Agent Decision Making Using Argumentation About Actions

Katie Atkinson, Trevor Bench-Capon and Peter McBurney

Department of Computer Science
University of Liverpool,
Liverpool,
UK
{k.m.atkinson,tbc,p.j.mcburney}@csc.liv.ac.uk

Abstract. In this paper we consider how a BDI agent might determine its best course of action. We draw on previous work which has presented a model of persuasion over action and extends the account of Walton. We propose a formalism based upon this model which extends the BDI agent architecture to include the notion of value functions. This enables us to make use of an argumentation framework in order to resolve conflicts over values. This formalism will allow BDI agents to reason and argue about practical action, in accordance with this model.

1 Introduction

The ability to reason effectively about what is the best or most appropriate course of action to take in a given situation is an essential activity for an agent. However, practical reasoning — reasoning about action — has not received, in either Computer Science or Philosophy, the attention that has been given to reasoning about beliefs. In this paper we provide an account of practical reasoning for agent systems by proposing a formalism for BDI agents to reason effectively and argue with themselves or other agents about practical action. Our account is based on previous work which has made use of an argumentation scheme and associated critical questions. We now take this general model of persuasion over action and tailor it for use by BDI agents.

The paper is structured as follows. Section 2 provides some motivations for this work by discussing the issues associated with dealing with practical reasoning. This section then goes on to briefly reprise our general theory of persuasion over action, which extends the work of Walton [12] who gave an initial account of practical reasoning in terms of presumptive justifications and critical questions. Section 3 presents a formalism to represent our theory of persuasion in a BDI agent architecture which we have augmented to incorporate the notion of value functions. Section 4 goes on to show 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. Section 5 gives an example which uses our theory and formalism and finally, Section 6 offers some concluding remarks and possible extensions for future work.

The primary contributions of this paper are two-fold. Firstly, we represent formally all the possible attacks on a proposed action. We do this using a recent representation of the justification for a proposed action as a presumptive argument together with a set of critical questions, from which a complete set of attacks on the argument can be generated. For each possible attack, we present sets of pre- and post-conditions, each condition being defined computationally. This work advances the computational modelling of practical reasoning. Secondly, we combine this formalism with a Value-based Argumentation Framework (VAF), to provide a mechanism to filter possible actions. The VAF allows us to represent the dialectical relationships between arguments for and against proposed actions, and, given the value preferences of an agent, to resolve these relationships into the agent's preferred set of arguments for action. By this means, a BDI agent can select from all possible justifiable actions a subset which comprise its intentions. We have thereby provided a novel qualitative mechanism for generation of agent intentions, using argumentation.

2 Motivation

One influential approach to practical reasoning in informal logic has been given by Walton [12] which regards practical reasoning as a species of presumptive argument. Given such an argument, we have a presumptive justification 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. Walton uses the notion of an argument scheme to present an argument giving 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.

The primary argument scheme for practical reasoning given in [12] is the *sufficient condition scheme*:

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

Walton associates four critical questions with the scheme:

CQW1: Are there alternative ways of realising G?
CQW2: Is it possible to do A?
CQW3: Does agent a have goals other than G which should be taken into account?
CQW4: Are there other consequences of doing A which should be taken into account?

We believe this argument scheme and its critical questions both need some elaboration. Firstly, we believe that the notion of a goal is ambiguous, because an action may be justified in terms of:

- its direct consequences (the effects of the action),
- a state of affairs following from the direct consequences, which the action was intended to realise (the goal of the action),
- the underlying social value promoted by performing the action, so as to realise the goal (the purpose of the action).

Secondly, apart from the possibility of the action, Walton does not consider other problems with soundness of W1, presupposing that the second premise is to be understood in terms of what an agent knows or reasonably believes. In [10] and [3] an argument scheme is proposed which extends Walton's scheme by unpacking his goal into *direct consequences*, *goals* and *purposes* and also makes explicit the factual context. This argument scheme is as follows:

AS1 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.

Additionally [10] and [3] make Walton's argument scheme more precise by giving relatively formal definitions of the critical questions associated with it. This allows for the separation of conflicting parts of the argument that can be resolved through verification of objective facts, from the parts which are a matter of subjective choice for each individual participant.

