An Argumentation-based Approach for Practical Reasoning

Iyad Rahwan,[1,2] Leila Amgoud[3]

[1]Institute of Informatics
British University in Dubai
POBox 502216, Dubai, UAE
irahwan@acm.org

[2](Fellow) School of Informatics
University of Edinburgh
Edinburgh, UK

[3]IRIT
118, route de Narbonne
31062, Toulouse, France
amgoud@irit.fr

ABSTRACT

We build on recent work on argumentation frameworks for generating desires and plans. We provide a rich instantiation of Dung's abstract argumentation framework for (i) generating consistent desires; and (ii) generating consistent plans for achieving these desires. This is done through three distinct argumentation frameworks: one (now standard) for arguing about beliefs, one for arguing about what desires the agent should adopt, and one for arguing about what plans to intend in order to achieve the agent's desires. More specifically, we refine and extend existing approaches by providing means for comparing arguments based on decision-theoretic notions (cf. utility). Thus, the *worth* of desires and the *cost* of resources are integrated into the argumentation frameworks and taken into account when comparing arguments.

Keywords

Argumentation in agent systems, formal models of agency

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*intelligent agents*

General Terms

Design, Languages, Theory

1. INTRODUCTION

Various frameworks have been proposed for formalising and mechanising the reasoning of autonomous software agents based on *mental attitudes* such as beliefs, desires and intentions (BDI). These range from theoretical models of mental attitudes using modal logics [13], to operational agent architectures such as AgentSpeak [5] and 3APL [8]. A central feature of reasoning with mental attitudes is that *conflict* may arise between various attitudes.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.

Argumentation is a promising approach for reasoning with inconsistent information, based on the construction and the comparison of arguments [6]. The basic idea is that it should be possible to say more about the certainty of a particular fact than just assessing a probabilistic certainty degree in the interval $[0, 1]$. In particular, it should be possible to assess the reasons (i.e. *arguments*) why a fact holds, and to combine and compare these arguments in order to reach a conclusion. The process of *argumentation* may be viewed as a kind of reasoning about arguments (considering attacks and conflicts among them, comparing their strengths etc.) in order to determine the most *acceptable* of them. Various argument-based frameworks have been developed in defeasible reasoning [12] for generating and evaluating arguments.

Classicaly, argumentation has been mainly concerned with *theoretical reasoning*: reasoning about *propositional* attitudes such as knowledge and belief. Recently, a number of attempts have been made to use argumentation to capture *practical reasoning*: reasoning about what to do. This requires capturing arguments about non-propositional attitudes, such as desires and goals. Some argument-based frameworks for practical reasoning are instantiations of Dung's abstract framework [6] (e.g. [1, 3, 9]). Others are operational and grounded in logic programming (e.g. [10, 14]).

In this paper, we build on recent work on argumentation frameworks for generating desires and plans [1, 3, 9]. We provide a rich, argumentation-based framework for (i) generating consistent desires; and (ii) generating consistent plans for achieving these desires. This is done through three distinct argumentation frameworks: one (now standard) for arguing about beliefs, one for arguing about what desires the agent should adopt, and one for arguing about what plans to intend in order to achieve the agent's desires. More specifically, we refine and extend existing approaches by providing means for comparing arguments based on decision-theoretic notions (cf. utility). Thus, the *worth* of desires and the *cost* of resources are integrated into the argumentation frameworks and taken into account when comparing arguments.

The paper is organised as follows. After some formal preliminaries in the next section, we present our three integrated argumentation frameworks in Section 3. We discuss related work in Section 4 and conclude in Section 5.

2. PRELIMINARIES

In this section we start by presenting the logical language which will be used throughout this paper, as well as the different mental states of the agents (their bases).

Let $\mathcal{L}$ be a propositional language, $\vdash$ stands for classical inference and $\equiv$ for logical equivalence. From $\mathcal{L}$ we can distinguish the three following sets of formulas:

- The set $\mathcal{D}$ which gathers all possible desires of agents.
- The set $\mathcal{K}$ which represents the knowledge.
- The set RES which contains all the available resources in a system.

From the above sets, two kinds of rules can be defined: *desire-generation* rules and *planning* rules.

DEFINITION 1. **(Desire-Generation Rules)** *A* desire-generation rule *(or a* desire rule*) is an expression of the form*

$$\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi_1 \wedge \cdots \wedge \psi_m \Rightarrow \psi$$

where $\forall\ \varphi_i \in \mathcal{K}$ *and* $\forall\ \psi_i,\ \psi \in \mathcal{D}$.

The meaning of the rule is "if the agent *believes* $\varphi_1, \ldots, \varphi_n$ and *desires* $\psi_1, \ldots, \psi_m$, then the agent will *desire* ψ as well". And let $\texttt{head}(\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi_1 \wedge \cdots \wedge \psi_m \Rightarrow \psi) = \psi$.

Let's now define the notion of *planning rule*, which is the basic building block for specifying plans.

DEFINITION 2. **(Planning Rules)** *A* planning *rule is an expression of the form*

$$\varphi_1 \wedge \cdots \wedge \varphi_n \wedge r_1 \cdots \wedge r_m \rightarrowtail \varphi$$

where $\forall\ \varphi_i \in \mathcal{D}$, $\varphi \in \mathcal{D}$ *and* $\forall r_i \in RES$.

A planning rule expresses that if $\varphi_1, \ldots, \varphi_n$ are achieved and the resources $r_1, \ldots, r_m$ are used then φ is achieved.[1]

Let DGR and PR be the set of all possible desire generation rules and planning rules, respectively. Each agent is equipped with four bases: a base $\mathcal{B}_b$ containing its *basic beliefs*, a base $\mathcal{B}_d$ containing its desire-generation rules, a base $\mathcal{B}_p$ containing its planning rules and finally a base $\mathcal{R}$ which will gather all the resources possessed by that agent. Beliefs can be uncertain, desires may not have equal priority and resources may have different costs.

