Picking Up the Best Goal
An Analytical Study in Defeasible Logic

Guido Governatori, Francesco Olivieri, Antonino Rotolo\textsuperscript{2}, Simone Scannapieco and Matteo Cristani\textsuperscript{1}

Department of Computer Science, University of Verona, Italy
CIRSFID and DSG, University of Bologna, Italy

RuleML 2013, 13 July 2013
Motivation: BDI Agents

BDI is a popular architecture to model autonomous agents:

- **B** Beliefs: How the agents perceives the environment
- **D** Desires: What the agent wants to achieve
- **I** Intentions: What the agent commits to
Motivation: BDI Agents

BDI is a popular architecture to model autonomous agents:

B Beliefs: How the agents perceives the environment
D Desires: What the agent wants to achieve
I Intentions: What the agent commits to

Beliefs, Intensions and Desires are called mental attitudes;
Motivation: BDI Agents

BDI is a popular architecture to model autonomous agents:

- **B** Beliefs: How the agents perceives the environment
- **D** Desires: What the agent wants to achieve
- **I** Intentions: What the agent commits to

Beliefs, Intentions and Desires are called **mental attitudes**; Intentions and Desires are also **motivational attitudes**.
Issues with BDI agents

- Is there anything missing in the BDI architecture?
Issues with BDI agents

- Is there anything missing in the BDI architecture?
- Is there something redundant in the BDI architecture?
BIOlogical Agents

Why BIOlogical agents? self-evident
BIO Logical Agents

Why BIO Logical agents?

B Beliefs: the description of the environment
BIO Logical Agents

Why BIO Logical agents?

- **B** Beliefs: the description of the environment
  - Intentions: the *internal* constraints/motivational attitudes
BIO Logical Agents

Why BIO Logical agents?

B Beliefs: the description of the environment

I Intentions: the internal constraints/motivational attitudes

O Obligations: the external constraints/motivational attitudes
Desires, Goals, Intentions, Social Intentions are nuances of a more general concept: Outcomes (the objectives of an agent)
Design Principles

• An agent is modelled by a set of rules;
• When an agent faces alternative outcomes in a given context, it is natural to rank them in a preference order;
• Beliefs prevail over conflicting motivational attitudes, thus avoiding various cases of wishful thinking;
• Norms and obligations are used to filter social motivational states (social intentions) and compliant agents;
• Goal-like attitudes can be derived via conversion using other mental states, such as beliefs (e.g., believing that Madrid is in Spain may imply that the goal to go to Madrid implies the goal to go to Spain).
Rule Types

- Belief rules
  \[ a_1, \ldots, a_n \Rightarrow c \]

- Obligation rules
  \[ a_1, \ldots, a_n \Rightarrow_O C \]

- Outcome rules
  \[ a_1, \ldots, a_n \Rightarrow_U C \]
Rule Types

- Belief rules
  \[ a_1, \ldots, a_n \Rightarrow c \]

- Obligation rules
  \[ a_1, \ldots, a_n \Rightarrow_{O} C \]

- Outcome rules
  \[ a_1, \ldots, a_n \Rightarrow_{U} C \]

\[ C = c_1 \circ c_2 \circ \cdots \circ c_n \]
Example

\[\text{holiday} \Rightarrow_U \text{visit\_friend} \circ \text{visit\_parents} \circ \text{stay\_home}\]
Desires as acceptable outcomes

\[
\begin{align*}
  r : a_1, \ldots, a_n &\Rightarrow \bigcirc b_1 \circ \cdots \circ b_m \\
  s : a'_1, \ldots, a'_n &\Rightarrow \bigcirc b'_1 \circ \cdots \circ b'_k
\end{align*}
\]

where \(a_1, \ldots, a_n\) and \(a'_1, \ldots, a'_n\) are mutually compatible. 
\(b_1\) and \(b'_1\) are mutually incompatible (\(b'_1 = \neg b_1\)).
Desires as acceptable outcomes

$r : a_1, \ldots, a_n \Rightarrow b_1 \circ \cdots \circ b_m$

$s : a'_1, \ldots, a'_n \Rightarrow b'_1 \circ \cdots \circ b'_k$

where $a_1, \ldots, a_n$ and $a'_1, \ldots, a'_n$ are mutually compatible

$b_1$ and $b'_1$ are mutually incompatible ($b'_1 = \neg b_1$).