In [4] sixteen critical questions associated with this scheme are identified. This allows us to question the presumptions made in the argument scheme in order to consider all alternative options and thus choose the "best" action based upon the justifications. Also worthy of note is that each of the critical questions falls into one of three distinct categories which relate to the nature of the attack: issues relating to the beliefs as to what is the case; issues relating to desires as to what should be the case; and issues relating to representation concerning the language being used and the logic being deployed in the argument. In [2] there is a discussion of each of these categories into which the critical questions fall and how resolution of the attack is dependant upon the category into which it falls. We will not discuss this further here. Nor will we discus the attacks falling into the representation category as in many common applications they do not arise. Details of these critical questions can be found in [4] and [1].

With the exception of issues of representation, the critical questions associated with the extended argument scheme are:

CQ1: Are the believed circumstances true?
CQ2: Assuming this, does the action have the stated consequences?

CQ3: Assuming all of these, will the action bring about the desired goal?
CQ4: Does the goal realise the value intended?
CQ5: Are there alternative ways of realising the same consequences?
CQ6: Are there alternative ways of realising the same goal?
CQ7: Are there alternative ways of promoting the same value?
CQ8: Does doing the action have a side effect which demotes the value?
CQ9: Does doing the action have a side effect which demotes some other value?
CQ10: Would doing the action promote some other value?
CQ11: Does doing the action preclude some other action which would promote some other value?

Therefore, in an argument about a matter of practical action, we should expect to see one or more *prima facie* justifications advanced stating, explicitly or implicitly, the current situation, an action, the situation envisaged to result from the action, the features of that situation for which the action was performed and the value promoted by the action. The critical questions can then be used to generate families of attacks on these justifications, of which we will give a formal definition in the next section. Because we see this situation as one of conflict, from now on we will refer to the various critical questions that can be posed in the given situation as "attacks". An attack is possible if the pre-conditions for a critical question are satisfied. We will also consider a number of variants on the basic attacks, which results in there being more attacks than critical questions. The reason for this is because 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 CQ1, 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, giving two variant attacks from the same critical question.

In the next section we formalise this notion of attack in terms applicable to BDI agents. However, before we do so we make some remarks regarding the underlying assumptions of our formalism and its relation to planning in agents. First we assume that the agents using our method have available to them a plan library containing a set of predefined plans as in [13]. In our formalism we use the word 'action' (corresponding to A in AS1) to denote some way for the agent to achieve the specified desire. Actions in this sense are not necessarily atomic: they could be a single action or a sequence of actions which form a plan. Each plan in the plan library has associated with it a set of pre- and post-conditions. In order for an action to be executable the pre-conditions for it must be satisfied. Such pre-conditions of actions are referred to in our model as the agent's beliefs (corresponding to R in AS1) about the world that need to hold in order for it to be able to select the action. Once the action has been executed the post-conditions of it are applied. The post-conditions of actions in the plan library correspond to the desires (corresponding to G in AS1) in our model. Executing the plan will also cause changes to the state of the world. The final state (corresponding to S in AS1) is what we refer to as 'Post' (i.e. post-plan execution) in definition 3 and following. The values (corresponding to V in AS1) with which we are augmenting the BDI model are the reasons for which agents have their desires. Once all possible actions have been iden-

tified through posing the critical questions against the proposed actions, the agent can then determine which of these actions is the best one to execute in the given situation. The chosen action will then form an intention of the agent. How the best action for the agent is chosen through the use of value-based argumentation frameworks is explained in Section 4.

3 Formalism

Firstly, we present some necessary definitions:

3.1 Definitions

Definition 1: *The Beliefs of an Agent.* The beliefs of an Agent J is a four tuple $<W_J$, A_J, D_J, $V_J>$ where,

W_J represents beliefs of Agent J about the world;
A_J represents beliefs of Agent J about actions;
D_J represents beliefs about the desires of Agent J;
V_J represents beliefs about the values of Agent J;

Definition 2: *Beliefs about the World.* The beliefs of an agent are used to determine which pre-conditions of plans in the agent's plan library are satisfied.

The beliefs about the world of Agent J is a set of triples <p, $cert_{pJ}$, t> where,

p is a proposition; $cert_{pJ}$ = -1 $\leq$ $cert_{pJ}$ $\leq$ 1; t is a time.