DEFINITION 3. **(Agent's bases)** *An agent is equipped with four bases* $\langle \mathcal{B}_b,\ \mathcal{B}_d,\ \mathcal{B}_p,\ \mathcal{R}\rangle$*:*

- $\mathcal{B}_b = \{(\beta_i, b_i) : \beta_i \in \mathcal{K}, b_i \in [0, 1], i = 1, \ldots, n\}$. *Pair* (β_i, b_i) *means belief* β_i *is certain at least to degree* b_i.[2]
- $\mathcal{B}_d = \{(dgr_i, w_i) : dgr_i \in DGR, w_i \in \mathbb{R}, i = 1, \ldots, m\}$. *Symbol* w_i *denotes the* worth *of the desire* $\texttt{head}(dgr)$. *Let* $Worth(\psi) = w_i$.
- $\mathcal{B}_p = \{pr_i : pr_i \in PR, i = 1, \ldots, l\}$.
- $\mathcal{R} = \{(r_i, c_i), i = 1, \ldots, n\}$ *where* $r_i \in RES$ *and* $c_i \in \mathbb{R}$ *is the cost of consuming* r_i. *Let* $Cost(r_i) = c_i$ *be a function which returns the cost of a given resource.*

In what follows, $\mathcal{B}_b^*$, $\mathcal{B}_d^*$, $\mathcal{B}_p^*$, $\mathcal{R}^*$ will denote the sets of formulas when the weights are ignored. Using desire-generation rules, we can characterise *potential desires*.[3]

[1]Note that the implications defined in desire-generation rules and planning rules are not material. So for example, from $\neg y$ and $x \rightarrowtail y$, we cannot deduce $\neg x$.

[2]The certainty degree can be seen as a necessity measure of possibility theory.

[3]Amgoud and Kaci [3] call them "potential initial goals."

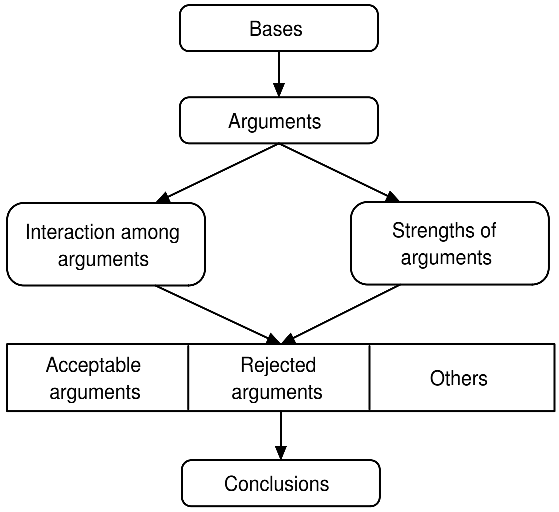

Figure 1: General view of argument-based decision making

DEFINITION 4. **(Potential Desire)** *The set of* potential desires *of an agent is* $\mathcal{PD} = \{\psi : \exists \varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi_1 \wedge \cdots \wedge \psi_m \Rightarrow \psi \in \mathcal{B}_d^*\}$.

These are "potential" desires because the agent does not know yet whether the antecedents (i.e. bodies) of the corresponding rules are true.

3. ARGUMENTATION FRAMEWORKS

The conceptual sketch of an argumentation framework is illustrated in Figure 1. It is essential to distinguish between arguing over beliefs and arguing over goals or desires. A proposition is believed because it is *true* and *relevant*. Desires, on the other hand, are adopted because they are *justified* and *achievable*. A desire is justified because the world is in a particular state that warrants its adoption. For example, one might desire to go for a walk because she believes it is a sunny day and may drop that desire if it started raining. A desire is *achievable*, on the other hand, if the agent has a plan that achieves that desire.

As a consequence of the different nature of beliefs and desires, they are supported by two different types of arguments. These arguments need to be treated differently, taking into account the different way they relate to one another. For example, a belief argument can be attacked by arguing that it is not consistent with observation, or because there is a reason to believe the contrary. Arguments for desires, on the other hand, could be attacked by demonstrating that the justification of that desire does not hold, or that the plan intended for achieving it is itself not achievable.

To deal with the different nature of the arguments involved, we present three distinct argumentation frameworks: one for reasoning about beliefs, another for arguing about what desires are justified and should be pursued, and a third for arguing about the best plan to intend in order to achieve these desires. The first framework is based on existing literature on argumentation over beliefs, originally proposed by Dung [6] and later extended by Amgoud and Cayrol [2]. For arguing about desires and plans, we draw on and extend work on argumentation-based desire-generation and planning [1, 3, 9].

3.1 Arguing over beliefs

Using beliefs, an agent can construct *belief arguments*, which have a deductive form. Indeed, from a set of beliefs, another belief is deduced as follows:

DEFINITION 5. **(Belief Argument)** *A* belief argument *A is a pair $A = \langle H, h\rangle$ such that:*

1. *$H \subseteq \mathcal{B}_b^*$;*
2. *H is consistent;*
3. *$H \vdash h$;*
4. *H is minimal (for set $\subseteq$) among the sets satisfying conditions 1, 2, 3.*

The support *of the argument is denoted by* $\mathtt{SUPP}(A) = H$. *The* conclusion *of the argument is denoted by* $\mathtt{CONC}(A) = h$. *$\mathcal{A}_b$ stands for the set of all possible belief arguments that can be generated from a belief base $\mathcal{B}_b$.*

In [2, 11], it has been argued that arguments may have forces of various strengths, and consequently different definitions of the force of an argument have been proposed. Generally, the force of an argument can rely on the information from which it is constructed. Belief arguments involve only one kind of information: *the beliefs*. Thus, the arguments using more certain beliefs are found stronger than arguments using less certain beliefs. A certainty level is then associated with each argument. That level corresponds to the less entrenched belief used in the argument. This definition is also used in belief revision [7].

