Computational Representation of Practical Argument

Katie Atkinson, Trevor Bench-Capon and Peter McBurney
Department of Computer Science
University of Liverpool
Liverpool L69 3BX
UK
{k.m.atkinson,tbc,p.j.mcburney}@csc.liv.ac.uk

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, extending the account of Walton [33]. From this perspective, we articulate an interaction protocol, which we call *PARMA*, for dialogues over proposed actions based on this theory. We outline both an axiomatic and denotational 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 not 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. This 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 extends the approach of Walton [33].

The paper is structured as follows. Section 2 discuss practical reasoning and highlights problems with the practical syllogism, a traditional method of viewing reasoning about actions as akin to reasoning about beliefs. This Section then goes on to examine and extend the work of Walton who gives an account of practical reasoning as presumptive justifications and critical questions. Section 3 articulates the schema and 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 axiomatic and denotational 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, for example [20]; the other is an online discussion forum realized using web technology. 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 can be used to filter options to decide on a course of action in the context of a multi-agent system. Finally, Section 8 offers some conclusions and possible extensions for future work.

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., [36], 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 [23, 25] and a book by John Searle [28]. Most of this work has taken as its starting point a version of the practical syllogism. Here is a typical example, taken from [18]:

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 this might appear to have the form of a deduction, on closer inspection it is evident that it may well be 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. Unlike beliefs as to what is true, when the world restricts us to a single rational choice, 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 person 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 in [28] (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 [36], 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) [10], as this is the most widely used.

The process of reasoning about action is described by Wooldridge in [36] 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. 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 developing a set of actions (or plans) in order for it to achieve 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, are avoided as the agent's plan library should contain only sensible plans.

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. 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 practically 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 it 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 in [33]. 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 the critical questions associated with the scheme.

In [33] Walton 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 to any of the direct results of the action, 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.

[1] In this and the next schema, we label each of Walton's symbols for clarity.

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.

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 values?

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 agent a knows or reasonably believes. In [13], 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 realize 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.

2.4. Walton and Krabbes' Typology of Dialogues

In [35], Walton and Krabbe identified a number of distinct dialogue types used in human communication: Persuasion, Negotiation, Inquiry, Information-Seeking, Deliberation, and Eristic Dialogues. This typology has become quite influential, and so we should consider how our work relates to it. These dialogue types are characterized by their initial positions, main goal and the aims of the participants. The dialogue types are summarised in Table 1.

As can be seen from Table 1, a number of these types could potentially involve conflict between participants over an action to be taken or decided upon. Whenever a dialogue involves such a conflict we refer to this part of the dialogue as *persuasion*; such persuasive dialogue may take place within any of the above dialogue types. Moreover, note that while Walton [34] reserves

Table I. **Types of Dialogue**

Type	Initial Situation	Main Goal	Participants' Aims
Persuasion	Conflicting points of view	Resolution of such conflicts by verbal means	Persuade the other(s)
Negotiation	Conflict of interests and need for cooperation: several actions	Making a deal	Get the best out of it for oneself
Inquiry	General ignorance	Growth of knowledge and agreement	Find a proof or destroy one
Deliberation	Need for action	Reach a decision	Find an outcome
Info-seeking	Personal ignorance	Spreading knowledge and revealing positions	Gain, pass on, show or hide personal knowledge
Eristic Dialogue	Conflict and antagonism	Reaching an accommodation in a relationship	Strike the other party and win in the eyes of onlookers

persuasion for arguments about beliefs, with respect to actions we distinguish between situations where both parties are indifferent to the action chosen, and those where at least one favours a course of action at the outset; we see the former as deliberation and the latter as persuasion. In our sense, then, persuasion can occur as part of any of the dialogue types identified by Walton and Krabbe. Our focus is on a particular aspect of such dialogues rather than the interaction as a whole.

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 realize some goal G
which will promote some value V.

We need recognize 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 realized and value promoted. We may then formalize 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 $< v, sign >$, where $v \in$ *Values* and $sign \in \{+, =, -\}$.
- A ternary relation *apply* on *Acts* $\times$ *States* $\times$ *States*, with *apply(A, R, S)* to be read as: *"Performing action A in state R results in state S."*

The argument schema G1 contains a number of problematic notions which are not readily formalized in classical logic. We can, however, see that there are four classical statements 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)* $=< v, + >$.