We interpret this as J has $cert_{pJ}$ regarding p at time t. If $cert_{pJ}$ = -1, J believes p to be definitely false, if $cert_{pJ}$ = 1, J believes p to be definitely true, and if $cert_{pJ}$ = 0, J has no opinion as to the truth of p.

Let M denote the set of all agents in the system and T the set of all times.
The set P denotes the set of all p such that <p, $cert_{pJ}$, t> $\in$ W_J for some agent J $\in$ M and some time t $\in$ T.

Definition 3: *Beliefs about Actions.* The actions available to an agent consist of plans from the plan library which are composed of one or more actions.

The beliefs about action of Agent J is a set of triples $<\alpha$, $Pre_{\alpha J}$, $Post_{\alpha J}>$ where,

α is an action; $Pre_{\alpha J}$ is a set of pairs <p, $threshold_{pJ}$> and $Post_{\alpha J}$ is a set of pairs <p, $truth_{pJ}$>, -1 $\leq$ $threshold_{pJ}$ $\leq$ 1, and -1 $\leq$ $truth_{pJ}$ $\leq$ 1.
$Pre_{\alpha J}$ is a set of preconditions for α recognised by agent J. The interpretation is that J believes that α can be performed at t if all elements of Pre_{pJ} are satisfied with respect to W_J at t.

<p, threshold$_{pJ}$> is satisfied with respect to W$_J$ if <p, cert$_{pJ}$, t> and if threshold$_{pJ}$ > 0, then cert$_{pJ} \geq$ threshold$_{pJ}$, else if threshold$_{pJ} < 0$, cert$_{pJ} \leq$ threshold$_{pJ}$. J believes that if α is performed at t, then for all <p, truth$_{pJ}$> $\in$ Post$_{\alpha J}$, <p, truth$_{pJ}$, t+1> will be an element of W$_J$.

W$_{J\alpha}$ is the state of the world that J believes will result from performing α.
Additionally, J may *assume* that α can be performed at t if all elements of Pre$_{\alpha J}$ can be *assumed to be satisfied* with respect to W$_J$ at t. <p, threshold$_{pJ}$> can be assumed satisfied with respect to W$_J$ if <p, cert$_{pJ}$, t> and if threshold$_{pJ} > 0$, then cert$_{pJ} \geq 0$ and if threshold$_{pJ} < 0$, cert$_{pJ} \leq 0$.

The set A denotes the set of all actions such that <α, Pre$_{\alpha J}$, Post$_{\alpha J}$> $\in$ A$_J$ for some agent J $\in$ M.

Definition 4: *Desires of an Agent.* The desires of an agent are the post-conditions of plans from the agent's plan library.

The desires of an Agent J is a set of pairs <d, Cond$_{dJ}$> such that,

d is a desire and Cond$_{dJ}$is a set of pairs <p, threshold$_{pJ}$>. The interpretation is that J believes that d is satisfied at t if Cond$_{dJ}$ is satisfied with respect to W$_J$ at t. The notions of satisfaction and assumed satisfaction for Cond$_{dJ}$ is the same as that for Pre$_{\alpha J}$.

The set D denotes the set of all desires such that <d, Cond$_{dJ}$> $\in$ D$_J$ for some agent J $\in$ M.

Definition 5: *Values of an Agent.* The values of an agent are associated with desires and they give the reasons as to why the agent wants to achieve the desire.

The values of an Agent J is a set of triples <v, d, prom$_{vJ}$> such that,

v is a value,
d is a desire,
prom$_{vJ}$ a number -1 $\leq$ prom$_{vJ} \leq 1$, representing the degree to which the satisfaction of d promotes v.

The set V denotes the set of all values such that <v, d, prom$_{vJ}$> $\in$ V$_J$ for some agent J $\in$ M.

Definition 6: Let satA(Formula, W$_J$) be true if Formula can be assumed to be satisfied with respect to W$_J$. A Formula can be assumed to be satisfied if it is not known to be false.

Let satS(Formula, W$_J$) be true if Formula can be satisfied with respect to W$_J$. A Formula is satisfied if it is known to be true. This allows us to distinguish between

the mere assumption that pre-conditions hold and the actual knowledge that they hold. This reflects the fact that agents often need to make assumptions because knowledge is typically incomplete.