DEFINITION 6. **(Certainty level)** *Let $A = \langle H, h\rangle \in \mathcal{A}_b$. The* certainty level *of A is $Level(A) = min\{a_i : \varphi_i \in H$ and $(\varphi_i, a_i) \in \mathcal{B}_b\}$.*

The different forces of arguments make it possible to compare pairs of arguments. Indeed, the higher the certainty level of an argument is, the stronger that argument is. Formally:

DEFINITION 7. **(Comparing arguments)** *Let A_1, $A_2 \in \mathcal{A}_b$. The argument A_1 is* preferred *to A_2, denoted $A_1 \succeq_b A_2$, if and only if $Level(A_1) \geq Level(A_2)$.*

Preference relations between belief arguments are used not only to compare arguments in order to determine the "best" ones, but also in order to refine the notion of *acceptability* of arguments. Since a belief base may be inconsistent, then arguments may be conflicting.

DEFINITION 8. **(Conflicts between Belief Arguments)** *Let $A_1 = \langle H_1, h_1\rangle, A_2 = \langle H_2, h_2\rangle \in \mathcal{A}_b$.*

- *A_1* undercuts *A_2 if $\exists h'_2 \in H_2$ such that $h_1 \equiv \neg h'_2$.*
- *A_1* attacks$_b$ *A_2 iff A_1 undercuts A_2 and not ($A_2 \succeq_b A_1$).*

Having defined the basic concepts, we are now ready to define the argumentation system for handling belief arguments.

DEFINITION 9. **(Belief Argumentation framework)** *An* argumentation framework *AF_b for handling belief arguments is a pair $AF_b = \langle \mathcal{A}_b,\ Attack_b\rangle$ where $\mathcal{A}_b$ is the set of belief arguments and $attack_b$ is the defeasibility relation between arguments in $\mathcal{A}_b$.*

Since arguments are conflicting, it is important to know what are the "good" ones, generally called *acceptable*. Beliefs supported by such arguments will be inferred from the base $\mathcal{B}_b$. Before defining the notion of acceptable arguments, let's first introduce a crucial notion of defence.

DEFINITION 10. **(Defence)** *Let $S \subseteq \mathcal{A}_b$ and $A_1 \in \mathcal{A}_b$. S* defends *A_1 iff for every belief argument A_2 where A_2 attacks$_b$ A_1, there is some argument $A_3 \in S$ such that A_3 attacks$_b$ A_2.*

An argument is acceptable either if it is not attacked, or if it is defended by acceptable arguments.

DEFINITION 11. **(Acceptable Belief Argument)** *A belief argument $A \in \mathcal{A}_b$ is* acceptable *with respect to a set of arguments $S \subseteq \mathcal{A}_b$ if either:*

- *$\nexists A' \in S$ such that A' attacks$_b$ A; or*
- *$\forall A' \in S$ such that A' attacks$_b$ A, we have an acceptable argument $A'' \in S$ such that A'' attacks$_b$ A'.*

This recursive definition enables us to characterise the set of acceptable arguments using a fixed-point definition.

PROPOSITION 1. *Let $AF_b = \langle \mathcal{A}_b,\ Attack_b\rangle$ be an argumentation framework. And let $\mathcal{F}$ be a function such that $\mathcal{F}(S) = \{A \in \mathcal{A}_b : S$ defends $A\}$. The set $Acc(\mathcal{A}_b)$ of* acceptable belief arguments *is defined as: $Acc(\mathcal{A}_b) = \bigcup \mathcal{F}_{i\geq 0}(\emptyset)$*

PROOF. Due to the use of propositional language and finite bases, the argumentation system is finitary, i.e each argument is attacked by a finite number of arguments. Since the argumentation system is finitary then the function $\mathcal{F}$ is continuous. Consequently, the least fixpoint of $\mathcal{F}$ is $\bigcup \mathcal{F}_{i\geq 0}(\emptyset)$. □

The set $Acc(\mathcal{A}_b)$ contains non-attacked arguments as well as arguments defended directly or indirectly by non-attacked ones.

3.2 Arguing over desires

Amgoud and Kaci have introduced *explanatory arguments* as a means for generating desires from beliefs [3]. We extend this framework in this section and refine it in order to resolve some problematic features caused by the fact that they combine belief argumentation with desire argumentation in a single framework. Moreover, we consider more general desire generation rules in the sense that a desire may not only be generated from beliefs as in [3], but it can also be generated from other desires.

In what follows, the functions $\mathtt{BELIEFS}(A)$, $\mathtt{DESIRES}(A)$ and $\mathtt{CONC}(A)$ return respectively, for a given argument A, the beliefs used in A, the desires supported by A and the conclusion of the argument A.

DEFINITION 12. **(Explanatory Argument)** *Let $\langle \mathcal{B}_b, \mathcal{B}_d\rangle$ two bases.*

- *If* $\exists(\Rightarrow \phi) \in \mathcal{B}_d^*$ *then* $\Rightarrow \phi$ *is an* explanatory argument *(A) with:*
  $\texttt{BELIEFS}(A) = \emptyset$
  $\texttt{DESIRES}(A) = \{\phi\}$
  $\texttt{CONC}(A) = \phi$

- *If* $B_1, \ldots, B_n$ *are belief arguments, and* $E_1, \ldots, E_m$ *are explanatory arguments, and* $\exists$ $\texttt{CONC}(B_1) \wedge \ldots \wedge \texttt{CONC}(B_n) \wedge \texttt{CONC}(E_1) \wedge \ldots \wedge \texttt{CONC}(E_m) \Rightarrow \psi \in \mathcal{B}_d^*$ *then* $B_1, \ldots B_n, E_1, \ldots E_m \Rightarrow \psi$ *is an* explanatory argument *(A) with:*[4]
  $\texttt{BELIEFS}(A) = \texttt{SUPP}(B_1) \cup \ldots \cup \texttt{SUPP}(B_n) \cup \texttt{BELIEFS}(E_1) \cup \ldots \cup \texttt{BELIEFS}(E_m)$
  $\texttt{DESIRES}(A) = \texttt{DESIRES}(E_1) \cup \ldots \cup \texttt{DESIRES}(E_m) \cup \{\psi\}$
  $\texttt{CONC}(A) = \psi$

