Journal of Universal Computer Science, vol. 15, no. 13 (2009), 2653-2675
submitted: 31/10/08, accepted: 13/6/09, appeared: 1/7/09 © J.UCS

Extended Defeasible Reasoning for Common Goals in n-Person Argumentation Games

Duy Hoang Pham
(Posts and Telecommunications Institute of Technology
Faculty of Information Technology
Hanoi, Vietnam
phamhduy@gmail.com)

Guido Governatori
(NICTA, Queensland Research Laboratory
Brisbane, Australia
guido.governatori@nicta.com.au)

Subhasis Thakur
(Griffith University and
NICTA, Queensland Research Laboratory
Brisbane, Australia
s.thakur@griffith.edu.au)

Abstract: Argumentation games have been proved to be a robust and flexible tool to resolve conflicts among agents. An agent can propose its explanation and its goal known as a claim, which can be refuted by other agents. The situation is more complicated when there are more than two agents playing the game.

We propose a weighting mechanism for competing premises to tackle with conflicts from multiple agents in an n-person game. An agent can defend its proposal by giving a counter-argument to change the "opinion" of the majority of opposing agents. Furthermore, using the extended defeasible reasoning an agent can exploit the knowledge that other agents expose in order to promote and defend its main claim.

Key Words: Artificial intelligence, Defeasible reasoning, Argumentation systems
Category: I.2.4

1 Introduction

In multi-agent systems, there are several situations requiring a group of agents to settle on common goals despite each agent's pursuit of individual goals which may conflict with other agents. To resolve the conflicts, an agent can argue to convince others about its pursued goal and provides evidence to defend its claim. This interaction can be modelled as an argumentation game [Prakken and Sartor 1996, Jennings et al. 1998, Parsons and McBurney 2003]. In an argumentation game, an agent can propose an explanation for its goal (i.e., an argument), which can be rejected by counter-evidence from other agents. This action can be iterated until an agent either successfully argues its proposal against other agents or drops its initial claim.

The argumentation game approach offers a robust and flexible tool to resolve conflicts by evaluating the status of arguments from agents. Dung's argumentation semantics [Dung 1995] is widely recognised to establish relationships (undercut, defeated, and accepted) among arguments. The key notion for a set of arguments is whether a set of arguments is self-consistent and provides the basis to derive a conclusion.

An argumentation game is more complicated when the group has more than two agents. It is not clear how to extend existing approaches to resolve conflicts from multiple agents, especially when agents have equal weight. In this case, the problem amounts to deciding which argument has precedence over competing arguments. The main idea behind our approach is the global collective preference over individual proposals, which enables an agent to identify the key arguments and premises from opposing agents in order to generate counter-arguments. These arguments cause a majority of opposing agents to reconsider their claims, therefore, an agent has an opportunity to change "attitudes" of others. In our approach, an agent in the argumentation game cooperates with the majority of the group if the agent fails to defend its proposal against those of the majority. At the end of the game, the agent follows the proposals accepted by the majority of the group.

Each of our agents is equipped with its private knowledge, background knowledge, and knowledge obtained from other agents. The private knowledge of an agent represents its own understanding about the game including its private goals. The background knowledge, commonly shared by the group, presents the expected behaviours of a member of the group and can include the description of the environment. Any argument violating the background knowledge is not supported by the group. The background knowledge also represents the primitive set of goals, which are recognised by the group of agents. The knowledge about other agents, growing during the game, enables an agent to efficiently convince others about its own goals.

Defeasible logic is chosen as our underlying logic for the argumentation game due to its efficiency and simplicity in representing incomplete and conflicting information. Furthermore, Defeasible Logic has a powerful and flexible reasoning mechanism [Antoniou et al. 2000, Maher et al. 2001] which enables our agents to flawlessly capture Dung's argumentation semantics by using two features of defeasible reasoning, namely the ambiguity propagating and ambiguity blocking.

Our paper is structured as follows. In section 2, we briefly introduce the basic notions of defeasible logic and the construction of the argumentation semantics. Section 3 introduces our n-person argumentation game framework using defeasible logic. We present firstly the external model of agents' interaction, which describes a basic procedure for an interaction between agents. Secondly, we elaborate the extension of the defeasible reasoning with a superior theory. Thirdly, we define the internal model, which shows how an agent can deal with individual knowledge sources to propose and defend its goal against other agents. Finally, we show the justification of arguments generated by an agent during the game with respect to (w.r.t.) the background knowledge of the group. Section 4 provides an overview of research works related to our approach. Section 5 concludes the paper.

2 Background

In this section, we briefly present the essential notions of defeasible logic (DL) and the construction of Dung's argumentation semantics by using two features of defeasible reasoning including ambiguity blocking and propagating.

2.1 Defeasible Logic

Following the presentation in [Billington 1993], a defeasible theory D consists of a finite set of facts F; a finite set of rules R; and a superiority relation $>$ on R. The language of defeasible theory is based on a finite set of literals. Given a literal l, we use $\sim l$ to indicate the complement of l.

A rule r in R is composed of an antecedent (body) $A(r)$, a consequent (head) $C(r)$, and a connective between $A(r)$ and $C(r)$. $A(r)$ consists of a finite set of literals and $C(r)$ contains a single literal. $A(r)$ can be omitted from the rule if it is empty. The set of connectives includes $\rightarrow$, $\Rightarrow$, and $\rightsquigarrow$, which represents strict rules R_s, defeasible rules R_d, and defeaters R_{dft} in R respectively. We define R_{sd} as the set of strict and defeasible rules, and $R[q]$ as the set of rules whose heads are q.

A conclusion derived from the theory D is a tagged literal and is categorised according to how the conclusion can be proved: $+\Delta q$: q is definitely provable in D; $-\Delta q$: q is definitely unprovable in D; $+\partial q$: q is defeasibly provable in D; $-\partial q$: q is defeasibly unprovable in D.

Provability is based on the concept of a derivation (or proof) in a defeasible theory $D = (F, R, >)$. Informally, definite conclusions can derive from strict rules by forward chaining, while defeasible conclusions can obtain from defeasible rules if and only if all possible "attacks" are rebutted due to the superiority relation or defeater rules. The set of conclusions of a defeasible theory is finite. This set is the Herbrand base that can be built from the literals occurring in the rules and the facts of the theory.

A derivation is a finite sequence $P = (P(1), \ldots, P(n))$ of tagged literals satisfying proof conditions (which correspond to inference rules for each of the four kinds of conclusions). $P(1..i)$ denotes the initial part of the sequence P of length i. In the follows, we present the proof conditions for definitely and defeasibly provable conclusions by [Antoniou et al. 2001].

Definition 1. The condition for a conclusion with tag $+\Delta$ is defined as:

$+\Delta$: If $P(i+1) = +\Delta q$ then

(1) $q \in F$ or
(2) $\exists r \in R_s[q]\ \forall a \in A(r) : +\Delta a \in P(1..i)$

The definition of $+\Delta$ describes just forward chaining of strict rules. For a literal q to be definitely provable there is a strict rule with head q, of which all antecedents have been definitely proved previously.

Definition 2. The condition for a conclusion with tag $-\Delta$ is defined as:

$-\Delta$: If $P(i+1) = -\Delta q$ then
(1) $q \notin F$ or
(2) $\forall r \in R_s[q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$

To show that q cannot be proven definitely, q must not be a fact. In addition, we need to establish that every strict rule with head q is known to be inapplicable. Thus, for every such rule r there must be at least one antecedent a for which we have established that a is not definitely provable $-\Delta q$.

Definition 3. The condition for a conclusion with tag $+\partial$ is defined as:

$+\partial$: If $P(i+1) = +\partial q$ then either
(1) $+\Delta q \in P(1..i)$ or
(2.1) $\exists r \in R_{sd}[q]\ \forall a \in A(r) : +\partial a \in P(1..i)$ and
(2.2) $-\Delta{\sim}q \in P(1..i)$ and
(2.3) $\forall s \in R_{sd}[{\sim}q]$ either
(2.3.1) $\exists a \in A(s) : -\partial a \in P(1..i)$ or
(2.3.2) $\exists t \in R_{sd}[q]$ such that $t > s$ and
$\forall a \in A(t) : +\partial a \in P(1..i)$

To show that q is provable defeasibly, it is more complicated since opposing chains of reasoning against q must be considered: (1) q is already definitely provable; or (2) the defeasible part of D is investigated. In particular, it is required that a strict or defeasible rule with head q which can be applied is in the theory (2.1). In addition, the possible "attacks" must be taken into account. To be more specific: q is defeasibly provable providing that ${\sim}q$ is not definitely provable (2.2); this is to ensure that the logic does not derive conflicting conclusions. Also (2.3) the set of all rules supporting ${\sim}q$ are considered. Essentially, each such a rule s attacks the conclusion q. The conclusion q is provable if each such rule s is not applicable or s must be counter-attacked by a rule t with head q and t must be stronger than s.