Now J has a presumptive argument for α at time t if:

there is an $<\alpha$, Pre$_{\alpha J}$, Post$_{\alpha J}> \in$ A$_J$ such that:
satA(Pre$_{\alpha J}$, J) at t;
satA(Cond$_{dJ}$, J) at t+1 and
Cond$_d J$ will be satisfied at t+1 with respect to W$_J$;
there is a <v, d, prom$_{vJ}>$, such that prom$_{vJ} > 0$.

The position is expressed as:

In circumstances r, where each r $\in$ R is the first term in each element of Pre$_\alpha J$,
Performing α,
Will result in s, where each s $\in$ S is the first term in each element of Post$_\alpha J$,
Which will realise d,
Which promotes v.

3.2 Pre-conditions for attacking a position

We now present the pre-conditions which must be satisfied for an attacking agent to question the presumptive argument presented to them in the opposing agent's initial statement of a position. Each attack derives from a critical question in the previous section and may have variants, as previously explained. All attacks are presented in the form of pre-conditions which must be met by the attacking agent A$_K$, in order for it to perform the attack in question on the presumptions in the scheme given by A$_J$. If all pre-conditions for the performance of the attack are met then A$_K$ can make the attack by presenting its argument. For each attack we give the critical question which motivates it, plus any variant, and a natural language definition of the corresponding argument. In this paper we present the definitions for the main form of attacks but, here we do not present all the possible variants on these attacks. Where such variant attacks are omitted it is noted in the definition below and these attacks can be found in [1]. However, all the attacks that are used in the example application in Section 5 do appear in the definitions below. In this section we will refer to situations which satisfy an agent's desires as 'goals'. This is because the underlying theory is independent of its realisation in the BDI model.

The definitions for the attacks are as follows:

There is an attacking agent K $\in$ M such that <W$_K$, A$_K$, D$_K$, V$_K>$ and agent K may attack the position put forward by agent J using the set of attacks subject to the

following conditions:

Source CQ: Are the believed circumstances true? (CQ1).

Attack 1a: *Pre-conditions for A_K to make an attack:*
satA(Pre$_{\alpha K}$, W$_K$) and,
not satS(Pre$_{\alpha K}$, W$_K$).

Argument: p may not be true.

Attack 1b: *Pre-conditions for A_K to make an attack:*
not satA(Pre$_{\alpha K}$, W$_K$).

Argument: p is not true.

Source CQ: Assuming the circumstances are true, does the action have the stated consequences? (CQ2).

Attack 2a: *Pre-conditions for A_K to make an attack:*
satA(Post$_{\alpha K}$, W$_K$) and,
not satS(Post$_{\alpha K}$, W$_K$).

Argument: α may not have the desired consequences.

Attack 2b: *Pre-conditions for A_K to make an attack:*
not satA(Post$_{\alpha K}$, W$_K$).

Argument: α will not have the desired consequences.

For attacks 2c – 2g see [1].

Attack 3a: *Pre-conditions for A_K to make an attack:*
for no $<$ d, Cond$_{dK}>$ does satA(Cond$_{dK}$, W$_{K\alpha}$) hold.

Argument: the state of affairs resulting from performing action α will not bring about the goal.

For attacks 3b – 3f see [1].

Source CQ: Does the goal realise the value intended? (CQ4).

Attack 4a: *Pre-conditions for A_K to make an attack:*
$<$v, d, prom$_{vK}>$ and,
prom$_{vK} \leq 0$.

Argument: the goal may not promote the value.

For attacks 4b – 4d see [1].

Source CQ: Are there alternative ways of realising the same consequences? (CQ5).

Attack 5: *Pre-conditions for A_K to make an attack:*
satA(Pre$_{\beta K}$, W$_K$) and,
satA(Post$_{\alpha K}$, W$_{K\beta}$) and $\beta \neq \alpha$.

Argument: there is an alternative action β which will realise the same consequences.

Source CQ: Are there alternative ways of realising the same goal? (CQ6).

Attack 6: *Pre-conditions for A_K to make an attack:*
satA(Pre$_{\beta K}$, W$_K$) and,
satA(Cond$_{dK}$, W$_{K\beta}$) and $\beta \neq \alpha$.

Argument: there is an alternative action β which will realise the same goal.