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 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)* $=< v, + >$.

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.

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 $\in$ 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)* $=< v, + >$.

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)* $=< v, - >$.

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)* $=< w, + >$, 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)* $=< w, - >$, 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)* $\neq< v, + >$.

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)* $=< v, - >$.

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)* $=< w, + >$.

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)* $=< w, - >$.

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

Attack 4a: It is not the case that *value(G)* $=< v, + >$.

Attack 4b: It is not the case that *value(G)* $=< v, + >$ and *value(G)* $=< v, - >$.

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

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

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 [13].

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)* $=< v, + >$.

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)* $=< v, + >$.

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)* $=< v, - >$.

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)* $=< w, - >$.

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)* $=< w, + >$.

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)* $=< w, + >$. However, there is no state X $\in$ *States* such that *apply(A&B,R,X)* $\in$ *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)* $=< w, + >$. 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)* $=< w, + >$, 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 2. The last column will be discussed in Section 6.

Table II. **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 realized by some alternative action	CQ1a	What is best
6	Believe that the desired features can be realized through some alternative action	CQ1b	What is best
7	Believe that the desired value can be realized 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 realized	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 axiomatic and a denotational 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 [19], 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 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 3 and 4, grouped into five classes. Fifteen locutions are shown in Table 3, 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. **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 4 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 IV. **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 [1].

Locutions for Attacks

The attacks listed in Section 3.2 can be realized in the *PARMA* protocol by means of one or more of the 25 primitive locutions listed in Tables 3 and 4. If more than one locution is required to be uttered for an attack to be realized, the order of utterance is not important. As illustration, we show which locutions are required for four attacks:

– Locution for Attack 1a: Deny circumstances(R)
– Locutions for Attack 1b: State circumstances(Q) AND Deny circumstances(R)
– Locution 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 realize each of the attacks of Section 3.2 can also be found in [1].

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 [15] and in accordance with most work on dialogue games (e.g.[20]) 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 — ie, statements which an agent must defend in the dialogue if attacked — and may bear no relation to the agent's real beliefs or intentions [15]. We thus make no assumptions about the private mental states of the agents involved in the dialogue. Full details of the axiomatic semantics can be found in [3, 1].

In the dialogue game built on these semantics [1], 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.

4.3. DENOTATIONAL SEMANTICS FOR *PARMA*

We also propose, in outline form, a denotational semantics for the *PARMA* protocol, that is a semantics which maps statements in the syntax to mathematical entities [32]. Our approach draws on the semantics proposed by Charles Hamblin for imperative statements [16], which itself may be viewed as a process theory of causality. The main proponent of such theories has been Wesley Salmon, whose theory of causal processes *"identifies causal connections with physical processes that transmit causal influence from one spacetime location to another"* [27, p. 191]. Our approach draws on elements of category theory, namely topos theory. Our reason for using this, rather than (say) a Kripkean possible worlds framework or a labelled transition system, is that topos theory enables a natural representation of logical consequence ($S \models G$) in the same formalism as mappings between spaces ($R \stackrel{A}{\to} S$ and $G \uparrow v$). To our knowledge, no other non-categorical denotational semantics currently proposed for action formalisms permits this.

We begin by representing proposals for action. We assume, as in Section 3.1, finite sets of Acts, Propositions, States, Goals, and Values, and various mappings. For simplicity, we assume there are n propositions. Each State may be considered as being equivalent to the set of propositions which are true in that State, and so there are 2^n States. We consider the space $\mathcal{C}$ of these States, with some additional structure to enable the representation of actions and truth-values. We consider values as mappings from Goals to some space of evaluations, called $\mathcal{S}$. This need not be the three-valued set *Sign* $= \{+, =, -\}$ we assumed in Section 3.1, although we assume that $\mathcal{S}$ admits at least one partial order. The structures we assume on $\mathcal{C}$ and $\mathcal{S}$ and between them is intended to enable us to demonstrate that these are categorical entities [11]. We begin by listing the mathematical entities, along with informal definitions.