$TOP(A) = \texttt{CONC}(B_1) \wedge \ldots \texttt{CONC}(B_n) \wedge \texttt{CONC}(E_1) \wedge \ldots \wedge \texttt{CONC}(E_m) \Rightarrow \psi$ *is the* TOP rule *of the argument.*
Let $\mathcal{A}_d$ *denote the set of all explanatory arguments that can be generated from* $\langle \mathcal{B}_b, \mathcal{B}_d \rangle$*, and* $\mathcal{A} = \mathcal{A}_d \cup \mathcal{A}_b$*.*

Example 1. *Let* $waic \in \mathcal{K}, aic \in \mathcal{D}$*; waic denotes "there is a relevant workshop at the Sydney AI conference;" aic denotes "attend the Sydney AI conference." Suppose we have:*
$\mathcal{B}_b = \{(waic, 0.8)\}$
$\mathcal{B}_d = \{(waic \Rightarrow aic, 6)\}$
$\mathcal{B}_p = \emptyset$
$\mathcal{R} = \emptyset$
The agent can construct the explanatory argument A_1 *in favour of its desire to attend the Sydney AI conference:*
B_1: $\langle \{waic\}, waic \rangle$
A_1: $B_1 \Rightarrow aic$
with $\texttt{BELIEFS}(A_1) = \{waic\}$, $\texttt{DESIRES}(A_1) = \{aic\}$, $\texttt{CONC}(A_1) = \{aic\}$.

Note that the above example involves a desire-generation rule that contains beliefs only in its body. The following extended example shows how a desire can follow from another, already generated desire.

Example 2. *Extending example 1, let: keynote denote "interesting key note speech"; attendkey denote "attend the key note speech". Suppose we have the following additional desire-generation rule, which states that if there is an interesting keynote speech at a conference I already desire to attend, then I would also desire to attend that speech:* $(keynote \wedge aic \Rightarrow attendkey, 8)$*. Suppose also that the agent believes that there is an interesting key note speech. Thus, we have the following new bases:*
$\mathcal{B}_b = \{(waic, 0.8), (keynote, 0.7)\}$
$\mathcal{B}_d = \{(waic \Rightarrow aic, 6), (keynote \wedge aic \Rightarrow attendkey, 8)\}$
$\mathcal{B}_p = \emptyset$
$\mathcal{R} = \emptyset$.
The agent can construct the explanatory argument A_2 *for the desire to attend the keynote speech:* B_1: $\langle \{waic\}, waic \rangle$
B_2: $\langle \{keynote\}, keynote \rangle$
A_1: $B_1 \Rightarrow aic$
A_2: $B_2, A_1 \Rightarrow attendkey$
with $\texttt{BELIEFS}(A_1) = \{waic\}$, $\texttt{BELIEFS}(A_2) = \{waic, keynote\}$, $\texttt{DESIRES}(A_1) = \{aic\}$, $\texttt{DESIRES}(A_2) = \{aic, attendkey\}$, $\texttt{CONC}(A_1) = \{aic\}$ *and* $\texttt{CONC}(A_2) = \{attendkey\}$.

[4] Note that B_i and E_i are comma-separated argument labels, not a conjunction of formulae (as in desire generation rules).

As with belief arguments, explanatory arguments may have different forces. However, since explanatory arguments involve two kinds of information: *beliefs* and *desires*, their strengths depend on both the quality of beliefs (using the notion of certainty level) and the importance of the supported desire. Formally:

Definition 13. **(The force of explanatory arguments)** *Let* $A \in \mathcal{A}_d$ *be an explanatory argument. The* force *of* A *is* $Force(A) = \langle Level(A), Weight(A) \rangle$ *where:*

- $Level(A) = \min\{a_i : \varphi_i \in \texttt{BELIEFS}(A)$ *and* $(\varphi_i, a_i) \in \mathcal{B}_b\}$*. If* $\texttt{BELIEFS}(A) = \emptyset$ *then* $Level(A) = 1$*;*
- $Weight(A) = w_i$ *such that* $(TOP(A), w_i) \in \mathcal{B}_d$*.*

In order to avoid any kind of wishful thinking, belief arguments are supposed to take precedence over explanatory ones. Formally:

Definition 14. **(Comparing mixed arguments)** $\forall A_1 \in \mathcal{A}_b$ *and* $\forall A_2 \in \mathcal{A}_d$*, it holds that* A_1 *is* preferred *to* A_2*, denoted* $A_1 \succeq_d A_2$*.*

Concerning explanatory arguments, one may prefer an argument which will, for sure, justify an important desire. This suggests the use of a conjunctive combination of the certainty level of the argument and its weight. However, a simple conjunctive combination is open to discussion since it gives an equal weight to the importance of the desire and to the certainty of the set of beliefs that establishes that the desire takes place. Indeed, since beliefs verify the validity and the feasibility of desires, it is important that beliefs take precedence over the desires. This is translated by the fact that the certainty level of the argument is more important than the priority of the desire. Formally:

Definition 15. **(Comparing explanatory arguments)** *Let* $A_1, A_2 \in \mathcal{A}_d$*.* A_1 *is* preferred *to* A_2*, denoted by* $A_1 \succeq_d A_2$*, iff*

- $Level(A_1) > Level(A_2)$*, or*
- $Level(A_1) = Level(A_2)$ *and* $Weight(A_1) > Weight(A_2)$*.*

An explanatory argument for some desire can be defeated either by a belief argument (which undermines the truth of the underlying belief justification), or by another explanatory argument (which undermines one of the existing desires the new desire is based on). Figure 2 summaries this notion of attack.