Source CQ: Are there alternative ways of promoting the same values? (CQ7).

Attack 7a: *Pre-conditions for A_K to make an attack:*
satA(Pre$_{\beta K}$, W$_K$) and,
for some e, e $\neq$ d, satA(Cond$_{eK}$, W$_{K\beta}$) and $\beta \neq \alpha$ and,
<v, e, prom$_{vK}$> and,
prom$_{vK} > 0$.

Argument: there is an alternative action β, leading to an alternative goal, which will promote the value.

For attack 7b see [1].

Source CQ: Does doing α have a side effect which demotes the value V? (CQ8).

Attack 8: *Pre-conditions for A_K to make an attack:*
satA(Cond$_{eK}$, W$_{K\alpha}$), e $\neq$ d and,
<v, e, prom$_{vK}$> and,
prom$_{vK} < 0$.

Argument: α has a side effect which satisfies an alternative goal, which demotes the value.

Source CQ: Does doing α have a side effect which demotes some other value? (CQ9).

Attack 9: *Pre-conditions for A_K to make an attack:*
satA(Cond$_{eK}$, W$_{K\alpha}$), e $\neq$ d and,
there is a w, w $\neq$ v such that $<$w, e, prom$_{wK}>$ and,
prom$_{wK} < 0$.

Argument: α has a side effect which satisfies an alternative goal, which demotes some other value.

Source CQ: Would doing α promote some other value? (CQ10).

Attack 10: *Pre-conditions for A_K to make an attack:*
satA(Cond$_{eK}$, W$_{K\alpha}$), e $\neq$ d and,
there is a w, w $\neq$ v such that $<$w, e, prom$_{wK}>$ and,
prom$_{wK} > 0$.

Argument: α has a side effect which satisfies an alternative goal, which promotes some other value.

Source CQ: Does doing α preclude some other action which would promote some other value? (CQ11).

Attack 11a: *Pre-conditions for A_K to make an attack:*
satA(Pre$_{\alpha K}$, W$_K$) and,
satA(Cond$_{eK}$, W$_{K\beta}$), e $\neq$ d and,
there is a w, w $\neq$ v such that $<$w, e, prom$_{wK}>$ and,
prom$_{wK} > 0$ and,
not satA(Pre$_{\alpha K}$, W$_{K\beta}$) and,
not satA(Pre$_{\beta K}$, W$_{K\alpha}$).

Argument: doing α precludes some other action which would promote some other value.

For attacks 11b – 11c see [1]. Also given in [1] are attacks 12–16 based on disagreements in the representation.

4 Argumentation Frameworks

The mechanism used by a BDI agent[1] to reason about action is described by Wooldridge in [13] as "the Deliberation Process". 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, given its beliefs and desires. In our model corresponding to the generation of options we generate a set of presumptive arguments for actions, and the critical questions/attacks which

[1] Because the BDI model has a number of proposed realisations we will take as our model the popular Procedural Reasoning System (PRS) [9].

can be used against these arguments. These critical questions themselves give rise to arguments attacking the original argument. The agent can now move on to the filtering phase. To perform the filtering in our model we form these arguments into a *Value-based Argumentation Framework* in the manner of [5] (an extension of Dung's framework [8] to accommodate arguments based on values). Whereas in [8] an argument is always defeated by an attacker, unless that attacker can itself be defeated, in [5], attack is distinguished from *defeat for an audience*[2]. This allows a particular audience to choose to reject an attacking argument, even if that argument cannot itself be defeated. This requires that audience to rank the purpose motivating the attacked argument, the value cited in its justification, as more important than that motivating the attacker. Within a value-based argument framework, therefore, which arguments are accepted depends on the ranking that the audience (characterised by a particular preference ordering on the values) to which they are addressed gives to these motivating purposes. Note, however, that if the attacker has the same value, the attack always succeeds.

A value-based argumentation framework with a given ordering on values can be mapped to a pair $(Args, Defeat)$, where Defeat is the subset of attacks which succeed for that audience. From this we can determine which arguments in $Args$ are acceptable to a particular audience by determining the preferred extension for that audience. The preferred extension for an audience is the maximal subset S of $Args$ such that no two arguments in S defeat each other given the value ordering of that audience, and all arguments in S are acceptable with respect to S, i.e., for any argument A in S, if A is defeated by an argument A' that is not in S, then there exists an argument in S that defeats A' on the given value ordering. The preferred extension thus represents the maximal consistent set of acceptable arguments with respect to the argumentation framework and a given value ordering, which is the maximal consistent position for an audience with that value ordering. In [5] it is shown that the preferred extension for a given value ordering is unique and non-empty, provided it contains no cycles in which every argument relates to the same value. The preferred extension will form the intentions of the agent.