Definition 4. The condition for a conclusion with tag $-\partial$ is defined as:

$-\partial$: If $P(i+1) = -\partial q$ then either
(1) $-\Delta q \in P(1..i)$ or
(2.1) $\forall r \in R_{sd}[q]\ \exists a \in A(r) : -\partial a \in P(1..i)$ or

(2.2) $+\Delta{\sim}q \in P(1..i)$ or
(2.3) $\exists s \in R_{sd}[{\sim}q]$ such that
  (2.3.1) $\forall a \in A(s) : +\partial a \in P(1..i)$ and
  (2.3.2) $\forall t \in R_{sd}[q]$ either $t \not> s$ or
    $\exists a \in A(t) : -\partial a \in P(1..i)$

The similar explanation is applied for proving $-\partial q$. In short, the theory D does not have any strict rule supporting q and one of following conditions: all defeasible rules for q are not applicable; there is a strict support for ${\sim}q$; at least one defeasible rule for ${\sim}q$ is applicable and successfully overrides the "attack" from those rules for q.

The defeasible reasoning has the properties of coherence (Definition 5) and consistency (Definition 6).

Definition 5. A defeasible theory is *coherent* if it is impossible to derive from it a pair $-\Delta q$ and $+\Delta q$, or $-\partial q$ and $+\partial q$.

Definition 6. A defeasible theory is *consistent* if it is possible to derive $+\partial q$ and $+\partial{\sim}q$ if and only if the theory derives both $+\Delta q$ and $+\Delta{\sim}q$.

2.2 Defeasible Logic with Ambiguity Propagation

The version presented in the previous section is ambiguity blocking. However, it is possible to provide an ambiguity propagating variant of the logic [Governatori et al. 2004, Antoniou et al. 2000]. The superiority relation is not considered in the inference process of this variant. The extension introduces a new tag Σ, which shows a support for a literal in a defeasible theory. $+\Sigma p$ means that there is a monotonic chain of reasoning that would lead to conclude p in the absence of conflicts. Thus, a defeasibly provable literal tagged with $+\partial$ is also supported. In contrast, a literal may be supported even though it is not defeasibly provable. Therefore, support is a weaker notion than defeasible provability. In the follows, we present the extension conditions for $\pm\Sigma$ conclusions with respect to the superiority relationship among defeasible rules.

Definition 7. The positive support for a literal is defined as:

$+\Sigma$: If $P(i+1) = +\Sigma q$ then
  $\exists r \in R_{\mathrm{sd}}[q]$: $\forall a \in A(r) : +\Sigma a \in P(1..i)$ and
  $\forall s \in R_{\mathrm{sd}}[{\sim}q]$ either $\exists a \in A(s) : -\partial a \in P(1..i)$ or
    $\exists t \in R_{\mathrm{sd}}[q]$ such that $t > s$ and $\forall a \in A(t) : +\Sigma a \in P(1..i)$

Definition 8. The negative support for a literal is constructed as:

$-\Sigma$:If $P(i+1) = -\Sigma q$ then
  $\forall r \in R_{sd}[q]$: $\exists a \in A(r) : -\Sigma a \in P(1..i)$ or
  $\exists s \in R_{\mathrm{sd}}[\neg q]$ and $\forall a \in A(s) : +\partial a \in P(1..i)$ and
    $\forall t \in R_{\mathrm{sd}}[q]$ such that $t \not> s$ or $\exists a \in A(t) : -\Sigma a \in P(1..i)$

We can achieve ambiguity propagation behaviour by making a minor change to the inference conditions for $+\partial_{AP}$ and $-\partial_{AP}$.

Definition 9. The condition for a positive defeasible conclusion with respect to ambiguity is defined as:

$+\partial_{AP}$: If $P(i+1) = +\partial_{AP}q$ then either
(1) $+\Delta q \in P(1..i)$ or
  (2.1) $\exists r \in R_{sd}[q]\ \forall a \in A(r) : +\partial_{AP}a \in P(1..i)$ and
  (2.2) $-\Delta{\sim}q \in P(1..i)$ and
  (2.3) $\forall s \in R_{sd}[{\sim}q]$
    $\exists a \in A(s) : -\Sigma a \in P(1..i)$r1

The conditions for $-\partial_{AP}$ is obtained from that of $+\partial_{AP}$ using the principle of the strong negation [Antoniou et al. 2000, Antoniou et al. 2006].

Definition 10. The condition for a unprovable defeasible conclusion with respect to ambiguity is constructed as:

$-\partial_{AP}$: If $P(i+1) = -\partial q$ then either
(1) $-\Delta q \in P(1..i)$ and
  (2.1) $\forall r \in R_{sd}[q]\ \exists a \in A(r) : -\partial_{AP}a \in P(1..i)$ or
  (2.2) $+\Delta{\sim}q \in P(1..i)$ or
  (2.3) $\exists s \in R_{sd}[{\sim}q]$ such that
    $\forall a \in A(s) : +\Sigma a \in P(1..i)$

In the following example, we illustrate the use of support notion and the inference with ambiguity.

Example 1. Considering a defeasible theory D where $R_d = \{r_1 :\Rightarrow a;\ r_2 :\Rightarrow {\sim}a;\ r_3 :\Rightarrow b;\ r_4 : a \Rightarrow {\sim}b\}$

Without the superiority relationship, there is no means to decide between a and ${\sim}a$ and both r_1 and r_2 are applicable. In a setting where the ambiguity is blocked, b is not ambiguous because r_3 for b is applicable whilst r_4 is not since its antecedent is not provable. If the ambiguity is propagated, we have evidence supporting all of four literals since all of the rules is applicable. $+\Sigma a, +\Sigma{\sim}a, +\Sigma b$ and $+\Sigma{\sim}b$ are included in the conclusion set. Moreover we can derive $-\partial a$, $-\partial{\sim}a$, $-\partial b$ and $-\partial{\sim}b$ showing that the resulting logic exhibits an ambiguity propagating behaviour. In the second setting b is ambiguous, and its ambiguity depends on that of a.

2.3 Argumentation Semantics

In what follows, we briefly introduce the basic notions of an argumentation system using defeasible reasoning [Governatori et al. 2004]. We also present the acceptance of an argument w.r.t. Dung's semantics.

Definition 11. An argument A for a literal p based on a set of rules R is a (possibly infinite) tree with nodes labelled by literals such that the root is labelled by p and for every node with label h:

1. If $b_1,\dots,b_n$ label the children of h then there is a rule in R with body $b_1,\dots,b_n$ and head h.
2. If this rule is a defeater then h is the root of the argument.
3. The arcs in a proof tree are labelled by the rules used to obtain them.

A (proper) sub-argument of an argument A is a (proper) subtree of the proof tree associated to A.

DL requires a more general notion of proof tree that admits infinite trees, so that the distinction is kept between an infinite un-refuted chain of reasoning and a refuted chain. Depending on the rules used, there are different types of arguments: (1) A supportive argument is a finite argument in which no defeater is used; (2) A strict argument is an argument in which only strict rules are used; (3) An argument that is not strict is called defeasible.

Example 2. Consider the following defeasible theory

$$D = \{\Rightarrow e;\ e \leadsto f;\ \Rightarrow p;\ p \Rightarrow q;\ \rightarrow \neg q\}$$

From D we can build a set of argument $Args^{\mathrm{D}} = \{\Rightarrow e \leadsto f;\ \Rightarrow p \Rightarrow q;\ \rightarrow \neg q\}$. The argument for f is not supportive because of the defeater rule $e \leadsto f$ whilst the argument for q is a supportive and defeasible. There is one strict argument for $\neg q$ in $Args^{\mathrm{D}}$.

Relationships between two arguments A and B are determined by literals constituting these arguments. An argument A *attacks* a defeasible argument B if a literal of A is the complement of a literal of B, and that literal of B is not part of a strict sub-argument of B. A set of arguments $\mathscr{S}$ attacks a defeasible argument B if there is an argument A in $\mathscr{S}$ that attacks B.