$b_1, \ldots, b_m, b'_1, \ldots, b'_k$ are all desires (acceptable outcomes)
Desires as acceptable outcomes

\[ r : a_1, \ldots, a_n \Rightarrow U b_1 \odot \cdots \odot b_m \]

\[ s : a'_1, \ldots, a'_n \Rightarrow U b'_1 \odot \cdots \odot b'_k \]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible
\( b_1 \) and \( b'_1 \) are mutually incompatible (\( b'_1 = \neg b_1 \)).

\( b_1, \ldots, b_m, b'_1, \ldots, b'_k \) are all desires (acceptable outcomes)

If \( s > r \), then \( b_2, \ldots, b_m, b'_1, \ldots, b'_k \) are desires (acceptable outcomes)
Goals as preferred outcomes

\[ r : a_1, \ldots, a_n \Rightarrow U \ b_1 \circ \cdots \circ b_m \]
\[ s : a'_1, \ldots, a'_n \Rightarrow U \ b'_1 \circ \cdots \circ b'_k \]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible \( b_1 \) and \( b'_1 \) are mutually incompatible \( (b'_1 = \neg b_1) \).
\( s > r \)
Goals as preferred outcomes

\[
\begin{align*}
  r : a_1, \ldots, a_n & \Rightarrow U b_1 \odot \cdots \odot b_m \\
  s : a'_1, \ldots, a'_n & \Rightarrow U b'_1 \odot \cdots \odot b'_k 
\end{align*}
\]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible
\( b_1 \) and \( b'_1 \) are mutually incompatible \((b'_1 = \neg b_1)\).

\( s > r \)

\( b_2 \) and \( \neg b_1 \) are the goals (most preferred outcomes)
Intentions as feasible outcomes

\[ r : a_1, \ldots, a_n \Rightarrow \bigcirc b_1 \bigcirc \cdots \bigcirc b_m \]

and the agent knows \( \neg b_1 \)
Intentions as feasible outcomes

\[ r : a_1, \ldots , a_n \Rightarrow \bigcirc b_1 \circ \cdots \circ b_m \]

and the agent knows \( \neg b_1 \)

\( b_2 \) is the intention (the preferred feasible outcome)
Intentions as feasible outcomes (2)

\[
\begin{align*}
  r & : a_1, \ldots, a_n \Rightarrow \bigcirc b_1 \circ \cdots \circ b_m \\
  s & : a'_1, \ldots, a'_n \Rightarrow \bigcirc b'_1 \circ \cdots \circ b'_k
\end{align*}
\]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible
\( b_1 \) and \( b'_1 \) are mutually incompatible (\( b'_1 = \neg b_1 \)).
\( s > r \)
and the agent knows \( b_1 \)
Intentions as feasible outcomes (2)

\[ r : a_1, \ldots, a_n \Rightarrow \bigcirc b_1 \circ \cdots \circ b_m \]
\[ s : a'_1, \ldots, a'_n \Rightarrow \bigcirc b'_1 \circ \cdots \circ b'_k \]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible
\( b_1 \) and \( b'_1 \) are mutually incompatible \((b'_1 = \neg b_1)\).
\( s > r \)
and the agent knows \( b_1 \)
\( b_1 \) and \( b_2 \) are the intentions (the preferred feasible outcomes)
Intentions as legal outcomes

\[ r : a_1, \ldots, a_n \Rightarrow b_1 \circ \cdots \circ b_m \]
\[ s : a'_1, \ldots, a'_n \Rightarrow b'_1 \circ \cdots \circ b'_k \]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible, \( b_1 \) and \( b'_1 \) are mutually incompatible (\( b'_1 = \neg b_1 \)).
Intentions as legal outcomes

\[ r : a_1, \ldots, a_n \Rightarrow_U b_1 \odot \cdots \odot b_m \]
\[ s : a'_1, \ldots, a'_n \Rightarrow_O b'_1 \odot \cdots \odot b'_k \]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible,
\( b_1 \) and \( b'_1 \) are mutually incompatible (\( b'_1 = \neg b_1 \)).