Now that the proposing agent's position has been stated, the attacking agent may question this position through the use of the attacks from our model, stemming from the critical questions. However, not all of the critical questions will be applicable to arguments across all domains. For example, CQ7 would be of use in the legal domain, as it concerns the justification of a past action which is taken as a precedent supporting some future action, but it would not be applicable to many other domains. Therefore, the critical questions need to be analysed to discover which ones apply to the domain in question. Once this list has been determined the agent must then check which specific attacks have their pre-conditions satisfied for making an attack on the other agent's position. The attacking agent will then go on to actually state the attacks for which all pre-conditions hold. On completion of this phase we can then modify the argumentation framework to include these attacks on the initial position. We illustrate this with an example in the next section. The entire deliberation process as described above is shown schematically in Figure 1.

[2] The term audience is taken from [11].

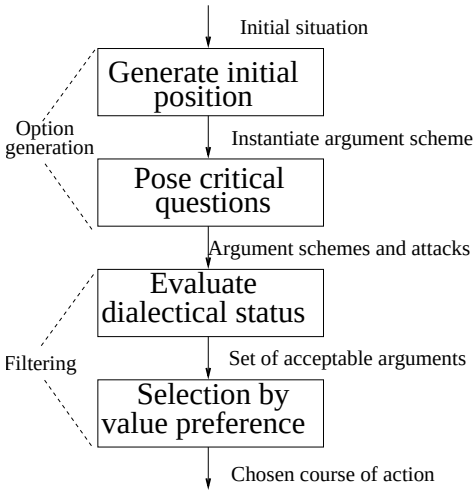

Figure 1.

5 Example Application

We now present an example situation where two agents are arguing over what action should be taken in a particular context and the dispute is resolved using the processes shown above.

The example is taken from a classic moral dilemma discussed by e.g. Coleman [7] and Christie [6]. There are two agents, called Hal and Carla, both of whom are diabetic. However, Hal, through no fault of his own, has lost his supply of insulin and needs to urgently take some to stay alive. Hal is aware that Carla has some insulin kept in her house, but Hal does not have permission to enter Carla's house. The question is whether or not Hal is justified in breaking into Carla's house in order to get some insulin to save his life. The desires are (a) Hal does not die, (b) Carla does not die, and (c) Carla's property (her house and her insulin) remain intact. Both (a) and (b) promote the value of respect for life and (c) promotes the value of respect for property. Table 1 shows the beliefs of the two agents in the initial situation (i.e. before Hal has taken the insulin) and it also shows the beliefs of each individual agent in the consequential situation (i.e. after Hal has taken the insulin). In the table T represents "true", F represents "false" and U represents "unknown", with respect to the agents' beliefs about the propositions.

Hal's view shows the state of the world according to his views after he has taken Carla's insulin and Carla's view shows the state of the world according to her views after Hal has taken her insulin. We can see from Table 1 that Hal, unlike Carla, knows that he can meet his needs whilst leaving enough insulin for Carla.

There are a number of possible desires which may be satisfied by the outcome of the action. These desires either promote or demote the two values in question, which are 'respect for life' and 'respect for property'.

The desires are summarised in Table 2. Here U represents "unimportant" as the desire is satisfied whether these attributes are true or false.

Table 1. Beliefs of the agents

Situation	H has insulin	C has insulin	H is alive	C is alive	Property of C intact
Before Hal takes the insulin					
Hal & Carla	F	T	T	T	T
After Hal takes the insulin					
Hal	T	T	U	T	T
Carla	T	U	U	U	F

Table 2. Desires of the agents

Desires	H is alive	C is alive	Property of C intact	Values: '+' = promoted, '–' = demoted
D1	T	U	U	+life
D2	U	T	U	+life
D3	U	U	T	+property
D4	U	U	F	–property
D5	F	U	U	–life
D6	U	F	U	–life

D1 represents the situation where Hal is alive and thus promotes the value of respect for life. Similarly, D2 promotes respect for life as this is the situation where Carla is alive. D3 represents the situation where Carla's property remains undiminished and this promotes the value respect for property. Conversely, D4 is the situation where Hal does break into Carla's house and this demotes the value of respect for property. This leaves D5 and D6 and they are respectively the situations where Hal is not alive and Carla is not alive and these both demote the value respect for life.