Definition 16. **(Attack among Explanatory and Belief Arguments)**
Let $A_1, A_2 \in \mathcal{A}_d$ *and* $A_3 \in \mathcal{A}_b$*.*

- A_3 b-undercuts A_2 *iff* $\exists h' \in \texttt{BELIEFS}(A_2)$ *such that* $\texttt{CONC}(A_3) =\equiv \neg h'$*;*
- A_1 d-undercuts A_2 *iff* $\exists h' \in \texttt{DESIRES}(A_2)$ *such that* $\texttt{CONC}(A_1) \equiv \neg h'$*;*
- *An argument* $A' \in \mathcal{A}$ attacks$_d$ $A_2 \in \mathcal{A}_d$ *iff* A' *b-undercuts or d-undercuts* A_2 *and not* $(A_2 \succeq_d A')$*.*

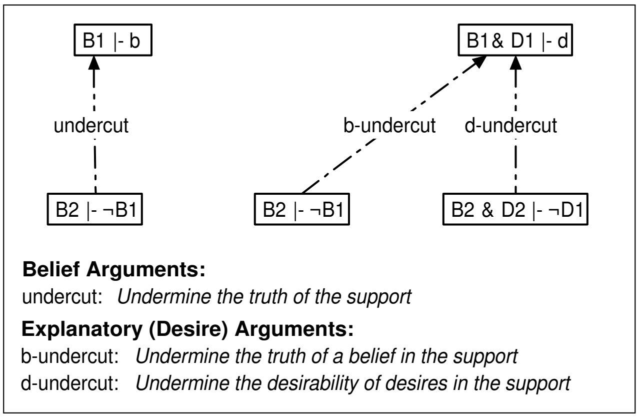

Figure 2: Summary of attacks involving belief and explanatory arguments

The following example illustrates the above concepts.

Example 3. *(Builds on example 1) The agent finds out that the workshop has been cancelled (wcancel). That agent does not desire to go to the AI conference if it is not of international standing (int). Unfortunately the Sydney AI conference is not a good one. So the new bases are:*
$\mathcal{B}_b = \{(waic, 0.8), (wcancel, 1), (wcancel \rightarrow \neg waic, 0.8), (\neg int, 1)\}$
$\mathcal{B}_d = \{(waic \Rightarrow aic, 6), (\neg int \Rightarrow \neg aic, 9)\}$
$\mathcal{B}_p = \emptyset$
$\mathcal{R} = \emptyset$.
The following arguments can be built:
B_1: $\langle\{waic\}, waic\rangle$
B_2: $\langle\{wcancel, wcancel \rightarrow \neg waic\}, \neg waic\rangle$
B_3: $\langle\{\neg int\}, \neg int\rangle$
A_1: $B_1 \Rightarrow aic$
A_2: $B_3 \Rightarrow \neg aic$
It is clear that the argument B_2 b-undercuts the argument A_1 since $waic \in$ BELIEFS(A_1) and CONC$(B_2) = \neg waic$. The argument A_2 d-undercuts the argument A_1 since CONC(A_2) $= \neg aic$ and $aic \in$ DESIRES(A_1).

Now that we have defined the notions of argument and defeasibility relation, we are ready to define the argumentation framework that should return the justified/valid desires.

Definition 17. **(Argumentation framework)** *An* argumentation framework AF_d *for handling explanatory arguments is a tuple $AF_d = \langle \mathcal{A}_b, \mathcal{A}_d, Attack_b, Attack_d\rangle$ where $\mathcal{A}_b$ is the set of belief arguments, $\mathcal{A}_d$ the set of explanatory arguments, and $attack_d$ is the defeasibility relation between arguments in $\mathcal{A}$ and $attack_b$ is the defeasibility relation between arguments in $\mathcal{A}_b$.*

The definition of acceptable explanatory arguments is based on the notion of defence. Unlike belief arguments, an explanatory argument can be defended by either a belief argument or an explanatory argument. Formally:

Definition 18. **(Defence among Explanatory and Belief Arguments)**
Let $S \subseteq \mathcal{A}$ and $A \in \mathcal{A}$. S defends *A iff $\forall A' \in \mathcal{A}$ where A' attacks$_b$ (or attacks$_d$) A, there is some argument $A'' \in S$ which attacks$_b$ (or attacks$_d$) A'.*

$\mathcal{F}'$ is a function such that $\mathcal{F}'(S) = \{A \in \mathcal{A}$ such that S defends $A\}$.

One can show easily that the function $\mathcal{F}$ is monotonic. Thus, it admits a least fixpoint. This last captures the acceptable arguments of AF_d.

Proposition 2. *Let $AF_d = \langle \mathcal{A}_b, \mathcal{A}_d, Attack_b, Attack_d\rangle$ be an argumentation framework. The set $Acc(\mathcal{A}_d)$ of* acceptable explanatory arguments *is defined as*

$$Acc(\mathcal{A}_d) = (\bigcup \mathcal{F}'_{i\geq 0}(\emptyset)) \cap \mathcal{A}_d$$

Proof. Due to the use of propositional language and finite bases, the argumentation system is finitary, i.e each argument is attacked by a finite number of arguments. Since the argumentation system is finitary then the function $\mathcal{F}'$ is continuous. Consequently, the least fixpoint of $\mathcal{F}'$ is $\bigcup \mathcal{F}'_{i\geq 0}(\emptyset)$. □

One can show that the above argumentation framework captures the results of the first framework which handles belief arguments.

Proposition 3. *Let $AF_d = \langle \mathcal{A}_b, \mathcal{A}_d, Attack_b, Attack_d\rangle$ be an argumentation framework. $\bigcup \mathcal{F}'_{i\geq 0}(\emptyset) = Acc(\mathcal{A}_b) \cup Acc(\mathcal{A}_d)$*

Proof. This follows directly from the definitions of $\mathcal{F}$ and $\mathcal{F}'$, and the fact that belief arguments are not attacked by explanatory arguments since we suppose that belief arguments are preferred to explanatory ones. □

Definition 19. **(Justified desire)** *A desire ψ is* justified *iff $\exists$ $A \in \mathcal{A}_d$ such that* CONC$(A) = \psi$*, and $A \in Acc(\mathcal{A}_d)$.*

Desires supported by acceptable explanatory arguments are justified and hence the agent will pursue them (if they are achievable).

