

# Conflict-free normative agents using assumption-based argumentation

**Dorian Gaertner** and **Francesca Toni**

Imperial College London, UK  
{dg00, ft}@doc.ic.ac.uk

May 11, 2007

# Motivation

- 1 formalised agent models (e.g. BDI agents, Elec. Institutions)
  - 2 conflict detection and resolution (e.g. Prakken and Sartor)
- combine the two and argue internally about **conflicts**:
    - between norms
    - between agent-internal knowledge
    - between norms and agent-internal knowledge

# Motivation

- 1 formalised agent models (e.g. BDI agents, Elec. Institutions)
  - 2 conflict detection and resolution (e.g. Prakken and Sartor)
- combine the two and argue internally about **conflicts**:
    - between norms
    - between agent-internal knowledge
    - between norms and agent-internal knowledge

## Conflicting norms

- *You should dance with your mother-in-law !*
- *You should not dance with someone of the same gender !*

## Conflicting internal knowledge

- *I desire to dance with Mary.*
- *I do not desire to dance with drunken people and Mary is drunk.*

## Conflict between a norm and internal knowledge

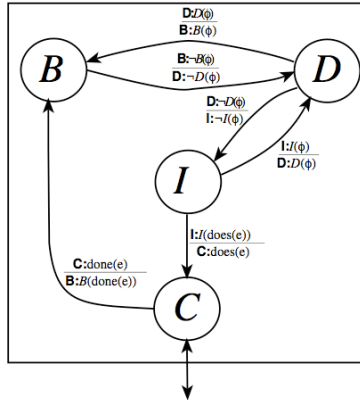
- *You should not dance with someone of the same gender !*
- *I want to dance with attractive people and I consider Bob to be attractive.*

- 1 Normative BDI Agents
  - Theory of MCS
  - Architecture
  - Norm Language
- 2 Assumption-based Argumentation
  - Background Theory
  - CaSAPI System
- 3 Mapping
  - Naive Translation
  - First improvement
- 4 Qualitative Preferences
  - Total Ordering
  - Partial Ordering
  - Dynamic Preferences

# Multi-context systems

- allow to structure knowledge into contexts and model relationships between these context
- each context is characterised by
  - a language  $L_i$
  - axioms  $A_i$
  - inference rules  $\Delta_i$
- bridge rules  $BR$  relate formulae in different contexts

# Example



MCS representation of a *realistic* agent by Parsons et al.

# BDI+N agent architecture

- uses set of context indices  $\mathcal{I} = \{\mathbf{B}, \mathbf{D}, \mathbf{I}\}$  to represent beliefs, desires and intentions
- $Agent = \langle \mathcal{I}, \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle, \mathcal{I} \rightarrow T_i, BR \rangle$
- internalised norms are represented as bridge rules

# Norm language

$$\begin{aligned}
 \textit{InternalNorm} & ::= \varphi \Rightarrow \psi \\
 \varphi & ::= \textit{ConjLiterals} \\
 \textit{ConjLiterals} & ::= \textit{MLiteral} \mid \textit{MLiteral} \wedge \textit{ConjLiterals} \\
 \psi & ::= \textit{MLiteral} \\
 \textit{MLiteral} & ::= \textit{MentalAtom} \mid \neg \textit{MentalAtom} \\
 \textit{MentalAtom} & ::= \text{B}(\textit{term}) \mid \text{D}(\textit{term}) \mid \text{I}(\textit{term})
 \end{aligned}$$

# Norm language

$$\begin{aligned}
 \textit{InternalNorm} & ::= \varphi \Rightarrow \psi \\
 \varphi & ::= \textit{ConjLiterals} \\
 \textit{ConjLiterals} & ::= \textit{MLiteral} \mid \textit{MLiteral} \wedge \textit{ConjLiterals} \\
 \psi & ::= \textit{MLiteral} \\
 \textit{MLiteral} & ::= \textit{MentalAtom} \mid \neg \textit{MentalAtom} \\
 \textit{MentalAtom} & ::= \text{B}(\textit{stateterm}) \\
 & \quad \mid \text{B}(\textit{Eitherterm} \rightarrow \textit{Eitherterm}) \\
 & \quad \mid \text{D}(\textit{Eitherterm}) \mid \text{I}(\textit{actionterm}) \\
 \textit{Eitherterm} & ::= \textit{actionterm} \mid \textit{stateterm}
 \end{aligned}$$

## Example bridge rules

- general agent bridge rule:

$$\frac{\neg D : (X)}{\neg I : (X)}$$

- internalised norm:

$$\frac{B : (\text{married}(\text{self}, W)) \wedge B : (\text{mother}(X, W))}{I : (\text{danceWith}(X))}$$

# Formal definition

## Definition

An assumption-based framework is a tuple  $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \bar{\cdot} \rangle$  where