- The space $\mathcal{C}$ comprises a finite collection $\mathcal{C}_0$ of objects and a finite collection $\mathcal{C}_1$ of arrows between objects.

- $\mathcal{C}_0$ includes 2^n objects, each of which may be considered as representing a State. We denote these objects by the lower-case Greek letters, $\alpha, \beta, \gamma, \ldots$, and refer to them collectively as *state objects* or *states*. We may consider each state to be equivalent (in some sense) to the set of propositions which are true in the state.

- $\mathcal{C}_1$ includes arrows between state objects, denoted by lower case Roman letters, $f, g, h, \ldots$. If f is an arrow from object α to object β, we also write $f : \alpha \rightarrow \beta$. Some arrows between the state objects may be considered as representing actions leading from one state to another, while other arrows are causal processes (not actions of the dialogue participants) which take the world from one state to another. There may be any number of arrows between the same two objects: zero, one, or more than one.

- Associated with every object $\alpha \in \mathcal{C}_0$, there is an arrow $1_\alpha \in \mathcal{C}_1$ from α to α, called the identity at α. In the case where α is a state object, this arrow may be considered as that action (or possibly inaction) which preserves the status quo at a state α.

- If $f : \alpha \rightarrow \beta$ and $g : \beta \rightarrow \gamma$ are both arrows in $\mathcal{C}_1$, then we assume there is an arrow $h : \alpha \rightarrow \gamma$. We denote this arrow h by $g \circ f$ (*"g composed with f")*. In other words, actions and causal processes may be concatenated.

- We assume that $\mathcal{C}_0$ includes a special object *Prop*, which represents the finite set of all propositions. We further assume that for every object $\alpha \in \mathcal{C}_0$ there is a monic arrow $f_\alpha : \alpha \rightarrow Prop$. Essentially, a monic arrow is an injective (one-to-one) mapping.

- We assume that $\mathcal{C}_0$ has a terminal object, **1**, ie, an object such that for every object $\alpha \in \mathcal{C}_0$, there is precisely one arrow $\alpha \to \mathbf{1}$.

- We assume that $\mathcal{C}$ has a special object Ω, and an arrow $true : \mathbf{1} \to \Omega$, called a *sub-object classifier*. The object Ω may be understood as the set comprising $\{True, False\}$.

- We assume that $\mathcal{S}$ is space of objects over which there is a partial order $<_i$ corresponding to each participant i in the dialogue. Such a space may be viewed as a category, with an arrow between two objects α and β whenever $\alpha <_i \beta$. For each participant, we further assume the existence of one or more mappings v between $\mathcal{C}$ and $\mathcal{S}$, which takes objects to objects, and arrows to arrows. We denote the collection of all these mappings by $\mathcal{V}$.

The assumptions we have made here enable us to show that $\mathcal{C}$ is a category [11], and we can thus represent the statement $R \xrightarrow{A} S$, for states R and S, and action A. Moreover, the presence of a sub-object classifier structure enables us to represent statements of the form $S \models G$, for state S and goal G, inside the same category $\mathcal{C}$. This structure we have defined for $\mathcal{C}$ creates some of the properties needed for $\mathcal{C}$ to be a topos [11]. Finally, each space $\mathcal{S}$ with partial order $<_i$ is also a category, and the mappings v are functors (structure-preserving mappings) between $\mathcal{C}$ and $\mathcal{S}$. This then permits us to represent statements of the form $G \uparrow v$, for goal G and value v.

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

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

is understood semantically as the assertion of the existence of objects representing R and S in $\mathcal{C}$, the existence of an arrow representing A between them, the existence of an arrow with certain properties[2] between *Prop* and Ω, and the existence of a functor $v \in \mathcal{V}$ from $\mathcal{C}$ to $\mathcal{S}$. Attacks on this position then may be understood semantically as denials of the existence of one or more of these elements, and possibly also, if the attack is sufficiently strong, the assertion of the existence of other objects, arrows or functors in the appropriate spaces.