A defeasible argument A is *undercut* by a set of arguments $\mathscr{S}$ if $\mathscr{S}$ supports an argument B attacking a proper non-strict sub-argument of A. An argument A is undercut by $\mathscr{S}$ means some literals of A cannot be proven if we accept the arguments in $\mathscr{S}$.

The concepts of the attack and the undercut concern only defeasible arguments and sub-arguments. A defeasible argument is assessed as valid if we can show that the premises of all arguments attacking it cannot be proved from the valid arguments in $\mathscr{S}$. The concepts of provability depend on the method used by the reasoning mechanism to tackle ambiguous information. According to the features of the defeasible reasoning, we have the definition of acceptable arguments (Definition 12).

Definition 12. An argument A for p is acceptable w.r.t. a set of arguments $\mathscr{S}$ if A is finite, and

1. If reasoning with the ambiguity propagation is used: (a) A is strict, or (b) every argument attacking A is attacked by $\mathscr{S}$.

2. If reasoning with the ambiguity blocking is used: (a) A is strict, or (b) every argument attacking A is undercut by $\mathscr{S}$.

The status of an argument is determined by the concept of acceptance. If an argument can resist a reasonable refutation, this argument is justified (Definition 13). If an argument cannot overcome attacks from other arguments, this argument is rejected (Definition 15). We define the sets of justified arguments w.r.t. a set of arguments constructed from a defeasible theory D as follows:

Definition 13. Let D and $Args^{\mathrm{D}}$ be a defeasible theory and a set of arguments from D respectively. We define J_i^{D} as follows.

- $J_0^{\mathrm{D}} = \emptyset$
- $J_{i+1}^{\mathrm{D}} = \{a \in Args^{\mathrm{D}} | \ a$ is acceptable w.r.t. $J_i^{\mathrm{D}}\}$

The set of justified arguments in a defeasible theory D is $JArgs^{\mathrm{D}} = \bigcup_{i=1}^{\infty} J_i^{\mathrm{D}}$.

Typically argumentation semantics frameworks just consider justified arguments and justified conclusions. However, as we have seen in the previous sections, Defeasible Logic has both provable and unprovable conclusions. To capture that a conclusion is not provable, we have to introduce the notion of rejected argument.

Definition 14. An argument A for p is rejected w.r.t. to a set of arguments $\mathscr{S}$ (set of already rejected arguments) and a set of arguments $\mathscr{T}$ (set of accepted arguments), if

1. If reasoning with ambiguity propagation is used: (a) a proper subargument of A is in $\mathscr{S}$ or (b) it is attacked by a finite argument.

2. If reasoning with ambiguity blocking is used: (a) a proper subargument of A is in $\mathscr{S}$ or (b) it is attacked by an argument supported by $\mathscr{T}$.

Notice that in the above definition, for the case of ambiguity propagation that the set of accepted argument is not used.

Definition 15. Let D be a defeasible theory and $\mathscr{T}$ be a set of arguments. We define $R_i^{\mathrm{D}}(\mathscr{T})$ as follows.

- $R_0^{\mathrm{D}}(\mathscr{T}) = \emptyset$
- $R_{i+1}^{\mathrm{D}}(\mathscr{T}) = \{$ a $\in Args^{\mathrm{D}}|$ a is rejected by $R_i^{\mathrm{D}}(\mathscr{T})$ and $\mathscr{T}\}$.

The set of rejected arguments in a defeasible theory D w.r.t. $\mathscr{T}$ is $RArgs^{\mathrm{D}}(\mathscr{T}) = \bigcup_{i=1}^{\infty} R_i^{\mathrm{D}}(\mathscr{T})$.

For a defeasible theory the set of rejected argument is given by $RArgs^{\mathrm{D}}(JArgs^{\mathrm{D}})$, i.e., we have the set of arguments rejected by $JArgs^{\mathrm{D}}$.

We are now able to extend the notions of justified and rejected to conclusions of a defeasible theory.

Definition 16. Given a defeasible theory D:

- a literal a is justified if it is the conclusion of an argument in $JArgs^{\mathrm{D}}$.
- A literal a is rejected by $\mathscr{S}$ if there is no argument in $JArgs^{\mathrm{D}} - RArgs^{\mathrm{D}}(\mathscr{S})$, the top rule of which is a strict or defeasible rule with head a. A literal is rejected if it is rejected by $JArgs^{\mathrm{D}}$.

3 n-Person Argumentation Game

In this section, we utilise the argumentation semantics presented in Section 2.3 to model agents' interactions in an n-person argumentation game. Also, we propose a knowledge structure which enables an agent to construct its arguments w.r.t. knowledge from other agents as well as to select a defensive argument.

3.1 Agents' Interactions

In our approach to n-person argumentation, a group of agents $\mathscr{A}$ shares a set of primitive goals $\mathscr{G}$ and a set of external constraints T_{bg} represented as a defeasible theory, known as a background knowledge. This knowledge provides common expectations and restrictions to the agents in $\mathscr{A}$. An agent has its own view on the working environment, therefore, can autonomously pursue its own goals. The agent convinces the other agents by presenting the arguments to defend its goals. An agent can recall the arguments played by the others but does not know the knowledge, which generates those arguments. That is an agent has a partial view on the knowledge of the other agents.

In this work, we model interactions between agents to settle on goals commonly accepted by the group. Also, at each step of the game, we show how an agent can identify a goal and sub-goals for its counter arguments. This information is critical for those agents whose main claims are refuted either directly by arguments from other agents or indirectly by the combination of these arguments.

3.1.1 Settling on Common Goals

An agent can pursue a goal in the set of common goals $\mathscr{G}$ by proposing an explanation for its goal. The group justifies proposals from individual agents in order to identify commonly-accepted goals using a dialogue as follows:

1. Each agent broadcasts an argument for its goal. The system can be viewed as an argumentation game with n-players corresponding to the number of agents.

2. An agent checks the status of its argument against those from the other agents. There are three possibilities, the argument is: (a) *directly refuted* if its argument conflicts with those from others; (b) *collectively refuted* if its argument does not conflict with individual arguments but violates the combination of individual arguments (see Section 3.3.1); (c) *collectively accepted* if its argument is justified by the combination (see Section 3.4.2).

3. According to the status of its main claim, an agent can: (a) defend its claim; (b) attack a claim from other agents; (c) rest.

4. The dialogue among agents is terminated if all agents pass their claims. For a dispute, agents stop arguing if they do not have any more argument to propose.

3.1.2 Weighting Opposite Premises

In a dialogue, at each step an agent is required to identify goals and sub-goals which are largely shared by other agents. This information is highly critical for agents, whose main claims are refuted either directly by other agents or collectively by the combination of arguments from others in order to effectively convince other agents.

To achieve that an agent, A_{me}, identifies a sub-group of agents, namely "opp-group", which directly or collectively attacks its main claim at step i. A_{me} creates $Args_i^{\mathrm{opp}}$ as the set of opposing arguments from the opp-group and P_i^{opp} as the set of premises in $Args_i^{\mathrm{opp}}$. Essentially, $Args_i^{\mathrm{opp}}$ contains arguments attacking A_{me}'s claim. Each element of P_i^{opp} is weighted by its frequency in $Args_i^{\mathrm{opp}}$. We define the preference over P_i^{opp} as given $p_1, p_2 \in P_i^{\mathrm{opp}}$, $p_2 \succeq p_1$ if the frequency of p_2 in $Args_i^{\mathrm{opp}}$ is greater than that of p_1. Because $Args_i^{\mathrm{opp}}$ is the set of arguments played by the opp-group, the more frequent an element $q \in P_i^{\mathrm{opp}}$ is the more agents use this premise in their arguments. Therefore, the refutation of q challenges other agents better than the premises having lower frequency since this refutation causes a larger number of agents to reconsider their claims.

3.1.3 Defending the Main Claim

At iteration i, $Args_i^{\mathrm{opp}}$ represents the set of arguments played by the opp-group:

$$Args_i^{\mathrm{opp}} = \bigcup_{j=0}^{|\mathscr{A}|} Arg_i^{\mathrm{A_j}} | Arg_i^{\mathrm{A_j}} \text{ directly attacks } Arg_i^{\mathrm{A_{me}}}$$

where $Arg_i^{\mathrm{A_j}}$ is the argument played by agent A_j at step i. If A_j rests at iteration i, its last argument (at iteration k) is used $Arg_i^{\mathrm{A_j}} = Arg_k^{\mathrm{A_j}}$. $Arg_i^{\mathrm{A_{me}}}$ is the argument played by A_{me} at step i. The set of opposite premises at iteration i is:

$$P_i^{\mathrm{opp}} = \{p | p \in Args_i^{\mathrm{opp}} \text{ and } p \notin Arg_i^{\mathrm{A_{me}}}\}$$

The preference over elements of P^{opp} provides a mechanism for A_{me} to select arguments for defending its main claim.