3.3 Arguing over plans

In the previous section, we have presented a framework for arguing about desires and producing a set of *justified* desires. In what follows we will show, among these justified desires, which ones will be pursued and with which plan.
The basic building block of a plan is the notion of "partial plan," which corresponds to a planning rule.

Definition 20. **(Partial Plan)** *A* partial plan *is a pair $[H, \varphi]$ where*

- *$\varphi \in \mathcal{R}$ and $H = \emptyset$, or*
- *$\varphi \in \mathcal{D}$ and $H = \{\varphi_1, \ldots, \varphi_n, r_1 \ldots, r_m\}$ such that $\exists \varphi_1 \wedge \cdots \wedge \varphi_n \wedge r_1 \cdots \wedge r_m \rightarrowtail \varphi \in \mathcal{B}_p$.*

A partial plan $[H, \varphi]$ is elementary *iff $H = \emptyset$.*

Definition 21. **(Instrumental Argument, or Complete Plan)** *An* instrumental argument *is a pair $\langle G, d\rangle$ such that $d \in \mathcal{D}$, and G is a finite tree such that:*

- *the root of the tree is a partial plan $[H, d]$;*
- *a node $[\{\varphi_1, \ldots, \varphi_n, r_1 \ldots, r_m\}, h']$ has exactly $n + m$ children $[H'_1, \varphi_1], \ldots [H'_n, \varphi_n]$, $[\emptyset, r_1], \ldots [\emptyset, r_m]$ where each $[H'_i, \varphi_i]$, $[\emptyset, r_k]$ is a partial plan;*
- *the leaves of the tree are elementary partial plans.*

Nodes(G) *is a function which returns the set of all partial plans of tree* G*, Des*(G) *is a function which returns the set of desires that plan* G *achieves, and Resources*(G) *is a function which returns the set of all resources needed to execute* G*.*

Let $\mathcal{A}_p$ *denotes the set of all instrumental arguments that can be built from agent's bases.*

An instrumental argument may achieve one or several desires of different worths with a certain cost. So the strength of that argument is the "benefit" or "utility" which is the difference between the worths of the desires and the cost of the plan. Formally:

Definition 22. **(Strength of Instrumental Arguments)** *Let* $A = \langle G, g\rangle$ *be an instrumental argument. The* utility *of* A *is*

$$\texttt{Utility}(A) = \sum_{d_i \in Des(G)} Worth(d_i) - \sum_{r_j \in Resources(G)} Cost(r_j).$$

In [3], the strength of an instrumental argument is defined only on the basis of the weight of the corresponding desire. That definition does not account for the cost of executing the plan.

Example 4. *A customer requires a car hire (a resource) in order to go to Sydney (a goal), which in turn achieves the agent's wish to attend an Artificial Intelligence conference (a desire). The customer desires to attend the AI conference because he believes it includes a workshop related to his research (a belief that justifies the desire). Let:*
aic = "attend the Sydney AI conference";
syd = "go to Sydney";
reg = "pay conference registration";
rent = "rent a car";
ford = "get a particular car of make Ford";
pay\$100 = "pay \$100";
pay\$200 = "pay \$200";[5]
We can now specify the following, for the buyer agent B and seller agent S:

1. $\mathcal{B}_b^B = \{(waic, 1)\}$
2. $\mathcal{B}_d^B = \{(waic \Rightarrow aic, 6)\}$
3. $\mathcal{B}_p^B = \begin{cases} syd \wedge reg \rightarrowtail aic \\ rent \rightarrowtail syd \\ ford \wedge pay\$200 \rightarrowtail rent \\ pay\$100 \rightarrowtail reg \end{cases}$
4. $RES = \{pay\$100, pay\$200, ford\}$
5. $\mathcal{R}^B = \{pay\$100, pay\$200\}$
6. $\mathcal{R}^S = \{ford\}$

Figure 3 shows an instrumental argument, for attending the Sydney AI conference, that agent B can construct using the above information. Note that this plan involves the execution of action ford by agent S, because B does not have "ford" as one of its resources. Without getting the car from S, B cannot make it to Sydney using this plan.

[5]Realistically, one requires a more elaborate treatment of actions, e.g. the agent must also be able to pay \$300, or pay \$50 six times. For simplicity, we suffice with these illustrative unique actions.

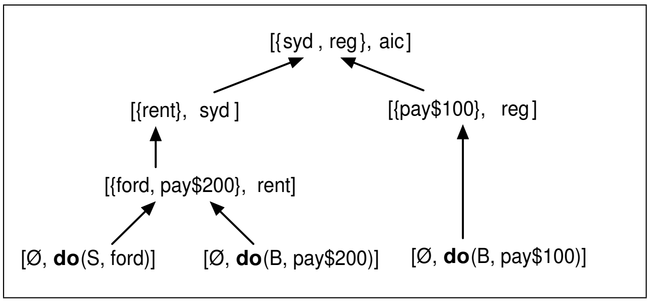

Figure 3: Complete plan for example 4

In [1], it has been shown that there are four great families of conflicts between partial plans. In fact, two partial plans $[H_1, h_1]$ and $[H_2, h_2]$ may be conflicting for one of the following reasons:

- *desire-desire* conflict, *ie* $\{h_1\} \cup \{h_2\} \vdash \bot$
- *plan-plan* conflict, *ie* $H_1 \cup H_2 \vdash \bot$.
- *consequence-consequence* conflict, *ie* the consequences of achieving the two desires h_1 and h_2 are conflicting.
- *plan-consequence* conflict, *ie* the plan H_1 conflicts with the consequences of achieving h_2.

The above conflicts are captured when defining the notion of *conflict-free* sets of instrumental arguments.