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

$$\langle \mathcal{C}_1,\ \mathcal{S}_1,\ \mathcal{V}_1 \rangle,\ \langle \mathcal{C}_2,\ \mathcal{S}_2,\ \mathcal{V}_2 \rangle,\ \langle \mathcal{C}_3,\ \mathcal{S}_3,\ \mathcal{V}_3 \rangle,\ \ldots,$$

[2] This arrow is the characteristic function for the object representing G, and the properties are that a certain diagram commutes in $\mathcal{C}$.

where the k-th triple is created from the k-th utterance in the dialogue according to the representation rules just described. Then, our denotational semantics for the *PARMA* Protocol itself is defined as the collection of all such countable sequences of triples for valid dialogues conducted under *PARMA*. This approach views the semantics of the protocol as a space of mathematical objects, which are created incrementally and jointly by the participants in the course of their dialogue together. The approach derives from the constructive view of human language semantics of Discourse Representation Theory [17], and is similar in spirit to the denotational semantics, called a *trace semantics*, defined for deliberation dialogues in [21], and the *dialectical graph* recording the statements of the participants in the Pleadings Game of Thomas Gordon [12].

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. 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. The program checks the legality of the participants' chosen moves by verifying that all pre-conditions, expressed in terms of the commitment stores and dialogue history, for the move hold. Thus, the participants are able to state and attack each other's positions with the program verifying that the dialogue always complies with the Protocol. If a participant attempts to make an illegal move, the system informs them of this, and provides them with the opportunity to chose an alternative move. After a legal move has been uttered, the commitment store of the participant who made the move is updated to contain any new commitments created by the utterance. All moves, whether legal or illegal, are entered into the history, which records which moves were made by which participant and the legality of the move chosen. After a legal move has been made, the commitment store of the player who made the move is printed to the screen to show all previous commitments and any new ones that have consequently been added. By publicly displaying the commitment stores in this way each participant is able to see their own and each other's commitments. Thus, participants can determine which of their commitments overlap with those of the other participant, and thereby identify points of agreement. 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 disagrees with the attack. If the first participant refuses to accept the reasons for the attack then a point of disagreement has been identified and the dialogue terminates. Dialogues may also reach a natural end with agreement between the two participants on a course of action. If this occurs, both players may choose to exit the dialogue. 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 analyzed, 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. Further details of the implementation can be found in [1].

Implementing the dialogue game has proved to be a very useful task, as we have shown that our general theory of persuasion can be conducted via computer mediated dialogues of this form. This implementation has also raised a number of interesting issues in relation to our underlying argumentation 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 involving 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.

A more detailed description of the issues encountered through the implementation of the Java program can be found in [1].

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 well 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.

We 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 guided through a series of web pages 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 analyze the information in order to identify what elements of an argument are more strongly supported than others.

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 [4]. [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.

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 2. 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 2. 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 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. [29]).

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

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. Thus a participant may propose an alternative action which will not bring about such undesired side effects (Attacks 8, 9 and 10) and they may prefer the goals and values endorsed by this alternative action. Finally, a participant may deem an action as undesirable if it interferes with other actions in question, with respect to the promotion of another value, previously not considered (Attack 11). In such cases, disputes must be decided by determining the party whose wishes are to be represented, 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 is through access to the same ontologies, such as those used in [31, 30, 6], to establish the common language of the topic in question.

Our model assumes that such matters of meaning and context are agreed upon by the participants of a dialogue and therefore such attacks concerning representation should not occur frequently in dialogue exchanges. However, these attacks remain possible, especially in systems which permit encounters with unfamiliar or unpredictable agents, and should not be overlooked.

6.4. Clarification of a Position

In everyday conversations disputes often arise due to participants making ill-informed assumptions about each other's positions. As conversations progress the players' positions become clearer and more explicit and earlier ill-informed assumptions may be dissolved. However, players may not be aware of their opponents' full position about an issue. If the position is not fully explicit then the players may 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 participant's ontologies to ensure a shared understanding of the concepts in the given topic of conversation. Both disagreements about representation and disagreements about facts should be resolved before disagreements about choice can be addressed.