Example 3. Suppose that agent A_1 and A_2 respectively propose $Arg^{A_1} = \{\Rightarrow e \Rightarrow b \Rightarrow a\}$ and $Arg^{A_2} = \{\Rightarrow e \Rightarrow c \Rightarrow a\}$ whilst agent A_3 claims $Arg^{A_3} = \{\Rightarrow d \Rightarrow \sim a\}$. From A_3's view, its claim directly conflicts with those of A_1 and A_2. The arguments and premises of the opp-group are:

$$Args_i^{\text{opp}} = \{\Rightarrow e \Rightarrow b \Rightarrow a;\ \Rightarrow e \Rightarrow c \Rightarrow a\} \text{ and } P_i^{\text{opp}} = \{a^2, b^1, c^1, e^2\}$$

The superscript of elements in P_i^{opp} represents the frequency of a premise in $Args_i^{\text{opp}}$. A_3 can defend its claim by providing a counter-argument that refute $\sim a$ – the major claim of the opp-group. Alternatively, A_3 can attack either b or c or e in the next step. An argument against e is the better selection compared with those against b or c since A_3's refutation of e causes both A_1 and A_2 to reconsider their claims.

3.1.4 Attacking an Argument

In this situation, individual arguments of other agents do not conflict with that of A_{me} but the integration of these arguments does. Agent A_{me} should argue against one of these arguments in order to convince others about its claim.

At iteration i, let the integration of arguments be $T_{\text{INT}}^i = T_{\text{bg}} \bigcup_{j=0}^{|\mathscr{A}|} T_j^i$, where T_j^i is the knowledge from agent j supporting agent j's claim, and $JArgs^{T_{\text{INT}}^i}$ be the set of justified arguments from integrated knowledge of other agents (see Section 3.4.2). The set of opposite arguments is defined as:

$$Args_i^{\text{opp}} = \{a | a \in JArgs^{T_{\text{INT}}^i} \text{ and } a \text{ is attacked by } Arg_i^{A_{\text{me}}}\}$$

and the set of opposite premises is:

$$P_i^{\text{opp}} = \{p | p \in Args_i^{\text{opp}} \text{ and } (p \notin Arg_i^{A_{\text{me}}} \text{ or } p \text{ is not attacked by } Arg_i^{A_{\text{me}}})\}$$

The second condition is that A_{me} is self-consistent and does not play any argument against itself. In order to convince other agents about its claim, A_{me} is required to provide arguments against any premise in P^{opp}. In fact, the order of elements in P^{opp} offers a guideline for A_{me} on selecting its attacking arguments.

3.2 Extended Defeasible Reasoning

In this section, we propose a simple method to integrate two independent defeasible theories. Note that a defeasible theory has finite sets of facts and rules, and a derivation from the theory can be computed in linear time [Maher 2001]. In addition, we revise the argumentation semantics with respect to the extended reasoning mechanism.

3.2.1 Defeasible Reasoning with a Superior Theory

Suppose that an agent considers two knowledge sources represented by defeasible theories labeled as T_{sp} – the superior theory, and T_{in} – the agent's internal theory. The agent considers that T_{sp} has higher level of importance than T_{in}. That relationship is represented as $T_{\text{sp}} \succ T_{\text{in}}$. Thus, conclusions from the internal theory should be withdrawn if they conflict with the superior theory; the agent prefers the superior theory's conclusions to its own. We denote that reasoning process over T_{sp} and T_{in} as $T_{\text{sp}} \ni T_{\text{in}}$. It is noticed that T_{sp} and T_{in} are coherent and consistent.

Thanks to the transformations of [Antoniou et al. 2001] to remove the superiority relation and defeater rules, we can assume that the two theories contain only strict and defeasible rules. To perform the defeasible reasoning, the agent generates a superiority relation over sets of rules as in $R^{sp} > R^{in}$, that is the rules from T_{sp} are stronger than the rules from T_{in}. In this scheme, the subscript denotes the type of rules while the superscript indicates the type of the theory which contains the rules.

A definite conclusion e.g. $+\Delta q$ is derived by performing forward chaining with the strict rules in the superior theory, or in the internal theory if the complementary literals cannot be positively proved by the superior theory. In other words, $T_{\text{sp}} \ni T_{\text{in}}$ derives $+\Delta q$ if T_{sp} has a proof for q or T_{in} can definitely prove q provided that there is no (definite) support for $\sim q$ from T_{sp}.

$+\Delta$: If $P(i+1) = +\Delta q$ then
(1) $q \in F$ or
(2) $\exists r \in R_s^{sp}[q]\ \forall a \in A(r) : +\Delta a \in P(1..i)$ or
(3) $\exists t \in R_s^{in}[q]\ \forall a \in A(t) : +\Delta a \in P(1..i)$ and
$\forall r \in R_s^{sp}[\sim q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$

The conclusions tagged with $-\Delta$ mean that the extended mechanism cannot produce a positive proof for the corresponding literals from the strict parts of the both theories or the superior theory holds the complements.

$-\Delta$: If $P(i+1) = -\Delta q$ then
(1) $q \notin F$ and
(2) $\forall r \in R_s^{sp}[q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$ and
(3) $\forall r \in R_s^{in}[q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$ or
$\exists t \in R_s^{sp}[\sim q]\ \forall a \in A(t) : +\Delta a \in P(1..i)$

A defeasible conclusion $+\partial q$ can either be drawn directly from definite conclusions, or by investigating the defeasible part of the integrated theory. In particular, it is required that a strict or defeasible rule with an "applicable head" q is in the theory (2.1). In addition, the possible "attacks" must be either unprovable (2.2 and 2.3.1) or counter-attacked by "stronger" rules (2.3.2).

$+\partial$: If $P(i+1) = +\partial q$ then either

(1) $+\Delta q \in P(1..i)$ or
 (2.1) $\exists r \in R^{sp}_{sd}[q] \cup R^{in}_{sd}[q]\ \forall a \in A(r) : +\partial a \in P(1..i)$ and
 (2.2) $-\Delta{\sim}q \in P(1..i)$ and
 (2.3) $\forall s \in R^{sp}_{sd}[{\sim}q] \cup R^{in}_{sd}[{\sim}q]$ either
  (2.3.1) $\exists a \in A(s) : -\partial a \in P(1..i)$ or
  (2.3.2) $\exists t \in R^{sp}_{sd}[q] \cup R^{in}_{sd}[q]$ such that $t > s$ and
   $\forall a \in A(t) : +\partial a \in P(1..i)$

The conclusions tagged with $-\partial$ mean that the extended mechanism cannot retrieve a positive proof for the corresponding literals from the strict and defeasible rules of both theories or these conclusions are rebutted because of "stronger" conclusions. The proof for $-\partial$ derives from that of $+\partial$ by using the strong negation principle [Antoniou et al. 2000].

$-\partial$: If $P(i+1) = -\partial q$ then
(1) $-\Delta q \in P(1..i)$ and
 (2.1) $\forall r \in R^{sp}_{sd}[q] \cup R^{in}_{sd}[q]\ \exists a \in A(r) : -\partial a \in P(1..i)$ or
 (2.2) $+\Delta{\sim}q \in P(1..i)$ or
 (2.3) $\exists s \in R^{sp}_{sd}[{\sim}q] \cup R^{in}_{sd}[{\sim}q]$ such that
  (2.3.1) $\forall a \in A(s) : +\partial a \in P(1..i)$ and
  (2.3.2) $\forall t \in R^{sp}_{sd}[q] \cup R^{in}_{sd}[q]$ either $t \not> s$ or
   $\exists a \in A(t) : -\partial a \in P(1..i)$

The extended mechanism reconstructs the superiority relation among rules from that of the theories T_{sp} and T_{in}. Therefore, the standard conditions for defeasible conclusions can be applied for the defeasible part of the combined theory. It is noticed that the conditions for $\pm\Delta$ are different. Essentially, the meaning of definite conclusions is slightly changed. Only definite conclusions from the superior theory T_{sp} are not refutable but those conclusions from the internal theory T_{in} can be rejected by those from T_{sp}.