Definition 23. **(Conflict-free sets of instrumental arguments)** *Let* $S \subseteq \mathcal{A}_p$*.* S *is* conflict-free*, with respect to the agent's beliefs* $\mathcal{B}_b^*$*, iff* $\nexists\ \mathcal{B}' \subseteq \mathcal{B}_b^*$ *such that:*

1. $\mathcal{B}'$ *is consistent, and*
2. $\bigcup_{\langle G,d\rangle \in S}[\bigcup_{[H,h]\in Nodes(G)}(H \cup \{h\})] \cup \mathcal{B}' \vdash \bot$

As with belief and explanatory arguments, we now present the notion of an *acceptable set of instrumental arguments.*

Definition 24. **(Acceptable Set of Instrumental Arguments)** *Let* $S \subseteq \mathcal{A}_p$*.* S *is* acceptable *iff:*

- S *is conflict-free.*
- S *is maximal for set inclusion among the sets verifying the above condition.*

Let $S_1, \ldots, S_n$ *be the different acceptable sets of instrumental arguments.*

Definition 25. **(Achievable desire)** *Let* $S_1, \ldots, S_n$ *be the different acceptable sets of instrumental arguments. A desire* ψ *is* achievable *iff* $\exists S' \in \{S_1, \ldots, S_n\}$*, such that* $\langle G, \psi\rangle \in S'$

Definition 26. **(Utility of Set of Instrumental Arguments)** *For an acceptable set of instrumental arguments* $S = \{\langle G_1, d_1\rangle, \ldots, \langle G_m, d_m\rangle\}$*, the set of all desires achieved by* S *and all resources consumed by* S *as follows:*
$DE(S) = \{g_l : g_l \in Des(G_k), l = 1, \ldots, h, k = 1, \ldots, m\}$
$RE(S) = \{r_l : r_l \in Res(G_k), l = 1, \ldots, h, k = 1, \ldots, m\}$
The utility *of a set of arguments* S *is:*

$$\texttt{Utility}(S) = \sum_{g_i \in DE(S)} Worth(g_i) - \sum_{r_j \in Resources(S)} Cost(r_j).$$

We can now construct a complete pre-ordering on the set $\{S_1, \ldots, S_n\}$ of acceptable sets of instrumental arguments. The basic idea is to prefer the set with a maximal total utility: a maximal set of consistent plans.

DEFINITION 27. **(Preferred set)**
Let $S_1, \ldots, S_n$ be the acceptable sets of instrumental arguments. S_i is preferred *to S_j iff* $\texttt{Utility}(S_i) \geq \texttt{Utility}(S_j)$

Note that the above definition allows for cases where a set with a single desire/plan pair is preferred to another set with two or more desire/plan pairs (because the utility achieved by this desire is higher than the other two). This is more flexible than the frameworks of Amgoud and of Hustijn and van der Torre [1, 9], where sets with maximal *number* of desires are privileged, with no regard to their priority or the cost of different plans.

In order to be pursued, a desire should be both justified (i.e supported by an acceptable explanatory argument) and also achievable. Such desires will form the intentions of the agent.

DEFINITION 28. **(Intention set)**
Let $T \subseteq \mathcal{PD}$. T is an intention set iff:

1. *$\forall d_i \in T$, d_i is justified and achievable.*
2. *$\exists S_l \in \{S_1, \ldots, S_n\}$ such that $\forall d_i \in T$, $\exists \langle G_i, d_i \rangle \in S_l$.*
3. *$\forall\ S_k \neq S_l$ with S_k satisfying condition 2, then S_l is preferred to S_k.*
4. *T is maximal for set inclusion among the subsets of $\mathcal{PD}$ satisfying the above conditions.*

The second condition ensures that the desires are achievable together. If there is more than one intention set, a single one must be selected (e.g. at random) to become the agent's *intention.* The chosen set is denoted by $\mathcal{I}$. Finally, the *intended resources*, denoted $\mathcal{IR} \subseteq RES$ denote the resources needed by plans in S_l for achieving $\mathcal{I}$. The example below, depicted in Figure 4, puts the above concepts together.

EXAMPLE 5. *(Extends example 4) Suppose the buyer also would like to go on holiday to New Zealand and must reason with a limited budget. Let:*
nz = "take a holiday in New Zealand";
flynz = "fly to New Zealand";
hotel = "book a hotel accommodation";
friend = "stay at a friend's place";
call = "call a friend";
Suppose the agent has the following new desire generation knowledge base: $\mathcal{B}_d^B = \{(waic \Rightarrow aic, 0.6), \Rightarrow nz, 0.5)\}$ and that desires aic and nz are justified.

Finally, suppose costs are assigned as follows: $Cost(pay\$200) = 0.2$, $Cost(pay\$100) = 0.1$, $Cost(pay\$200) = 0.2$, $Cost(call) = 0$, $Cost(ford) = 0$).[6]

Suppose the buyer has two instrumental arguments for going to New Zealand: one requires booking a hotel (and paying \$200), while the other involves calling a friend to arrange a stay at his place. There are no conflicts between the arguments A_1, A_2 and A_3. Thus, there exists a unique acceptable set of instrumental arguments $\{A_1, A_2, A_3\}$. Since the desires aic and nz are supposed justifies, then there is a unique intention set $I = \{aic, nz\}$.

[6]The cost of "*ford*" to the buyer is zero because this resource is possessed by the seller and hence would only incur a cost to the seller.

4. RELATED WORKS

Recently, a number of attempts have been made to use formal models of argumentation as a basis for practical reasoning. Some of these models (e.g. [1, 3, 9]) are instantiations of the *abstract* argumentation framework of Dung [6], and our work is a contribution to this approach. Other approaches are based on an encoding of argumentative reasoning in logic programs (e.g. [10, 14]) or on completely new theories of practical reasoning and persuasion (e.g. [4, 15]).

Amgoud [1] presented an argumentation framework for generating consistent plans from a given set of desires and planning rules. This was later extended with argumentation frameworks that generate the desires themselves (see below).

Amgoud and Kaci [3] have a notion of "conditional rule," which is meant to generate desires from beliefs. Our desire generation rules are more general. In particular, we allow the generation of desires not only from beliefs, but also on the basis of other desires. Hence, our desire generation rules are more general.

Another problem arises because Amgoud and Kaci's definition does not distinguish between desires and beliefs in the antecedent and consequent of these rules. This may lead to incorrect inferences where an agent may conclude beliefs on the basis of yet-unachieved desires, hence exhibiting a form of wishful thinking. Our approach resolves this by distinguishing between beliefs and desires in the rule antecedents, allowing desires only in the consequent, and refining the notion of attack among explanatory arguments accordingly.

