Norm Compliance of Rule-based Cognitive Agents

Antonino Rotolo

CIRSFID, University of Bologna, Italy

IJCAI 2011
22 July 2011, Barcelona, Spain

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

- Defeasible Logic (DL) for BIO agents (Governatori and Rotolo 2006 and 2008): rule-based cognitive agents

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

- Defeasible Logic (DL) for BIO agents (Governatori and Rotolo 2006 and 2008): rule-based cognitive agents
- Extending BIO-DL with paths

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

- Defeasible Logic (DL) for BIO agents (Governatori and Rotolo 2006 and 2008): rule-based cognitive agents
- Extending BIO-DL with paths
- Using and extending revision-theory techniques for DL (Billington *et al.* 1999) to handle intention reconsideration and compliance in BIO-DL

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

- Defeasible Logic (DL) for BIO agents (Governatori and Rotolo 2006 and 2008): rule-based cognitive agents
- Extending BIO-DL with paths
- Using and extending revision-theory techniques for DL (Billington *et al.* 1999) to handle intention reconsideration and compliance in BIO-DL
  - Removing intention rules remove intentions from theory extensions

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

- Defeasible Logic (DL) for BIO agents (Governatori and Rotolo 2006 and 2008): rule-based cognitive agents
- Extending BIO-DL with paths
- Using and extending revision-theory techniques for DL (Billington *et al.* 1999) to handle intention reconsideration and compliance in BIO-DL
  - Removing intention rules remove intentions from theory extensions
  - Adding new rules to remove intentions from theory extensions

Layout

Motivation: So far research in MAS (cognitive agents) has overlooked intention reconsideration to achieve norm compliance

- Defeasible Logic (DL) for BIO agents (Governatori and Rotolo 2006 and 2008): rule-based cognitive agents
- Extending BIO-DL with paths
- Using and extending revision-theory techniques for DL (Billington *et al.* 1999) to handle intention reconsideration and compliance in BIO-DL
  - Removing intention rules remove intentions from theory extensions
  - Adding new rules to remove intentions from theory extensions
- Properties and future work

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;
- $R^{\mathbf{I}}$ is a finite set of intention rules;

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;
- $R^{\mathbf{I}}$ is a finite set of intention rules;
- all types of rule may have the following formats ($X \in \{\mathbf{B}, \mathbf{I}, \mathbf{O}\}$):

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;
- $R^{\mathbf{I}}$ is a finite set of intention rules;
- all types of rule may have the following formats ($X \in \{\mathbf{B}, \mathbf{I}, \mathbf{O}\}$):
  - $a_1, \ldots, a_n \rightarrow_X b$

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;
- $R^{\mathbf{I}}$ is a finite set of intention rules;
- all types of rule may have the following formats ($X \in \{\mathbf{B}, \mathbf{I}, \mathbf{O}\}$):
  - $a_1, \ldots, a_n \rightarrow_X b$
  - $a_1, \ldots, a_n \Rightarrow_X b$

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;
- $R^{\mathbf{I}}$ is a finite set of intention rules;
- all types of rule may have the following formats ($X \in \{\mathbf{B}, \mathbf{I}, \mathbf{O}\}$):
  - $a_1, \ldots, a_n \rightarrow_X b$
  - $a_1, \ldots, a_n \Rightarrow_X b$
  - $a_1, \ldots, a_n \leadsto_X b$

BIO-DL without paths (Governatori and Rotolo 2008)

An agent theory D is a structure

$$(F, R^{\mathbf{B}}, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$$

- F set of facts (literals, e.g., p, or modal literals, e.g., $\mathbf{O}p$);
- $R^{\mathbf{B}}$ is a finite set of belief rules;
- $R^{\mathbf{O}}$ is a finite set of obligation rules;
- $R^{\mathbf{I}}$ is a finite set of intention rules;
- all types of rule may have the following formats ($X \in \{\mathbf{B}, \mathbf{I}, \mathbf{O}\}$):
  - $a_1, \ldots, a_n \rightarrow_X b$
  - $a_1, \ldots, a_n \Rightarrow_X b$
  - $a_1, \ldots, a_n \leadsto_X b$
- $\succ$ is an acyclic (superiority) relation over $(R^{\mathbf{B}} \times R^{\mathbf{B}}) \cup (R^{\mathbf{I}} \times R^{\mathbf{I}}) \cup (R^{\mathbf{O}} \times R^{\mathbf{O}})$.