Given two defeasible theories T and S and a proof tag #, we use $T \vdash \#q$ to mean that $\#q$ can be proved from theory T using the basic proof conditions of DL. Meanwhile $T \ni S \vdash \#q$ means that there is a derivation of $\#q$ from the theory integrating T and S using the proof conditions given in this section and where T plays the role of T_{sp} and S the role of T_{in}.

The properties of the extended reasoning with the superior knowledge are represented as propositions. If a strict conclusion is derived from the superior theory, the extended mechanism does not provide any proof for its negation (Proposition 17). The conclusions from the extended mechanism can violate defeasible conclusions obtained from the superior theory if the agent has a strong evidence of the contradiction in its internal knowledge (Proposition 18). The extended reasoning maintains the coherence and consistent of the standard reasoning (Proposition 19).

Proposition 17. *If $T_{\text{sp}} \vdash +\Delta q$ then $T_{\text{sp}} \ni T_{\text{in}} \not\vdash +\Delta{\sim}q$ and $T_{\text{sp}} \ni T_{\text{in}} \not\vdash +\partial{\sim}q$.*

Proof. This result directly draws from the proof conditions of our reasoning mechanism. Assume that q is derived from the strict rules of the superior theory, $T_{\text{sp}} \vdash +\Delta q$, while $+\Delta{\sim}q$ is computed by the integrated theory, $T_{\text{sp}} \ni T_{\text{in}} \vdash +\Delta{\sim}q$. With regard to the proof for $+\Delta$, $+\Delta{\sim}q$ can be derived either from (1) or (2) that is $T_{\text{sp}} \vdash +\Delta{\sim}q$. But this violates the assumption. If $+\Delta{\sim}q$ is proved by (3), this proof condition requires the strict rules from the superior theory do not have any support for q, $T_{\text{sp}} \not\vdash +\Delta q$. Again, this contradicts the assumption. Hence the first part of the proposition is proved. $T_{\text{sp}} \vdash +\Delta q$ blocks the derivation of $+\Delta{\sim}q$ and $-\Delta{\sim}q$ from the integrated theory. That means (1) and (2.2) are not satisfied in the proof of $+\partial$. Consequently, $+\partial{\sim}q$ is blocked.

Proposition 18. *If $T_{\text{sp}} \vdash -\Delta{\sim}q$ and $T_{\text{in}} \vdash +\Delta q$ then $T_{\text{sp}} \ni T_{\text{in}} \vdash +\partial q$.*

Proof. Assume that the integrated theory derives the contradiction, $T_{\text{sp}} \ni T_{\text{in}} \vdash +\partial{\sim}q$, given the conditions of the proposition. According to the proof of $+\partial$, $+\partial{\sim}q$ requires $T_{\text{sp}} \ni T_{\text{in}}$ to prove either $+\Delta{\sim}q$ or $-\Delta q$ since the conditions of (2.1) and (2.3) can be met due to the superiority of T_{sp} over T_{in}. As shown in the proof of Δ, the derivation of $+\Delta{\sim}q$ and $-\Delta q$ from $T_{\text{sp}} \ni T_{\text{in}}$ respectively requires $T_{\text{sp}} \vdash +\Delta{\sim}q$ or $T_{\text{in}} \vdash +\Delta{\sim}q$, $T_{\text{sp}} \not\vdash +\Delta q$ and $T_{\text{in}} \not\vdash +\Delta q$. Clearly, these requirements violate the assumptions of $T_{\text{sp}} \vdash {\sim}\Delta{\sim}q$ and $T_{\text{in}} \vdash +\Delta q$.

Proposition 19. *The extended reasoning mechanism is coherent and consistent (See Definition 5 and 6).*

Proof. First, we investigate the derivation of strict conclusions. By the definition of the standard defeasible logic, for each individual theory it is not possible to prove both $-\Delta q$ and $+\Delta q$. The only case leading to conflicting conclusions is where the superior theory supports q, $T_{\text{sp}} \vdash +\Delta q$, while the internal theory proves the contradictory, $T_{\text{in}} \vdash +\Delta{\sim}q$. According to the proof condition for $+\Delta$, the contradiction is always rejected. Therefore, the derivation of strict conclusions is coherent.

The coherence of the defeasible part of the reasoning with the superior theory directly inherits from that of the standard defeasible reasoning in the case that both theories have empty sets of facts and strict rules. The only case that can lead to the incoherence is $T_{\text{sp}} \vdash +\partial q$ and $T_{\text{in}} \vdash +\Delta{\sim}q$. That means the defeasible rules of the superior theory prove a conclusion conflicting with that supported by the strict rules of the internal theory. However, as mentioned above, the defeasible part of the superior theory is defeated by the strict part of the internal theory. That is, the derivation of defeasible conclusions is also coherent.

The consistency property follows from the coherence property. It is impossible to have conflicting conclusions from the proof conditions for defeasible conclusions. The only source of inconsistency comes from the proof for strict conclusions.

The extended mechanism computes a consistent set of conclusions with respect to the superior theory. The mechanism goes beyond the standard defeasible reasoning because it extends the superiority relation of rules to that of theories. This increases the size of theory to be investigated, but does not change the complexity of the reasoning mechanism, since the inference mechanism is not changed.

3.2.2 Argumentation Semantics

This section shows the properties of arguments built by the extended reasoning mechanism. First, we revise the proof for strict conclusions derived from the combination, therefore, strict arguments. Regard to the standard defeasible reasoning, we introduce the notion of defeasibility into the strict part of the combined theory. Essentially, a strict argument can be rejected if and only if that argument is constructed from a theory with lower priority. Otherwise, the argument is not rejected by any argument. We define the acceptance of a strict argument w.r.t. the extended reasoning in Definition 20.

Definition 20. In the extended reasoning, a strict argument A for p is strictly acceptable w.r.t. a set of strict arguments $\mathscr{S}$ if A is finite, and every argument attacking A is undercut by $\mathscr{S}$.

We now present the property of strict arguments constructed by the extended reasoning over two defeasible theories T_{sp} and T_{in}, where T_{sp} has priority over T_{in}.

Proposition 21. *Let T_{sp} and T_{in} be defeasible theories such that $T_{\text{sp}} \succ T_{\text{in}}$ and p be a literal.*

1. *$T_{\text{sp}} \ni T_{\text{in}} \vdash +\Delta p$ iff there is a strictly acceptable argument for p from $T_{\text{sp}} \ni T_{\text{in}}$.*
2. *$T_{\text{sp}} \ni T_{\text{in}} \vdash -\Delta p$ iff there is no strictly acceptable argument for p from $T_{\text{sp}} \ni T_{\text{in}}$.*

Proof. We prove the only if ($\Rightarrow$) direction of the proposition by induction on the length of a derivation P of the extended reasoning over T_{sp} and T_{in}.

At the first step of the derivation, $P(1) = +\Delta p$. That implies there is a strict rule, r, supporting p in $T_{\text{sp}} \ni T_{\text{in}}$ (Notice that a fact can be considered as a strict rule with an empty body). If r is in T_{sp} then there is a strict argument for p constructed from T_{sp}. That argument is self acceptable within T_{sp} due to the priority of T_{sp}. If r is in T_{in}, there is a strict supportive argument A for p constructed from T_{in}. Within T_{in} there is no argument against A as the standard reasoning. Corresponding to the extended condition for $+\Delta$, $P(1)$ holds only if there is no strict rule in T_{sp} supporting $\sim p$. In other words, there is no argument supporting $\sim p$ constructed from T_{sp}. Therefore, the argument A from T_{in} is acceptable w.r.t. T_{sp}.

At the first step, if $P(1) = -\Delta p$ then there is not any strict rule r supporting p in both T_{sp} and T_{in}. Therefore, it is not possible to have a strict argument for p in both theories.

At the inductive step, we assume that the proposition holds for derivations with length up to n. $P(n+1) = +\Delta p$. That is there exists a supportive argument A for p, which is built from a strict rule $r \in T_{\text{sp}} \cup T_{\text{in}}$ such that $\forall a_r \in A(r), +\Delta a_r \in P(1..n)$. Every a_r must be justified by inductive hypothesis. According to the standard reasoning, the strict rules from a theory do not support for both a_r and $\sim a_r$. Therefore, the attack results from the other theory of the combination $T_{\text{sp}} \ni T_{\text{in}}$. If $a_r \in T_{\text{sp}}$, the attacks from T_{in} is rejected due to the priority of T_{sp} over T_{in}. In the case that $a_r \in T_{\text{in}}$, we have $\forall r \in R_s^{sp}[\sim a_r]\ \exists a \in A(r) : -\Delta a \in P(1..n)$. By inductive hypothesis, there is no strict argument for $\sim a_r$ from T_{sp}. The argument for a_r from T_{in} is not attacked, hence, it is acceptable.

