logical consequences(S, H) AND state purpose (H, W, D−) AND deny logical consequences(S, G).

Attack 4a: deny purpose(G, V, D+).

Attack 4b: state purpose(G, V, D−) AND deny purpose(G, V, D+).

Attack 4c: state purpose(G, W, D+) AND deny purpose(G, V, D+).

Attack 4d: state purpose(G, W, D−) AND deny purpose(G, V, D+).

Attack 5: state action(B) AND state consequences(B, R, S).

Attack 6: state action(B) AND state consequences(B, R, T) AND state logical consequences(T, G).

Attack 7a: state action(B) AND state consequences(B, R, T) AND state logical consequences(T, H) AND state purpose(H, V, D+).

Attack 7b: state consequences(A, R, S) AND state logical consequences(S, H) AND state purpose(H, V, D+).

Attack 8: state consequences(A, R, S) AND state logical consequences(S, H) AND state purpose(H, V, D−).

Attack 9: state consequences(A, R, S) AND state logical consequences(S, H) AND state purpose(H, W, D−).

Attack 10: state consequences(A, R, S) AND state logical consequences(S, H) AND state purpose(H, W, D+).

Attack 11a: state consequences(A, R, S) AND state action(B) AND state consequences(B, R, T) AND state logical consequences(T, H) AND state purpose(H, W, D+) AND deny consequences(A & B, R, X).

Attack 11b: state consequences(A, R, S) AND state purpose(H, W, D+) state logical consequences(S, ¬H).

Attack 11c: state consequences(A, R, S) AND IF state purpose(J, W, D+) THEN state logical consequences(S, ¬J).

Attack 12: deny action exists(A).

Attack 13: deny initial state exists(R).

Attack 14: deny resultant state exists(S).

Attack 15: deny goal exists(G).

Attack 16: deny value exists(R).

NOTES

[1] In this and the next schema, we label each of Walton's symbols for clarity. In an earlier account (Walton 1990), Walton gave a more detailed version of these schemes. We have built our work on the later account. The earlier scheme has five rather than four critical questions.

[2] *A&B* denotes the execution of two actions, *A* and *B*, which could be conducted sequentially or in parallel.

[3] A prototypical example can be seen at www.csc.liv.ac.uk/~katie/Parmenides.html.

REFERENCES

Atkinson, K. and T. Bench-Capon, and P. McBurney: 2004, 'PARMENIDES: Facilitating democratic debate', in *Electronic Government 2004*, Lecture Notes in Computer Science 3183, pp. 313–316.

Atkinson, K., T. Bench-Capon, and P. McBurney: 2005a, 'Arguing about cases as practical reasoning', in *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005)*, ACM Press, New York, NY, USA. pp. 35–44.

Atkinson, K., T. Bench-Capon, and P. McBurney: 2005b, 'A dialogue game protocol for multi-agent argument for proposals over action', *Journal of Autonomous Agents and Multi-Agent Systems*, 11(2), 155–171.

Bench-Capon, T.: 2003, 'Persuasion in practical argument using value based argumentation frameworks', *Journal of Logic and Computation* **13**(3), 429–448.

Bench-Capon, T. and G. Sartor: 2003, 'A model of legal reasoning with cases incorporating theories and values', *Artificial Intelligence* **150**(1–2), 97–143.

Beun, R-J. and R. M. van Eijk: 2004. 'A co-operative dialogue game for resolving ontological discrepancies', in F. Dignum (ed.), *Advances in Agent Communication*, LNAI 2922, Springer, Berlin, Germany. pp. 349–363.

Brewka, G. and T. F. Gordon: 1994, 'How to buy a porsche: an approach to defeasible decision making', in *Working Notes of the AAAI-94 Workshop on Computational Dialectics*, Seattle, Washington, USA, pp. 28–38.

Doutre, S., T. Bench-Capon, and P. E. Dunne: 2005, 'Explaining preferences with argument positions', in *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pp. 1560–1561.

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

Dunne, P. E. and T. Bench-Capon: 2004, 'Identifying audience preferences in legal and social domains', in F. Galindo and M. Takizawa and R. Traunmüller (eds.), *Database and Expert Systems Applications*, Lecture Notes in Computer Science 3180, Springer, Berlin, pp. 518–527.

Georgeff, M. P. and A. L. Lansky: 1987, 'Reactive reasoning and planning', in *Proceedings of the Sixth International Conference on Artificial Intelligence (AAAI-87)*, Seattle, WA. pp. 677–682.

Gordon, T. F. and N. I. Karacapilidis: 1997, 'The Zeno Argumentation Framework', in *Proceedings of 6th International Conference on AI and Law (ICAIL-2003)*, ACM Press, New York, pp. 10–18.