We now begin the discussion about the action to be taken and we assume there are two agents involved in the discussion: Hal and Carla. On the basis of his beliefs Hal can instantiate AS1 to give the following argument, argument A1:

Argument A1: Hal has no insulin and will die without some, so he should break into Carla's house to get access to some insulin, so that Hal remains alive, promoting the value of respect for life.

Carla's beliefs satisfy the conditions of attack 9, which states that performing this action has a side effect which demotes some other value, this value being respect for property. This instantiates AS1 to argument A2:

Argument A2: The insulin belongs to Carla, so Hal should not take it, so as to keep Carla's property intact and promote respect for property rights.

This situation is depicted in the argumentation graph given below in Figure 2. In this figure and in all the figures that follow, nodes represent arguments. They are labelled with the given argument identifier, the associated value, and on the right hand side, the agent introducing the argument. Arcs are labelled with the number of the attack they represent.

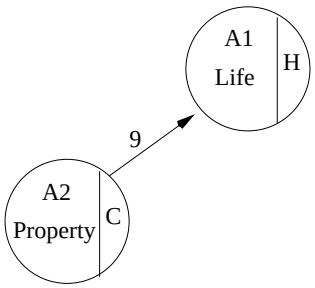

Figure 2.

Carla's beliefs also satisfy the conditions for attack 8, which states that there are unconsidered consequences of the action. This instantiates AS1 with argument A3:

Argument A3: Hal should not take the insulin as this will leave Carla without any and this will threaten her life, demoting the value of respect for life.

This adds a new node to the graph as shown below in Figure 3:

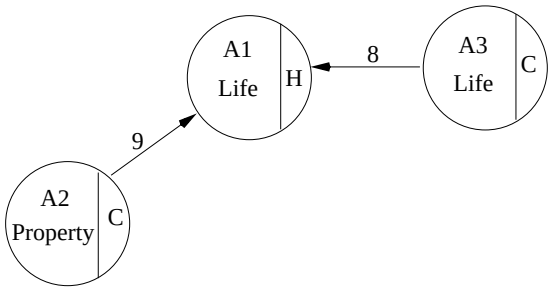

Figure 3.

Carla may herself use attack 2a since as Table 1 shows she has only assumed that Hal will not leave her with enough insulin, giving argument A4:

Argument A4: It is only an assumption that if Hal takes Carla's insulin she will be left with none and die.

From Table 1 we can see that Hal knows that Carla has plenty of insulin and so he can make the stronger attack 2b since the propositions that Carla has insulin and Carla

is alive following the action are both true rather than unknown. This gives argument A5:

Argument A5: It is only a presumption that if Hal takes Carla's insulin she will be left with none and die but we actually know that Carla has ample supplies of insulin and will not die if Hal takes some.

Since both these arguments concern matters of fact they are given the value 'truth', as in [5]. Since 'truth' is always given the highest ranking among values (we can't choose what is the case), this now means that Argument A3 is defeated, as shown in Figure 4:

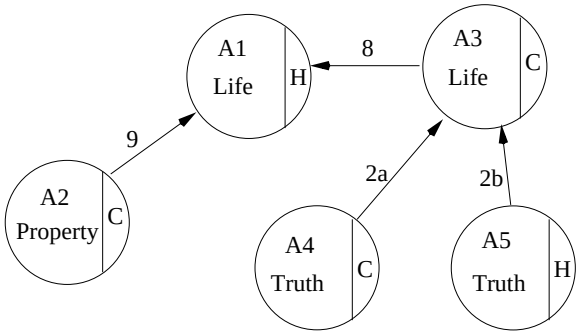

Figure 4.

The conflict[3] between A1 and A2 is resolved in accordance with the methods described in [5] by calculating the preferred extensions relative to the possible value orders. If we give respect for life a higher priority than respect for property, Hal may take the insulin, whereas if property is respected over life he may not. Thus, the acceptance of A1 is subjective and depends on how the audience to the debate ranks the two values involved.