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 then ministers make the choices between actions on the basis of these facts.

Naturally, resolution will also occur if one party allows themself to be persuaded that their preference ordering is wrong or they concede to the ordering of their opponent's preferences. 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 (Attacks 9 or 16), then resolution will require agreement between them on a preference ordering over values. Such resolution may require other types of dialogue, and some of these interactions have received considerable attention from philosophers, for example [14, 24, 26]. A formalism to represent disagreement involving arguments which rely on values is proposed in [5].

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 first is that we must allow for the possibility of rational disagreement here. 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." [28, p. xv]

Although addressed in many current agent systems by the use of 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."* [28, 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 in [7] 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 in [5]. 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 [8]. In both [7] and [5], the use of preferred semantics gives rise to the possibility of different but defensible choices, thus accommodating the possibility of rational disagreement.

To summarize, 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.

7. Values and the BDI Architecture

In this Section we will sketch how our proposal can be made computational within the framework of an agent based on the Belief-Desire-Intention model. 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 [5] and determine the preferred extension for our agent, using the ordering of values chosen by that agent as required. This preferred extension will form the set of intentions of the agent.

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 preconditions 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 preconditions 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 $BelA(R_{ja},B_j)$ & $BelA(G,S_{ja})$ & $V_j(D_{jg}) > 0$ 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 realize 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, and the complete set can be found in [2]. Attack 1b may be made by agent k if not 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 preconditions 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 [9]. The action to be chosen by the decision-maker concerns the appropriate treatment for a patient threatened by blood clotting. The choice of actions is between:

- *Administer-aspirin*: preconditions are that the patient has high platelet adhesion. The post-conditions are that platelet adhesion is low and that expense is medium.
- *Administer-chlopidogrel*: preconditions 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 preconditions, 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 treat dangerous acidity levels, which will result if there is a history of gastritis and no protein 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:

cost(expense(large)) = -1
cost(expense(medium))= 0
cost(expense(small)) = 1
safety(reduced blood clotting) = 1
safety(dangerous acidity levels) = -1

There are three agents with beliefs as follows:

Table IV.

Agent	platelet-adhesion (high)	platelet-adhesion (medium)	platelet-adhesion (low)	history of gastritis	protein 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 proposes the following argument:

Argument A1: Assuming that platelet-adhesion is low, we should do nothing, since this has small expense which promotes the value of cost.

Agent Jay will attack this with attacks 1b, 9 and 11a:

Argument A2: Your assumption that platelet adhesion is low is false. Platelet adhesion is high.

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

Argument A4: As platelet adhesion is high, we should administer aspirin since this will result in low platelet adhesion, so that blood clotting is reduced, which fact will promote the value of safety.

This situation can be depicted as a Value-Based Argumentation Framework [5] 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 labeled with the name of the argument they represent (A1, A2, etc), along with the value which that argument promotes (*Cost, Safety*, etc.) Note that argument A2 is a factual argument, and thus promotes the value *Truth*. This representation follows that of [5], where the value *truth* is ranked as the most important value by all audiences.

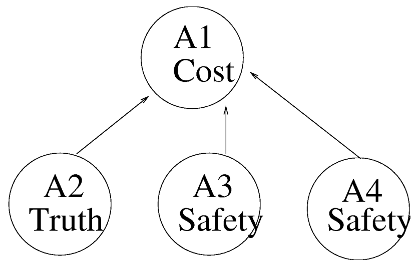

Figure 1.

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.

Argument A5: Since there is a history of gastritis and assuming no protein pump inhibitor is available, we should not administer aspirin as this will result in dangerous acidity levels, which will demote the value of safety.

The resulting argumentation framework is shown in Figure 2.

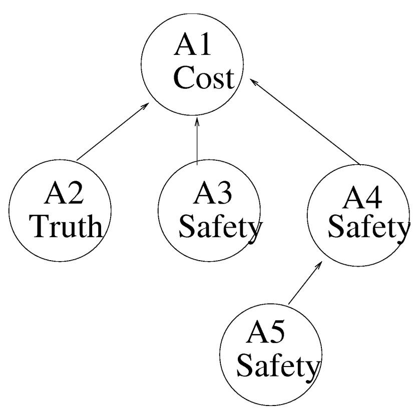

Figure 2.