Grasso, F., A. Cawsey, and R. Jones: 'Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition', *International Journal of Human-Computer Studies* **53**, 1077–1115.

Greenwood, K., T. Bench-Capon and P. McBurney: 2003, 'Towards a computational account of persuasion in law', in *Proceedings of the Ninth International Conference on AI and Law (ICAIL-2003)*, ACM Press, New York, NY, USA, pp. 22–31.

Habermas, J.: 1996, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, MIT Press, Cambridge, MA, USA. (Translation by W. Rehg).

Hamblin, C. L.: 1970, *Fallacies*, Methuen, London, UK.

Jarke, M., M. T. Jelassi and M. F. Shakun: 1987, 'MEDIATOR: Towards a negotiation support system', *European Journal of Operational Research* **31**, 314–333.

Kenny, A. J. P.: 1975, *Practical Reasoning and Rational Appetite* Reprinted in Raz (1978).

Labrou, Y., T. Finin, and Y. Peng: 1999, 'Agent communication languages: The current landscape', *IEEE Intelligent Systems* **14**(2), 45–52.

Lührs, R., S. Albrecht and M. Lübcke and B. Hohberg: 2003, 'How to grow? Online consultation about growth in the city of Hamburg: methods techniques success factors', in R. Traunmüller (ed.), *Electronic Government (EGOV 2003)*, Lecture Notes in Computer Science 2739, Springer, Berlin, Germany, pp. 79–84.

MacKenzie, J. D.: 1979, 'Question-begging in non-cumulative systems', *Journal of Philosophical Logic* **8**, 117–133.

McBurney, P. and S. Parsons: 2002, 'Games that agents play: A formal framework for dialogues between autonomous agents', *Journal of Logic Language and Information* **11**(3), 315–334.

McCarthy, J. and P. J. Hayes: 1969, 'Some philosophical problems from the standpoint of artificial intelligence', in B. Melzer and D. Michie (eds.), *Machine Intelligence 4*, Edinburgh University Press, PP. 463–502.
Milligram, E.(ed.): 2001, *Varieties of Practical Reasoning*, MIT Press: A Bradford Book, Cambridge, MA, USA.
Modgil, S. and J. Fox: 2004, 'A guardian agent approach to safety in medical multiagent systems', in proceedings of the First International Workshop on Safety and Security in Multi-Agent Systems (SASEMAS), AAMAS 2004, New York, NY, USA.
Perelman, C.: 1980, *Justice Law and Argument*, D. Reidel Publishing Company, Dordrecht, Holland.
Perelman, C. and L. Olbrechts-Tyteca: 1969, *The New Rhetoric: A Treatise on Argumentation*, University of Notre Dame Press, Notre Dame, IN, USA.
Raz, J. (ed.): 1978, *Practical Reasoning*, Oxford University Press, Oxford, UK.
Richardson, H. S.: 1994, *Practical Reasoning about Final Ends*, Cambridge University Press, Cambridge, UK.
Searle, J. R. 2001, *Rationality in Action*, MIT Press, Cambridge, MA, USA.
Sierra, C., N. R. Jennings, P. Noriega, and S. Parsons: 1998, 'A Framework for Argument Based Negotiation', in A. Rao, M. Singh and M. Wooldridge (eds.), *Intelligent Agents IV*, LNAI, Springer, Berlin, Germany.
Tamma, V. and T. J. M. Bench-Capon: 2001, '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.
Tamma, V., I. Blacoe, B. Lithgow-Smith, and M. Wooldridge: 2004, 'SERSE: Searching for Semantic Web Content', in R. López de Mántaras and L. Saitta (eds.), *Proceedings of the Sixteenth European Conference on Artificial Intelligence (ECAI-04)*, pp. 63–67.
Tennent, R. D.: 1991, *Semantics of Programming Languages*, Prentice-Hall, Hemel Hempstead, UK.
Verheij, B.: 1990, 'Dialectical argumentation with argumentation schemes: an approach to legal logic', *Artificial Intelligence and Law* **11**, 167–195.
Walton, D. N.: 1990, *Practical Reasoning: Goal-Driven Knowledge-Based Action-Guiding Argumentation*, Rowman and Littlefield Savage, Maryland, USA.
Walton, D. N.: 1996, *Argument Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, Mahwah, NJ, USA.
Walton, D. N. and E. C. W. Krabbe: 1995, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, SUNY Press, Albany, NY, USA.
Wooldridge, M. J.: 2000, *Reasoning about Rational Agents*, MIT Press, Cambridge, MA, USA.

Department of Computer Science
University of Liverpool
Liverpool L69 3BX
UK
E-mail: {k.m.atkinson,tbc,p.j.mcburney}@csc.liv.ac.uk