However, [6] proposes that Hal compensate Carla which would make use of attack 6 to instantiate AS1 with argument A6 as follows:

Argument A6: When Hal has taken Carla's insulin Hal should compensate Carla, so that Carla has insulin, promoting the value of respect for property.

This new argument is added to the framework as seen below in Figure 5:

[3] Note, had we started with A2 we could have used attack 10 to attack it with A1

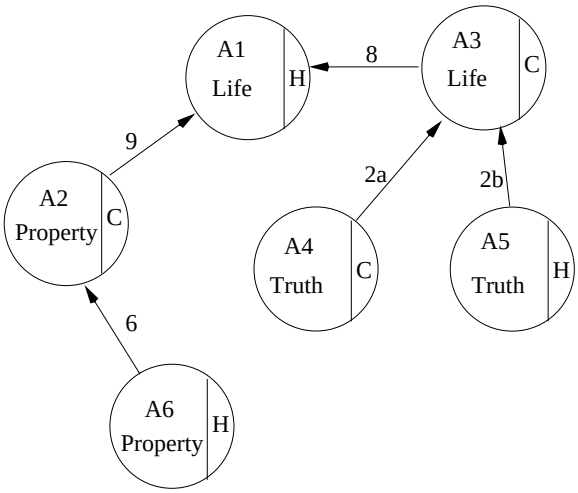

Figure 5.

Now, whatever the value ordering, A1 and A6 are accepted and so Hal can take the insulin but he must compensate Carla by replacing the insulin.

6 Concluding Remarks

In this paper we have proposed a formalism to allow BDI agents to reason and argue about the justification of proposed actions. This formalism is grounded upon a firm model of practical reasoning which uses argument schemes and associated critical questions to justify and critique a presumptive argument in favour of an action. We have used this model as a basis for our formalism which augments the BDI agent architecture to include the use of value functions. We have gone on to combine the formalism with a well understood method for resolving any conflicts over such values, through the use of Value-based Argumentation Frameworks. We have thereby provided a novel qualitative mechanism for generation of agent intentions, using argumentation.

In future work we will apply this method and formalism to specific domains and we are currently investigating how this work can be applied to the fields of medicine and law.

Acknowledgements

Katie Atkinson is grateful for support from the EPSRC. Trevor Bench-Capon and Peter McBurney acknowledge partial support received from the European Commission, through Project ASPIC (IST-FP6-002307).

References

1. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. Attacks on a presumptive argument scheme in multi-agent systems: pre-conditions in terms of beliefs and desires. Techni-

cal Report ULCS-04-015, Department of Computer Science, University of Liverpool, UK, 2004.
2. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. Computational representation of persuasive argument. Technical Report ULCS-04-006, Department of Computer Science, University of Liverpool, UK, 2004.
3. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. A dialogue game protocol for multi-agent argument for proposals over action. In I. Rahwan, P. Moraitis, and C. Reed, editors, *First International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2004)*, Lecture Notes in Artificial Intelligence, pages 149–161. Springer, Berlin, Germany, 2004. *An extended version of this paper is to appear in the Journal of Autonomous Agents and Multi-Agent Systems*.
4. K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. Justifying practical reasoning. In F. Grasso, C. Reed, and G. Carenini, editors, *Proceedings of the Fourth International Workshop on Computational Models of Natural Argument (CMNA 2004)*, pages 87–90, Valencia, Spain, 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. C. G. Christie. *The Notion of an Ideal Audience in Legal Argument*. Kluwer Academic Publishers, 2000.
7. J. Coleman. *Risks and Wrongs*. Cambridge University Press, 1992.
8. 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.
9. 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.
10. K. M. Greenwood, T. J. M. Bench-Capon, and P. McBurney. Towards a computational account of persuasion in law. In *Proceedings of Ninth International Conference on AI and Law (ICAIL-2003)*, pages 22–31, New York, NY, USA, 2003. ACM Press.
11. C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame, IN, USA, 1969.
12. D. N. Walton. *Argument Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.
13. M. J. Wooldridge. *Introduction to Multiagent Systems*. John Wiley and Sons, New York, NY, USA, 2001.