Assume that $P(n+1) = -\Delta p$, there are two possibilities. First, each strict rule r for p in T_{sp} and T_{in} has at least one literal a_r in the body such that $-\Delta a_r \in P(1..n)$. By inductive hypothesis, there is no strict argument for a_r, therefore, it is not possible to built a strict proof for p from T_{sp} and T_{in}. Second, there is a rule in T_{sp} supports the complement of p. By the inductive hypothesis for the positive proof, there is a strictly acceptable argument for $\sim p$. Hence, all of the arguments for p (from T_{in}) is not acceptable.

In what follows, we prove the if direction ($\Leftarrow$) of the proposition.

In the first part of the proposition, suppose that A is a strict argument for p having the height of 1. If A is built from T_{sp}, there is a strict rule with empty body for p in the combination $T_{\text{sp}} \cup T_{\text{in}}$. If A is built from T_{in} and accepted by T_{sp}, there is a strict rule for p in T_{in} and no rule for $\sim p$ in T_{sp}. In both case, there is a applicable rule for p in the combination, therefore, $T_{\text{sp}} \ni T_{\text{in}} \vdash +\Delta p$.

At the inductive step, we assume that the first part holds for arguments with height up to n and A is an argument for p. From A, we construct a strict rule as $A(r) \to p$. For every literal $a_r \in A(r)$, which is accepted by T_{sp}, we create sub-arguments of A having the height less than n. By inductive hypothesis we obtain $+\Delta a_r$, hence, the condition for $A(r) \to p$ is satisfied. Therefore, $T_{\text{sp}} \ni T_{\text{in}} \vdash +\Delta p$.

We prove the second part of the proposition by contradiction. Accordingly, assume that $T_{\text{sp}} \ni T_{\text{in}} \not\vdash -\Delta p$. That leads to: (1) $r \in R_s^{sp}[p] \forall a_r \in A(r)\ T_{\text{sp}} \ni T_{\text{in}} \not\vdash -\Delta a_r$; or (2) $s \in R_s^{in}[p] \forall a_s \in A(s)\ T_{\text{sp}} \ni T_{\text{in}} \not\vdash -\Delta a_r$ and $\forall t \in R_s^{sp}[\sim p] \exists a_t \in A(t)\ T_{\text{sp}} \ni T_{\text{in}} \not\vdash +\Delta a_t$.

For a strict rule for p in T_{sp}, we construct a partial argument A for p by expanding r. The expansion of the argument ends with three instances:

1. A rule with the empty body. That is there a strict argument for the literal from T_{sp}. That contradicts the assumption.

2. No more rule to expand, therefore, we have $-\Delta a_r$. That also contradicts with the assumption.

3. A loop. None of the literals of the loop can prove the adjacent literal. Therefore, we have $-\Delta a_r$. That also contradicts with the assumption.

For a strict rule for s in T_{in}, we construct a partial argument B for p by expanding s. Considering an argument C attacking B at q. If q is supported by T_{sp}, the attack is rejected because of the priority of T_{sp}. Hence, q is supported by a rule T_{in}. If E for $\sim q$ is constructed from T_{sp}, then that violates the assumption $\forall t \in R_s^{sp}[\sim q] \exists a_t \in A(t)\ T_{\text{sp}} \ni T_{\text{in}} \not\vdash +\Delta a_t$. If E for $\sim q$ is derived from T_{in}, that violates the coherence and consistency of the strict part of T_{in}. Therefore, E is not an acceptable argument. Also, B is not attack by any argument.

For both cases, the assumption is not valid, hence, the second part of the proposition is proved.

Example 4. This example shows the result of extending the superiority relation between theories to the strict parts of defeasible theories. Suppose that we have two defeasible theories $T_{\text{sp}} = \{R_s = \{r_1 :\to a;\ r_2 : a \to b\}\}$ and $T_{\text{in}} = \{R_s = \{r_1 :\to c;\ r_2 : c \to \sim b\}\}$ such that $T_{\text{sp}} \succ T_{\text{in}}$.

The extended $T_{\text{sp}} \ni T_{\text{in}}$ reasoning proves $+\Delta a$, $+\Delta b$, $+\Delta c$, and $-\Delta \sim b$. Correspondingly, the combined theory justifies the strict arguments $JArgs_{T_{\text{sp}} \ni T_{\text{in}}} = \{\to a \to b; \to c\}$. Due to the priority of T_{sp} over T_{in}, the argument $\to c \to \sim b$ is rejected and undercut by $\to a \to b$. Therefore, there is no justified argument for $\sim b$.

The combination of $T_{\text{sp}} \ni T_{\text{in}}$ extends the priority among defeasible theories to that of rules in the combination. Therefore, the set of arguments constructed from the combination inherits the justification property from that of the standard defeasible logic. This is also due to the coherent property of the extended conditions for strictly provable conclusions.

Proposition 22. *In the combination of two independent theories $T_{\text{sp}} \ni T_{\text{in}}$*

1. *$T_{\text{sp}} \ni T_{\text{in}} \vdash +\partial p$ iff arguments for p are justified by $T_{\text{sp}} \ni T_{\text{in}}$.*
2. *$T_{\text{sp}} \ni T_{\text{in}} \vdash -\partial p$ iff arguments for p are rejected.*

Due to coherent and consistent properties of the extended defeasible reasoning, the set of arguments constructed from the combination of two independent theories satisfies Proposition 23. The extended reasoning over the combination does not simultaneously provide proof for $+\partial p$ and $-\partial p$, or $+\Delta p$ and $-\Delta p$. As a result, it is not possible to construct the arguments both for and against a literal and its complement.

Proposition 23. *In the integration of two defeasible theories $T_{\text{sp}} \ni T_{\text{in}}$*

1. *No argument is both justified and rejected.*
2. *No literal is both justified and rejected.*

3.3 Knowledge Representation

Agent A_{me} has three types of knowledge including the background knowledge T_{bg}, its own knowledge about the working environment T_{me}, and the knowledge about others:

$$\mathscr{T}_{\text{other}} = \{T_j : 1 \le j \le |\mathscr{A}| \text{ and } j \ne \text{me}\}$$

where T_j is obtained from agent A_j during iterations and T_j is represented in DL. At iteration i, the knowledge obtained from A_j is accumulated from previous steps:

$$T_j^i = \bigcup_{k=0}^{i-1} T_j^k + Arg_i^{\mathrm{A_j}}$$

In our framework, the knowledge of an agent can be rebutted by other agents. It is reasonable to assume that defeasible theories contain only defeasible rules and defeasible facts (defeasible rules with empty body).

3.3.1 Knowledge Integration

To generate arguments, an agent integrates knowledge from different sources. Given ambiguous information between two sources, there are two possible methods to combine them: ambiguity blocking and ambiguity propagation.

3.3.1.1 Ambiguity Blocking Integration.

This method extends the standard defeasible reasoning by creating a new superiority relation from that of the knowledge sources, i.e., given two sources as T_{sp}, the superior theory, and T_{in}, the inferior theory, we generate a new superiority relation $R_{\mathrm{d}}^{\mathrm{sp}} > R_{\mathrm{d}}^{\mathrm{in}}$ based on rules from two sources. The integration of the two sources is denoted as $T_{\mathrm{INT}} = T_{\mathrm{sp}} \ni T_{\mathrm{in}}$ (see Section 3.2). Now, the standard defeasible reasoning can be applied for T_{INT} to produce a set of arguments $Args_{\mathrm{AB}}^{\mathrm{T_{INT}}}$.