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: As platelet adhesion is high, we should administer chlopidogrel since this will result in low platelet adhesion, so that blood clotting is reduced, which fact 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 [5], 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 [5], 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. We therefore add the preference as:

Argument A6a: Aspirin is preferable to chlopidogrel.

As chlopidogrel is also an expensive drug, however, attack 9 can be directed at it:

Argument A7: We should not administer chlopidogrel as this will result in large expense, which will demote the value of cost.

This produces the following argumentation framework:

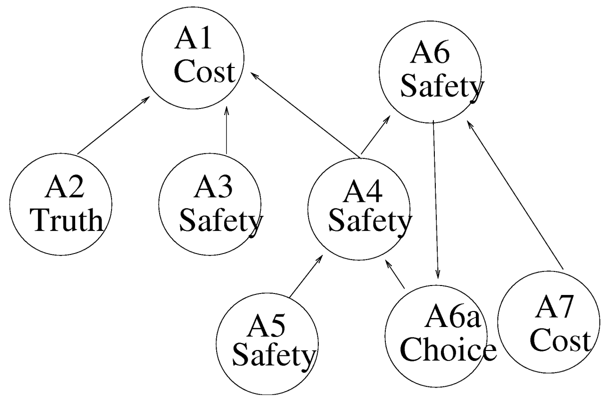

Figure 3.

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 3a:

Argument A8: Given the availability of the protein pump inhibitor, dangerous acidity levels will not result from the administration of aspirin.

This gives the final framework:

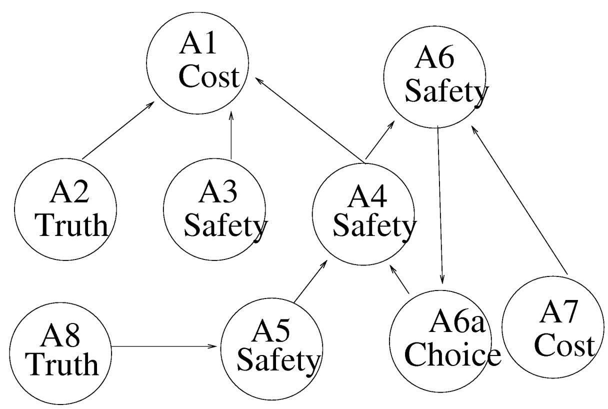

Figure 4.

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, since we have a preference for 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 protein 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. We have proposed a perspective on practical reasoning as presumptive justification and critical questions, giving an extension to the account proposed by Walton [33]. We believe that these argument schemes, such as the one we have incorporated, give us a much richer model, since Walton's original account did not allow some important distinctions to be made. In addition, the notion of a goal could be unpacked to identify further critical questions to be addressed. We have thus extended and made more specific Walton's account of practical reasoning.

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 and a denotational 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 [22]. 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 formalize our model of attacks for use in BDI agents.

9. 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.

References

1. K. M. Atkinson, T. J. M Bench-Capon, and P. McBurney. Implementation of a dialogue game for persuasion over action. Technical Report ULCS-04-005, Department of Computer Science, University of Liverpool, UK, 2003.
2. K. M. Atkinson, T. J. M Bench-Capon, and P. McBurney. Arguing about actions with BDI agents: basic definitions. Technical report, Department of Computer Science, University of Liverpool, UK, 2004. *To appear*.
3. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. A dialogue game protocol for multi-agent argument for proposals over action. In I. Rahwan, P. Moraitis, and C. Reed, editors, *Proceedings of the First International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2004)*, Berlin, Germany, 2004. Springer. *To appear*.
4. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. Parmenides: Facilitating democratic debate. In R. Traunmüller, editor, *Electronic Government 2004*, LNCS. Springer, Berlin, Germany, 2004.
5. T. J. M. Bench-Capon. Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation*, 13 3:429–48, 2003.
6. R-J Beun and R. M. van Eijk. A co-operative dialogue game for resolving ontological discrepancies. In F. Dignum, editor, *Advances in Agent Communication*, LNAI 2922, pages 349–363. Springer, Berlin, Germany, 2004.
7. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
8. P. E. Dunne and T. J. M. Bench-Capon. Identifying audience preferences in legal and social domains. In *Database and Expert Systems Applications*, LNCS. Springer, Berlin, 2004.