Hulstijn and van der Torre [9], on the other hand, have a notion of "desire rule," which contains only desires in the consequent. But their approach is still problematic. It requires that the selected goals[7] are supported by goal trees[8] which contain both desire rules and belief rules that are deductively consistent. This consistent deductive closure again does not distinguish between desire literals and belief literals (see Proposition 2 in [9]). This means that one cannot both believe $\neg p$ and desire p. In our framework, on the other hand, the distinction enables us to have an acceptable belief argument for believing $\neg p$ and, at the same time, an acceptable explanatory argument for desiring p.

Another advantage of our framework is that it derives preferences among explanatory and instrumental arguments using both worth and cost measures. This contrasts with Amgoud's and Hulstijn and van der Torre's frameworks, which privilege extensions with maximal number of desires without regard to desire priorities and resource cost. And while [3] does incorporate the weight of desires when calculating the strength of an instrumental argument, the cost of executing plans is not taken into account.

5. CONCLUSIONS

We presented a formal model for reasoning about desires (generating desires and plans for achieving them) based on argumentation theory. We adapted the notions of attack and preference among arguments in order to capture the differences in arguing about beliefs, desires and plans. We incorporated both the worth of desires and cost of resources in order to produce intentions that maximise utility.

One of the main advantages of our framework is that, being grounded in argumentation, it lends itself naturally

[7]Similar to our justified desires

[8]Similar to our explanatory arguments.

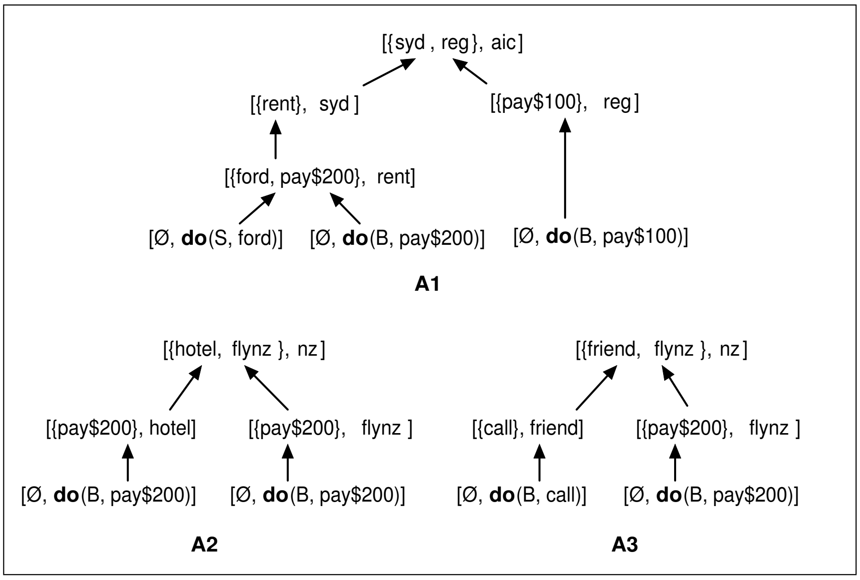

Figure 4: Plans for example 5

to facilitating dialogues about desires and plans. Indeed, we are currently extending our framework with dialogue game protocols in order to facilitate negotiation and persuasion among agents. Another interesting area of future work is investigating the relationship between our framework and axiomatic approaches to BDI agents.

6. ACKNOWLEDGMENTS

The authors thank Liz Sonenberg, Frank Dignum and Rogier van Eijk for discussions surrounding the paper topic.

7. REFERENCES

[1] L. Amgoud. A formal framework for handling conflicting desires. In T. D. Nielsen and N. L. Zhang, editors, *Proc. ECSQARU*, volume 2711 of *LNCS*, pages 552–563. Springer, Germany, 2003.

[2] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1–3):197–215, 2002.

[3] L. Amgoud and S. Kaci. On the generation of bipolar goals in argumentation-based negotiation. In I. Rahwan et al, editor, *Proc. 1st Int. Workshop on Argumentation in Multi-Agent Systems (ArgMAS)*, volume 3366 of *LNCS*. Springer, Germany, 2005.

[4] K. Atkinson, T. Bench-Capon, and P. McBurney. Justifying practical reasoning. In C. R. F. Grasso and G. Carenini, editors, *Proc. Workshop on Computational Models of Natural Argument (CMNA)*, pages 87–90, 2004.

[5] R. H. Bordini and J. F. Hübner. Jason: A Java-based AgentSpeak interpreter used with saci for multi-agent distribution over the net, 2005. http://jason.sourceforge.net/.

[6] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[7] P. Gärdenfors, editor. *Belief Revision.* Cambridge University Press, 1992.

[8] K. V. Hindriks, F. S. de Boer, W. van der Hoek, and J.-J. Meyer. Agent programming in 3apl. *Autonomous Agents and Multi-Agent Systems*, 2(4):357–401, 1999.

[9] J. Hulstijn and L. van der Torre. Combining goal generation and planning in an argumentation framework. In A. Hunter and J. Lang, editors, *Proc. Workshop on Argument, Dialogue and Decision, at NMR*, Whistler, Canada, June 2004.

[10] A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proc. 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 883–890, Melbourne, Australia, 2003.

[11] H. Prakken and G. Sartor. Argument-based logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75, 1997.

[12] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer, Netherlands, 2002.

[13] A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–342, June 1998.

[14] G. R. Simari, A. J. Garcia, and M. Capobianco. Actions, planning and defeasible reasoning. In *Proc. 10th International Workshop on Non-Monotonic Reasoning*, pages 377–384, 2004.

[15] Y. Tang and S. Parsons. Argumentation-based dialogues for deliberation. In F. Dignum et al, editor, *Proc. AAMAS, Utrecht, The Netherlands*, pages 552–559, New York NY, USA, 2005. ACM Press.