Example 5. Given two defeasible theories

$$T_{\mathrm{bg}} = \{R_{\mathrm{d}} = \{r_1 : e \Rightarrow c;\ r_2 : g, f \Rightarrow {\sim}c;\ r_3 :\Rightarrow e\}; >= \{r_2 > r_1\}\}$$
$$T_{\mathrm{me}} = \{R_{\mathrm{d}} = \{r_1 :\Rightarrow d;\ r_2 : d \Rightarrow {\sim}a;\ r_3 :\Rightarrow g\}\}$$

The integration of $T_{\mathrm{bg}} \ni T_{\mathrm{me}}$ produces:

$$\begin{aligned} T_{\mathrm{INT}} = \{R_{\mathrm{d}} =&\{r_1^{T_{\mathrm{bg}}} : e \Rightarrow c;\ r_2^{T_{\mathrm{bg}}} : g, f \Rightarrow {\sim}c;\ r_3^{T_{\mathrm{bg}}} :\Rightarrow e; \\ & r_1^{T_{\mathrm{me}}} :\Rightarrow d;\ r_2^{T_{\mathrm{me}}} : d \Rightarrow a;\ r_3^{T_{\mathrm{me}}} :\Rightarrow g\}; \\ >=&\{r_2^{T_{\mathrm{bg}}} > r_1^{T_{\mathrm{bg}}}\}\} \end{aligned}$$

The integrated theory inherits the superiority relation from T_{bg}. That means the new theory reuses the blocking mechanism from T_{bg}.

3.3.1.2 Ambiguity Propagation Integration.

Given two knowledge sources T_1 and T_2, the reasoning mechanism with ambiguity propagation can directly apply to the combination of theories denoted as $T_{\mathrm{INT}}^{'} = T_1 + T_2$. The preference between two sources is unknown, therefore, there is no method to solve conflicts between them. The supportive and opposing arguments for any premise are removed from the final set of arguments. The set of arguments obtained by this integration is denoted by $Args_{\mathrm{AP}}^{\mathrm{T_{INT}^{'}}}$.

3.4 Argument Justification

The motivation of an agent to participate in the game is to promote its own goal. However, its claim can be refuted by different agents. To gain the acceptance of the group, at the first iteration, an agent should justify its arguments by common constraints and expectations of the group governed by the background knowledge T_{bg}. The set of arguments justified by T_{bg} determines arguments that an agent can play to defend its claim. In subsequent iterations, even if the proposal does not conflict with other agents, an agent should ponder knowledge from others to determine the validity of its claim. That is an agent is required a justification by collecting individual arguments from others.

3.4.1 Justification by Background Knowledge

Agent A_{me} generates the set of arguments for its goals by combining its private knowledge T_{me} and the background knowledge T_{bg}. The combination is denoted as $T_{\text{INT}} = T_{\text{bg}} \Cup T_{\text{me}}$ (see Section 3.2) and the set of arguments is $Args^{\text{T}_{\text{INT}}}$. Due to the non-monotonic nature of DL, the combination can produce arguments beyond individual knowledges. From A_{me}'s view, this can bring more opportunities to fulfil its goals. However, A_{me}'s arguments must be justified by the background knowledge T_{bg} since T_{bg} governs essential behaviours (expectations) of the group. Any attack to T_{bg} is not supported by members of $\mathscr{A}$. A_{me} maintains the consistency with the background knowledge T_{bg} by following procedure:

1. Create $T_{\text{INT}} = T_{\text{bg}} \Cup T_{\text{me}}$. The new defeasible theory is obtained by replicating all rules from common constraints T_{bg} into the internal knowledge T_{me} while maintaining the superiority of rules in T_{bg} over that in T_{me}.
2. Use the ambiguity blocking feature to construct the set of arguments $Args^{\text{T}_{\text{bg}}}$ from T_{bg} and the set of arguments $Args^{\text{T}_{\text{INT}}}_{\text{AB}}$ from T_{INT}.
3. Remove any argument in $Args^{\text{T}_{\text{INT}}}_{\text{AB}}$ attacked by those in $Args^{\text{T}_{\text{bg}}}$, obtaining the justified arguments by the background knowledge $JArgs^{\text{T}_{\text{INT}}} = \{a \in Args^{\text{T}_{\text{INT}}}_{\text{AB}}$ and a is accepted by $Args^{\text{T}_{\text{bg}}}\}$.

Example 6. Consider two defeasible theories:

$$T_{\text{bg}} = \{R_{\text{d}} = \{r_1 : e \Rightarrow c;\ r_2 : g, f \Rightarrow \sim c;\ r_3 :\Rightarrow e\}; >= \{r_2 > r_1\}\}$$
$$T_{\text{me}} = \{R_{\text{d}} = \{r_1 :\Rightarrow d;\ r_2 : d \Rightarrow \sim a;\ r_3 :\Rightarrow g\}\}$$

We have sets of arguments from the background theory and the integrated theory:

$$Args^{\text{T}_{\text{bg}}} = \{\Rightarrow e;\ \Rightarrow e \Rightarrow c\}$$
$$Args^{\text{T}_{\text{INT}}} = Args^{\text{T}_{\text{bg}} \Cup \text{T}_{\text{me}}} = \{\Rightarrow e;\ \Rightarrow e \Rightarrow c;\ \Rightarrow d;\ \Rightarrow g;\ \Rightarrow d \Rightarrow \sim a\}$$

In this example, there is not any attack between arguments in $Args^{\text{T}_{\text{bg}}}$ and $Args^{\text{T}_{\text{INT}}}_{\text{AB}}$. In other words, arguments from $Args^{\text{T}_{\text{INT}}}$ are acceptable by those from $Args^{\text{T}_{\text{bg}}}$. The set of justified arguments w.r.t. $Args^{\text{T}_{\text{bg}}}$ is $JArgs^{\text{T}_{\text{INT}}} = Args^{\text{T}_{\text{INT}}}_{\text{AB}}$.

3.4.2 Collective Justification

During the game, A_{me} can exploit the knowledge exposed by other agents in order to defend its main claims. Due to possible conflicts in individual proposals, an agent uses the sceptical semantics of the ambiguity propagation reasoning to retrieve the consistent knowledge. Essentially, given competing arguments an agent does not have any preference over them, therefore, these arguments will be rejected. The consistent knowledge from the others allows an agent to discover "collective wisdom" distributed among agents in order to justify its claim.

The justification of collective arguments, which are generated by integrating all knowledge sources, is done by the arguments from the background knowledge $Args^{T_{\text{bg}}}$. The procedure runs as follows:

1. Create a new defeasible theory $T^{'}_{\text{INT}} = (T_{\text{bg}} \ni T_{\text{me}}) + \mathscr{T}_{\text{other}}$.
2. Generate the set of arguments $Args_{\text{AP}}^{T^{'}_{\text{INT}}}$ from $T^{'}_{\text{INT}}$ using the feature of ambiguity propagation.
3. Justify the new set of arguments $JArgs^{T^{'}_{\text{INT}}} = \{a | a \in Args_{\text{AP}}^{T^{'}_{\text{INT}}}$ and a is accepted by $Args^{T_{\text{bg}}}\}$.

$JArgs^{T^{'}_{\text{INT}}}$ allows A_{me} to verify the status of its arguments for its claim $JArgs^{T_{\text{INT}}}$. If arguments in $JArgs^{T^{'}_{\text{INT}}}$ and $JArgs^{T_{\text{INT}}}$ do not attack one another, A_{me}'s claims are accepted by other agents. Any conflict between two sets shows that accepting arguments in $JArgs^{T^{'}_{\text{INT}}}$ stops A_{me} to achieve its claims in next steps. The set of arguments $Args^{\text{opp}}$ against A_{me} is identified as any argument in $JArgs^{T^{'}_{\text{INT}}}$ attacking A_{me}'s arguments. A_{me} also establishes P^{opp} to select its counter-argument. It is noticed that A_{me} is self-consistent.

Example 7. Suppose the background knowledge T_{bg} and the private knowledge T_{me} of A_{me} are:

$$T_{\text{bg}} = \{R_{\text{d}} = \{r_1 : e \Rightarrow c;\ r_2 : g, f \Rightarrow \sim c\}; >= \{r_2 > r_1\}\}$$
$$T_{\text{me}} = \{R_{\text{d}} = \{r_1 : \Rightarrow e;\ r_2 : c \Rightarrow d;\ r_3 : \Rightarrow g\}\}$$

Agent A_{me} currently plays $\{\Rightarrow e \Rightarrow c \Rightarrow d\}$ and knows about other agents:

$$\mathscr{T}_{\text{other}} = \{T_1, T_2\} \text{ where } T_1 = \{\Rightarrow h \Rightarrow f \Rightarrow b \Rightarrow a\} \text{ and } T_2 = \{\Rightarrow e \Rightarrow c \Rightarrow a\}$$

The claim of A_3 is acceptable w.r.t. arguments played by the other agents. However, the combination $T^{'}_{\text{INT}} = T_{\text{bg}} \ni T_{\text{me}} + \mathscr{T}_{\text{other}}$ shows the difference. This combination generates $\{\Rightarrow g;\ \Rightarrow e;\ \Rightarrow e \Rightarrow f \Rightarrow b;\ \Rightarrow g, f \Rightarrow \sim c\}$. $\{\Rightarrow g, f \Rightarrow \sim c\}$ is due to the superiority relation in T_{bg} which rebuts the claim of A_3. Therefore, the set of opposing arguments $Args^{\text{opp}} = \{\Rightarrow g, f \Rightarrow \sim c\}$ and $P^{\text{opp}} = \{f^1\}$. Given this information, A_3 should provide a counter-evidence to f in order to pursue c. Moreover, A_3 should not expose g to the other agents. Otherwise, A_3 has to drop its initial claim d.