9. J. Fox, S. Modgil, S. Hartop, and D. Grecu. Argumentation-based decision making and reasoning over preferences. In *First International Workshop on Safety and Security in Multi-Agent Systems (SASEMAS 2004), Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, New York, NY, USA, 2004. Unpublished Invited Talk.
10. M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In *Proceedings of the Sixth International Conference on Artificial Intelligence (AAAI-87)*, pages 677–682, Seattle, WA, 1987.
11. R. Goldblatt. *Topoi: The Categorial Analysis of Logic*. North-Holland, Amsterdam, The Netherlands, 1979.
12. T. F. Gordon. The Pleadings Game: An exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.
13. K. M. Greenwood, T. J. M. Bench-Capon, and P. McBurney. Towards a computational account of persuasion in law. In *Proceedings of the Ninth International Conference on AI and Law (ICAIL-2003)*, pages 22–31, New York, NY, USA, 2003. ACM Press.
14. J. Habermas. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press, Cambridge, MA, USA, 1996. (Translation by W. Rehg).
15. C. L. Hamblin. *Fallacies*. Methuen, London, UK, 1970.
16. C. L. Hamblin. *Imperatives*. Basil Blackwell, Oxford, UK, 1987.
17. H. Kamp and U. Reyle. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic, Dordrecht, The Netherlands, 1993. Two Volumes.
18. A. J. P. Kenny. *Practical Reasoning and Rational Appetite*. 1975. Reprinted in [25].
19. Y. Labrou, T. Finin, and Y. Peng. Agent communication languages: The current landscape. *IEEE Intelligent Systems*, 14(2):45–52, 1999.
20. J. D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.
21. P. McBurney and S. Parsons. A denotational semantics for deliberation dialogues. In N. R. Jennings, C. Sierra, E. Sonenberg, and M. Tambe, editors, *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004), New York City, NY, USA*, pages 86–93, New York City, NY, USA, 2004. ACM Press.
22. J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Melzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969.
23. E. Millgram, editor. MIT Press: A Bradford Book, Cambridge, MA, USA, 2001.
24. C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame, IN, USA, 1969.
25. J. Raz, editor. *Practical Reasoning*. Oxford University Press, Oxford, UK, 1978.
26. H. S. Richardson. *Practical Reasoning about Final Ends*. Cambridge University Press, Cambridge, UK, 1994.
27. W. C. Salmon. *Causality and Explanation*. Oxford University Press, New York, NY, USA, 1998.
28. J. R. Searle. *Rationality in Action*. MIT Press, Cambridge, MA, USA, 2001.
29. C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons. A framework for argument based negotiation. In A. Rao M. Singh and M. Wooldridge, editors, *Intelligent Agents IV*, LNAI. Springer, Berlin, Germany, 1998.
30. V. Tamma, I. Blacoe, B. Lithgow-Smith, and M. Wooldridge. Serse: Searching for semantic web content. In *Proceedings of the Sixteenth European Conference on Artificial Intelligence (ECAI-04)*, Valencia, Spain, 2004.

31. V. A. M. Tamma and T. J. M. Bench-Capon. A conceptual model to facilitate knowledge sharing in multi-agent systems. In *Proceedings of Autonomous Agents 2001 Workshop on Ontologies in Agent Systems (OAS 2001)*, pages 69–76, Montreal, Canada, 2001.
32. R. D. Tennent. *Semantics of Programming Languages*. Prentice-Hall, Hemel Hempstead, UK, 1991.
33. D. N. Walton. *Argument Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.
34. D. N. Walton. *The New Dialectic: Conversational Contexts of Argument*. University of Toronto Press, Toronto, Ontario, Canada, 1998.
35. D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, Albany, NY, USA, 1995.
36. M. J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge, MA, USA, 2000.