\( \neg b_1 \) is obligatory and \( b_2 \) is socially intended (most preferred feasible outcome that does not violate the obligations).
Intentions as legal outcomes

\[ r : a_1, \ldots, a_n \Rightarrow U b_1 \odot \cdots \odot b_m \]
\[ s : a'_1, \ldots, a'_n \Rightarrow O b'_1 \odot \cdots \odot b'_k \]

where \( a_1, \ldots, a_n \) and \( a'_1, \ldots, a'_n \) are mutually compatible

\( b_1 \) and \( b'_1 \) are mutually incompatible (\( b'_1 = \neg b_1 \)).

\( \neg b_1 \) is obligatory and \( b_2 \) is socially intended (most preferred feasible outcome that does not violate the obligations)

If the agent knows \( \neg b_2 \), then \( b_3 \) is socially intended
To prove that an agent believes $p$.

There is a belief rule

$$a_1, \ldots, a_n \implies_B p$$

- all $a_i$ are provable
- and all rules for $\neg p$ are either not applicable or weaker than an applicable rule for $p$
To prove that $p$ is obligatory

There is an obligation rule

$$a_1, \ldots, a_n \Rightarrow \bigcirc c_1 \circ \cdots \circ c_m$$

- $p = c_j, 1 \leq j \leq m$
- all $a_i$ are provable
- for all $c_i, i < j$:
  - $c_i$ is obligatory and
  - the agent does not believe $c_i$
- defeasibility
To prove that an agent desires $p$

There is an outcome rule

$$a_1, \ldots, a_n \Rightarrow \bigcirc c_1 \bigcirc \cdots \bigcirc c_m$$

- $p = c_j$, $1 \leq j \leq m$
- all $a_i$ are provable
- defeasibility
Proving Goals

To prove that \( p \) is a goal of the agent

There is an outcome rule

\[ a_1, \ldots, a_n \Rightarrow u \; c_1 \odot \cdots \odot c_m \]

- \( p = c_j, 1 \leq j \leq m \)
- all \( a_i \) are provable
- for all \( c_i, i < j, c_i \) is not a goal of the agent
- defeasibility
To prove that the agent intends $p$

There is an outcome rule

$$a_1, \ldots, a_n \Rightarrow \bigcup_{i=1}^{m} c_i \bigcirc \ldots \bigcirc c_m$$

- $p = c_j$, $1 \leq j \leq m$
- all $a_i$ are provable
- for all $c_i$, $i < j$,
  - $c_i$ is not an intention of the agent
  - the agent does not believe $\neg c_i$
- defeasibility
Proving Social Intentions

To prove that the agent intends $p$

There is an outcome rule

$$a_1, \ldots, a_n \Rightarrow u \ c_1 \odot \cdots \odot c_m$$

- $p = c_j, 1 \leq j \leq m$
- all $a_i$ are provable
- for all $c_i, i < j$,
  - $c_i$ is not an intention of the agent
  - $c_i$ is not forbidden (i.e., $\neg c_i$ is not obligatory)
  - the agent does not belief $\neg c_i$

- defeasibility
Theorem

The logic is coherent, i.e., it is not possible to prove $Xp$ and $\neg Xp$ (and $Xp$ and $X\neg p$) for $X \in \{G, I, SI, O\}$

Theorem

The extension of the logic can be computed in $O(|D|)$, where $D$ is the number of symbols occurring in a theory.
Conclusion

- A novel account of the notions of goals like attitudes for agents
- We have argued that the notions of desires, goals, intentions are facets of a more general concept (i.e., outcome/objective)
- The account can be formalised in Defeasible Logic in a computationally feasible way
Questions?