4 Related Works

Substantial works have been done on argumentation games in the artificial intelligence and law-field. [Prakken and Sartor 1996] introduces a dialectical model of legal argument, where arguments can be attacked with appropriate counterarguments. In the model, the factual premises are not arguable; they are treated as strict rules. An early specification and implementation of an argumentation game based on the Toulmin argument-schema without a specified underlying logic is given in [Bench-Capon 1998]. [Lodder 2000] presented the pleadings game as a normative formalization and fully implemented computational model, using conditional entailment.

Settling on a common goal among agents can be seen as a negotiation process where agents exchange information to resolve conflicts or to obtain missing information.

The work in [Amgoud et al. 2007] provides a unified and general formal framework for the argumentation-based negotiation dialogue between two agents. The work establishes a formal connection between the status of a argument (accepted, rejected, and undecided) with an agent's actions (accept, reject, and negotiate respectively). Moreover, an agent's knowledge is evolved by accumulating arguments during interactions.

[Parsons and McBurney 2003] presents an argumentation-based coordination, where agents can exchange arguments for their goals and plans to achieve the goals. The acceptance of an argument of an agent depends on the attitudes of this agent namely credulous, cautious, and sceptical. In [Rueda et al. 2002], agents collaborate with one another by exchanging their proposals and counter-proposals in order to reach a mutual agreement. During conversations, an agent can retrieve missing literals (regarded as sub-goals) or fulfil its goals by requesting collaboration from other agents.

We have advantages of using DL since it flawlessly captures the statuses of arguments, such as accepted, rejected, and undecided by the proof conditions of DL. The statuses are derived from the notions of $+\partial, -\partial$ and $+\Sigma$ corresponding to a positive proof, a negative proof, and a positive support of a premise. Consequently, an agent can take a suitable action either to provide more evidence or to accept an argument from others. In addition, DL provides a compact representation to accommodate new information.

Recenly DL has been proposed to model various argumentation game concepts. [Letia and Vartic 2006] focuses on persuasive dialogues for cooperative interactions among agents. It includes in the process cognitive states of agents such as knowledge and beliefs, and presents some protocols for some types of dialogues (e.g. information seeking, explanation, persuasion). [Hamfelt et al. 2005] provides an extension of DL to include the step of the adversarial dialogue by defining a meta-program for an alternative computational algorithm for ambiguity propagating DL while the logic presented here is ambiguity blocking. In [Thakur et al. 2007], arguments are generated by using the defeasible reasoning with ambiguity blocking. After each step in an argumentation game, an agent can upgrade the strength of its arguments if these arguments are not refuted by the opposing agent.

We tackle the problem of evolving knowledge of an agent during iterations, where the argument construction is an extension of [Thakur et al. 2007]. In our work, we define the notion of collective acceptance for an argument and a method to weight arguments defending against opposing arguments by using both features of ambiguity blocking and propagating.

The works in literature did not clearly show how an agent can tackle with conflicts from multiple agents, especially when the preference over arguments is unknown. The main difference in our framework is the external model where more than two agents can argue to settle on goals commonly accepted by the group. Our weighting mechanism enables an agent to build up a preference over premises constituting opposing arguments from other agents. As a result, an agent can effectively select an argument among those justified by the group's background knowledge to challenge other agents.

We also propose the notion of collective justification to tackle the side-effect of accepting claims from individual agents. Individual arguments for these claims may not conflict with one another, but the integration of these arguments can result in conflicting with an agent's claim. This notion is efficiently deployed in our work due to the efficiency of defeasible logic in handling ambiguous information.

5 Conclusions

We presented an n-person argumentation game based on the extension of defeasible logic, which enables a group of more than two agents to settle on goals commonly accepted by the group. During an argumentation game, using the extended defeasible reasoning each agent can use knowledge from multiple sources including the group's constraints and expectations, other agents' knowledge, and its own knowledge in order to argue to convince other agents about its goals. The knowledge about the group's constraints and expectations plays a critical role in our framework since this knowledge provides a basis to justify new arguments non-monotonically inferred from the integration of different sources.

Acknowledgements

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

References

[Amgoud et al. 2007] Amgoud, L., Dimopoulos, Y., Moraitis, P.: "A unified and general framework for argumentation-based negotiation"; Proceedings AAMAS 2007; 1–8; 2007.

[Antoniou et al. 2000] Antoniou, G., Billington, D., Governatori, G., Maher, M. J.: "A flexible framework for defeasible logics"; Proc. American National Conference on Artificial Intelligence (AAAI-2000); 401–405; 2000.

[Antoniou et al. 2001] Antoniou, G., Billington, D., Governatori, G., Maher, M. J.: "Representation results for defeasible logic"; ACM Transactions on Computational Logic; 2 (2001), 2, 255–287.

[Antoniou et al. 2006] Antoniou, G., Billington, D., Governatori, G., Maher, M. J.: "Embedding defeasible logic into logic programming"; Theory and Practice of Logic Programming; 6 (2006), 6, 703–735.

[Bench-Capon 1998] Bench-Capon, T. J.: "Specification and implementation of Toulmin dialogue game"; A. K. C. d. V. M. C. G. J.C. Hage, T.J.M. Bench-Capon, ed., Jurix 1998; 5–20; 1998.

[Billington 1993] Billington, D.: "Defeasible logic is stable"; Journal of Logic and Computation; 3 (1993), 370–400.

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

[Governatori et al. 2004] Governatori, G., Maher, M. J., Antoniou, G., Billington, D.: "Argumentation semantics for defeasible logic"; J. Logic Computation; 14 (2004), 5, 675–702.

[Hamfelt et al. 2005] Hamfelt, A., Eriksson Lundström, J., Nilsson, J. F.: "A metalogic formalization of legal argumentation as game trees with defeasible reasoning"; ICAIL '05: Proceedings of the 10th international conference on Artificial intelligence and law; 250–251; ACM, 2005.

[Jennings et al. 1998] Jennings, N. R., Parsons, S., Noriega, P., Sierra, C.: "On argumentation-based negotiation"; Proceedings of the International Workshop on Multi-Agent Systems; 1998.

[Letia and Vartic 2006] Letia, I. A., Vartic, R.: "Defeasible protocols in persuasion dialogues"; WI-IATW '06: Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology; 359–362; IEEE Computer Society, Washington, DC, USA, 2006.

[Lodder 2000] Lodder, A. R.: "Thomas F. Gordon, The Pleadings Game – an artificial intelligence model of procedural justice"; Artif. Intell. Law; 8 (2000), 2/3, 255–264.

[Maher 2001] Maher, M. J.: "Propositional defeasible logic has linear complexity"; Theory and Practice of Logic Programming; 1 (2001), 6, 691–711.

[Maher et al. 2001] Maher, M. J., Rock, A., Antoniou, G., Billington, D., Miller, T.: "Efficient defeasible reasoning systems"; International Journal of Artificial Intelligence Tools; 10 (2001), 4, 483–501.

[Parsons and McBurney 2003] Parsons, S., McBurney, P.: "Argumentation-based dialogues for agent coordination"; Group Decision and Negotiation; (2003), 12, 415–439.

[Prakken and Sartor 1996] Prakken, H., Sartor, G.: "A dialectical model of assessing conflicting arguments in legal reasoning"; Artif. Intell. Law; 4 (1996), 3-4, 331–368.

[Rueda et al. 2002] Rueda, S. V., Garcia, A. J., Simari, G. R.: "Argument-based negotiation among bdi agents"; Journal of Computer Science and Technology; 2 (2002), 7, 1–8.

[Thakur et al. 2007] Thakur, S., Governatori, G., Padmanabhan, V., Eriksson Lundström, J.: "Dialogue games in defeasible logic"; 20th Australian Joint Conference on Artificial Intelligence, AI 2007; volume 4830 of LNAI; 497–506; Springer, 2007.