How BIO-DL works

- Different rules for each kind of modality; e.g.,

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\mathit{MiddleEarth} \rightarrow_{\mathbf{O}} \mathit{DestroyRing}$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\mathit{MiddleEarth} \rightarrow_{\mathbf{O}} \mathit{DestroyRing}$$
$$\mathit{Mordor} \Rightarrow_{\mathbf{I}} \mathit{BackToShire}$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\begin{array}{l} \mathit{MiddleEarth} \rightarrow_{\mathbf{O}} \mathit{DestroyRing} \\ \mathit{Mordor} \Rightarrow_{\mathbf{I}} \mathit{BackToShire} \\ \mathit{Nazgul} \Rightarrow_{\mathbf{B}} \mathit{Danger} \end{array}$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\begin{aligned}&\mathit{MiddleEarth} \rightarrow_{\mathbf{O}} \mathit{DestroyRing}\\&\mathit{Mordor} \Rightarrow_{\mathbf{I}} \mathit{BackToShire}\\&\mathit{Nazgul} \Rightarrow_{\mathbf{B}} \mathit{Danger}\end{aligned}$$

- As in DL, different types of provability:

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\begin{array}{l} \mathit{MiddleEarth} \rightarrow_{\mathbf{O}} \mathit{DestroyRing} \\ \mathit{Mordor} \Rightarrow_{\mathbf{I}} \mathit{BackToShire} \\ \mathit{Nazgul} \Rightarrow_{\mathbf{B}} \mathit{Danger} \end{array}$$

- As in DL, different types of provability:

$$+\Delta^{\mathbf{O}} \mathit{DestroyRing}$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\begin{array}{l} \mathit{MiddleEarth} \rightarrow_{\mathbf{O}} \mathit{DestroyRing} \\ \mathit{Mordor} \Rightarrow_{\mathbf{I}} \mathit{BackToShire} \\ \mathit{Nazgul} \Rightarrow_{\mathbf{B}} \mathit{Danger} \end{array}$$

- As in DL, different types of provability:

$$+\Delta^{\mathbf{O}} \mathit{DestroyRing} \qquad +\partial^{\mathbf{I}} \mathit{BackToShire}$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\begin{aligned} &\mathit{MiddleEarth} \rightarrow_{\mathsf{O}} \mathit{DestroyRing} \\ &\mathit{Mordor} \Rightarrow_{\mathsf{I}} \mathit{BackToShire} \\ &\mathit{Nazgul} \Rightarrow_{\mathsf{B}} \mathit{Danger} \end{aligned}$$

- As in DL, different types of provability:

$$+\Delta^{\mathsf{O}} \mathit{DestroyRing} \qquad +\partial^{\mathsf{I}} \mathit{BackToShire}$$

$$-\Delta^{\mathsf{O}} \mathit{DestroyRing}$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$MiddleEarth \to_{\mathbf{O}} DestroyRing$$
$$Mordor \Rightarrow_{\mathbf{I}} BackToShire$$
$$Nazgul \Rightarrow_{\mathbf{B}} Danger$$

- As in DL, different types of provability:

$$+\Delta^{\mathbf{O}} DestroyRing \qquad +\partial^{\mathbf{I}} BackToShire$$
$$-\Delta^{\mathbf{O}} DestroyRing \qquad -\partial^{\mathbf{I}} BackToShire$$

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$\begin{aligned} &MiddleEarth \rightarrow_{\mathbf{O}} DestroyRing \\ &Mordor \Rightarrow_{\mathbf{I}} BackToShire \\ &Nazgul \Rightarrow_{\mathbf{B}} Danger \end{aligned}$$

- As in DL, different types of provability:

$$\begin{array}{ll} +\Delta^{\mathbf{O}} DestroyRing & +\partial^{\mathbf{I}} BackToShire \\ -\Delta^{\mathbf{O}} DestroyRing & -\partial^{\mathbf{I}} BackToShire \end{array}$$

- Provability in an agent theory D is used for introducing modalities in the theory extension

How BIO-DL works

- Different rules for each kind of modality; e.g.,

$$MiddleEarth \rightarrow_{\mathbf{O}} DestroyRing$$
$$Mordor \Rightarrow_{\mathbf{I}} BackToShire$$
$$Nazgul \Rightarrow_{\mathbf{B}} Danger$$

- As in DL, different types of provability:

$$+\Delta^{\mathbf{O}} DestroyRing \qquad +\partial^{\mathbf{I}} BackToShire$$
$$-\Delta^{\mathbf{O}} DestroyRing \qquad -\partial^{\mathbf{I}} BackToShire$$

- Provability in an agent theory D is used for introducing modalities in the theory extension

$$\frac{MiddleEarth \Rightarrow_{\mathbf{O}} DestroyRing}{+\partial^{\mathbf{O}} DestroyRing} \qquad \frac{MiddleEarth}{D \mid\sim \mathbf{O} DestroyRing}$$

How BIO-DL works (cont'd)

Facts: *Entrusted*, *Hobbit*

Rules: r_1:$\mathbf{O}$*Mordor* $\Rightarrow_{\mathbf{O}}$ *DestroyRing*
r_2:*RingBearer* $\Rightarrow_{\mathbf{O}}$ *Mordor*
r_3:*RingBearer* $\rightarrow_{\mathbf{I}}$ $\neg$*DestroyRing*
r_4:*Entrusted* $\rightarrow_{\mathbf{B}}$ *RingBearer*
r_5:*Hobbit* $\Rightarrow_{\mathbf{O}}$ $\neg$*Mordor*