- $(\mathcal{L}, \mathcal{R})$  is a deductive system
- $\mathcal{A} \subseteq \mathcal{L}$  is the set of candidate *assumptions*
- if  $c \in \mathcal{A}$ , then there exists no inference rule of the form  
$$c \leftarrow c_1, \dots, c_n \in \mathcal{R}$$
- $\bar{\cdot}$  is a (total) mapping from  $\mathcal{A}$  into  $\mathcal{L}$ , where  $\bar{\alpha}$  is the *contrary* of  $\alpha$

# CaSAPI

- system that determines acceptability and support of claims
- using assumption-based argumentation frameworks
- according to three semantics
- providing structured arguments and their inter-relationships

## Example of CaSAPI input

Let  $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \bar{\cdot} \rangle$  be the ABA framework:

- $\mathcal{L} = \{p, q, r, s, t, \neg p, \neg q, \neg r, \neg s, \neg t\}$
- $\mathcal{R}$  consists of

$$p \leftarrow q, r$$

$$q \leftarrow s$$

$$\neg r \leftarrow t$$

$$\neg t \leftarrow$$

- $\mathcal{A} = \{r, s, t\}$
- $\bar{r} = \neg r, \bar{s} = \neg s, \bar{t} = \neg t$
- $\{r, s\} \vdash p$

## Example of CaSAPI input

Let  $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \bar{\cdot} \rangle$  be the ABA framework:

- $\mathcal{L} = \{p, q, r, s, t, \neg p, \neg q, \neg r, \neg s, \neg t\}$

- $\mathcal{R}$  consists of

$$p \leftarrow q, r$$

$$q \leftarrow s$$

$$\neg r \leftarrow t$$

$$\neg t \leftarrow$$

- $\mathcal{A} = \{r, s, t\}$

- $\bar{r} = \neg r, \bar{s} = \neg s, \bar{t} = \neg t$

- $\{r, s\} \vdash p$

## Example of CaSAPI input

Let  $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \bar{\ } \rangle$  be the ABA framework:

- $\mathcal{L} = \{p, q, r, s, t, \neg p, \neg q, \neg r, \neg s, \neg t\}$

- $\mathcal{R}$  consists of

$$p \leftarrow q, r$$

$$q \leftarrow s$$

$$\neg r \leftarrow t$$

$$\neg t \leftarrow$$

- $\mathcal{A} = \{r, s, t\}$

- $\bar{r} = \neg r, \bar{s} = \neg s, \bar{t} = \neg t$

- $\{r, s\} \vdash p$

```

myRule(p, [q,r]).
myRule(q, [s]).
myRule(not(r), [t]).
myRule(not(t), []).

myAsm([r,s,t]).

toBeProved([p]).

contrary(r,not(r)).
contrary(s,not(s)).
contrary(t,not(t)).
    
```

## Example of CaSAPI output

CASE 1ii

Step 8:

Content of this element:

PropNodes: []

OppoNodes: []

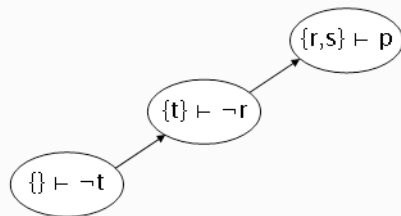
DefnceAss: [r,s]

Culprits : [t]

Arguments: [argument(1,[r,s],p),argument(2,t,not(r)),argument(4,[],not(t))]

Relations: [attacks(1,nothing),attacks(2,1),attacks(4,2)]

FINISHED, one defence set is: [r,s]



# Naive mapping

- assumption-based arg. inference rules (  $\leftarrow$  ) represent:
  - inference rules at context level (  $\text{---}$  )
  - bridge rules (  $\text{---}$  )
  - internalised norms (  $\Rightarrow$  )
- assumption-based arg. facts (  $\leftarrow$  ) represent:
  - axioms (e.g.  $\neg B(\perp)$ )
  - facts from initial theories (e.g.  $B(\text{attractive}(\text{bob}))$ )

## Problems due to naivety

- imagine two rules:
  - 1  $I(\text{danceWith}(\text{bob})) \leftarrow B(\text{attractive}(\text{bob}))$
  - 2  $\neg I(\text{danceWith}(\text{bob})) \leftarrow B(\text{sameSex}(\text{bob}, \text{self}))$
- querying CaSAPI with following input succeeds:  
 $I(\text{danceWith}(\text{bob})) \wedge \neg I(\text{danceWith}(\text{bob}))$
- this is clearly **not desirable**

# Mutual exclusivity

- add assumptions to make rules mutually exclusive:

①  $I(\text{danceWith}(\text{bob})) \leftarrow B(\text{attractive}(\text{bob})), \alpha$

②  $\neg I(\text{danceWith}(\text{bob})) \leftarrow B(\text{sameSex}(\text{bob}, \text{self})), \beta$

- define appropriate contraries:

①  $\bar{\alpha} = \neg I(\text{danceWith}(\text{bob}))$

②  $\bar{\beta} = I(\text{danceWith}(\text{bob}))$

- this **avoids conflicts** but does not resolve them

## Preliminaries

- add to agent definition a preference function mapping bridge rules, axioms, inference rules and theory elements to  $\mathbb{Q}$
- cluster rules according to the literal in their head
- order rules within a cluster by preference

$$l_1 \leftarrow r_1 \quad l_2 \leftarrow r_2 \quad l_3 \leftarrow r_3 \quad \dots \quad l_n \leftarrow r_n$$

## Use of preferences

- make rules defeasible by adding assumptions  $p_i$ :

$$l_1 \leftarrow r_1 \quad l_2 \leftarrow r_2, p_2 \quad l_3 \leftarrow r_3, p_3 \quad \dots \quad l_n \leftarrow r_n, p_n$$

- new sentences  $q_i$  are added:

$$\begin{array}{ccccccc} q_2 \leftarrow r_1 & q_3 \leftarrow r_2, p_2 & q_4 \leftarrow r_3, p_3 & \dots & q_n \leftarrow r_{n-1}, p_{n-1} & & \\ & q_3 \leftarrow q_2 & q_4 \leftarrow q_3 & \dots & q_n \leftarrow q_{n-1} & & \\ & & q_4 \leftarrow q_2 & \dots & q_n \leftarrow q_{n-2} & & \\ & & & \dots & \dots & & \\ & & & & & & q_n \leftarrow q_2 \end{array}$$

- contraries are defined:  $\overline{p_i} = q_i$

## Example

- from the ballroom example:
  - 1  $\neg I(\text{danceWith}(\text{bob})) \leftarrow B(\text{sameSex}(\text{bob}, \text{self}))$
  - 2  $I(\text{danceWith}(\text{bob})) \leftarrow B(\text{attractive}(\text{bob}))$
  - 3  $\neg I(\text{danceWith}(\text{bob})) \leftarrow B(\text{drunk}(\text{bob}))$
  - 4
  - 5
  - 6

•

## Example

- from the ballroom example:
  - 1  $\neg I(\text{danceWith}(\text{bob})) \leftarrow B(\text{sameSex}(\text{bob}, \text{self}))$
  - 2  $I(\text{danceWith}(\text{bob})) \leftarrow B(\text{attractive}(\text{bob})), p_2$
  - 3  $\neg I(\text{danceWith}(\text{bob})) \leftarrow B(\text{drunk}(\text{bob})), p_3$
  - 4  $q_2 \leftarrow B(\text{sameSex}(\text{bob}, \text{self}))$
  - 5  $q_3 \leftarrow B(\text{attractive}(\text{bob})), p_2$
  - 6  $q_3 \leftarrow q_2$
- contrary  $\overline{p_i} = q_i$

# Partial ordering

- replace preference function with binary preference relation  $pref$
- assume that  $pref(\pi_1, \pi_2)$  is **asymmetric** and **irreflexive**
- when rules  $\pi_1$  and  $\pi_2$  are in same cluster but neither  $pref(\pi_1, \pi_2)$  nor  $pref(\pi_2, \pi_1)$  hold, use conflict avoidance

# Partial ordering

- each rule  $\pi$  of the form  $l \leftarrow r$  is replaced with:

$l \leftarrow r, p$

$q \leftarrow r_1, \text{pref}(\pi_1, \pi)$

...

$q \leftarrow r_n, \text{pref}(\pi_n, \pi)$

- where each  $\pi_i$  is a rule with the complement of  $l$  as its head
- set  $\bar{p} = q$

## Allowing dynamic preferences

- replace facts of relation  $pref$  with meta-rules of the form:

$$pref(\pi_1, \pi_2) \leftarrow conditions$$

- more fine-grained control over arguments

## Example of dynamic preferences

- imagine two rules:
  - $\pi_1$  states that if one desires to dance, one should approach a partner
  - $\pi_2$  states that women should wait to be approached
- together with two meta-rules:
  - $pref(\pi_2, \pi_1) \leftarrow normal\_male\_choice$
  - $pref(\pi_1, \pi_2) \leftarrow ladies\_choice\_song$

# Conclusions

We have:

- presented an **agent architecture** based on MCS:  
 $Agent = \langle \mathcal{I}, \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle, \mathcal{I} \rightarrow T_i, BR, \mathcal{P} \rangle$
- with norms internalised as relations between mental attitudes
- suggested to use **argumentation** to solve (normative) conflicts
- proposed 3 preference-based **conflict resolution** mechanisms

## Current and future work

- apply assumption-based argumentation to other logic-based agent architectures (e.g. 3APL, KGP or Jason)
- adding automatic translation to CaSAPI
- apply assumption-based argumentation to solve conflicts between agents (with common notion of preference)

Thank you

Questions ?