Superiority relation:
$r_5 \succ r_2$

How BIO-DL works (cont'd)

Facts: *Entrusted*, *Hobbit*

Rules: r_1:$\mathbf{O}$*Mordor* $\Rightarrow_{\mathbf{O}}$ *DestroyRing*
r_2:*RingBearer* $\Rightarrow_{\mathbf{O}}$ *Mordor*
r_3:*RingBearer* $\rightarrow_{\mathbf{I}}$ $\neg$*DestroyRing*
r_4:*Entrusted* $\rightarrow_{\mathbf{B}}$ *RingBearer*
r_5:*Hobbit* $\Rightarrow_{\mathbf{O}}$ $\neg$*Mordor*

Superiority relation:
$r_5 \succ r_2$

Phase 1: Prove O ***Mordor***
Facts+r_4 + r_2
Phase 2: Attacks
Facts+r_5
Phase 3: Rebut attacks
r_5 weaker than r_2

Phase 1: Prove O ***DestroyRing***
Facts+r_4 + r_1
Phase 2: Attacks
See above: Facts+r_5 (undercut)
Phase 3: Rebut attacks
See above

Phase 1: Prove I$\neg$***DestroyRing***
Facts+r_4 + r_3
Phase 2: Attacks
No argument
Phase 3: Rebut attacks
Not needed

Intention reconsideration and compliance

Suppose we have $\mathbf{O}b$

Intention reconsideration and compliance

Suppose we have $\mathbf{O}b$

- Case 1—unchangeable intentions:

$$\mathbf{I}b \qquad \mathbf{I}\neg b \tag{1}$$

Intention reconsideration and compliance

Suppose we have $\mathbf{O}b$

- Case 1—unchangeable intentions:

$$\mathbf{I}b \qquad \mathbf{I}\neg b \tag{1}$$

- Case 2—strong intentions:

$$r_1 : a \to_{\mathbf{I}} b \qquad r_2 : c \to_{\mathbf{I}} \neg b \tag{2}$$

Intention reconsideration and compliance

Suppose we have $\mathbf{O}b$

- Case 1—unchangeable intentions:

$$\mathbf{I}b \qquad \mathbf{I}\neg b \tag{1}$$

- Case 2—strong intentions:

$$r_1 : a \rightarrow_{\mathbf{I}} b \qquad r_2 : c \rightarrow_{\mathbf{I}} \neg b \tag{2}$$

- Case 3—weak intentions:

$$r_1 : a \Rightarrow_{\mathbf{I}} b \qquad r_2 : c \Rightarrow_{\mathbf{I}} \neg b \tag{3}$$

Intention reconsideration and compliance

Suppose we have $\mathbf{O}b$

- Case 1—unchangeable intentions:

$$\mathbf{I}b \qquad \mathbf{I}\neg b \tag{1}$$

- Case 2—strong intentions:

$$r_1 : a \rightarrow_{\mathbf{I}} b \qquad r_2 : c \rightarrow_{\mathbf{I}} \neg b \tag{2}$$

- Case 3—weak intentions:

$$r_1 : a \Rightarrow_{\mathbf{I}} b \qquad r_2 : c \Rightarrow_{\mathbf{I}} \neg b \qquad r_2 \succ r_1 \tag{3}$$

Recovering from violations: reconsidering strong intentions

Recovering from violations: reconsidering strong intentions

$$F = \{a, \mathbf{I}b\}$$

$$R = \{r_1 : a \rightarrow_{\mathbf{I}} \neg c,$$

$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_3 : \mathbf{I}b \rightarrow_{\mathbf{I}} d,$$

$$r_4 : \mathbf{I}d, a \rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \emptyset$$

Recovering from violations: reconsidering strong intentions

$$F = \{a, \mathbf{I}b\}$$

$$R = \{r_1 : a \rightarrow_{\mathbf{I}} \neg c,$$

$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_3 : \mathbf{I}b \rightarrow_{\mathbf{I}} d,$$

$$r_4 : \mathbf{I}d, a \rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \emptyset$$

Recovering from violations: reconsidering strong intentions

$$F = \{a, \mathbf{I}b\}$$

$$R = \{r_1 : a \rightarrow_{\mathbf{I}} \neg c,$$

$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_3 : \mathbf{I}b \rightarrow_{\mathbf{I}} d,$$

$$r_4 : \mathbf{I}d, a \rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \emptyset$$

Recovering from violations: reconsidering strong intentions

$$F = \{a, \mathbf{I}b\}$$

$$R = \{r_1 : a \to_{\mathbf{I}} \neg c,$$
$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$
$$r_3 : \mathbf{I}b \to_{\mathbf{I}} d,$$
$$r_4 : \mathbf{I}d, a \to_{\mathbf{I}} \neg c\}$$

$$\succ = \emptyset$$

Intention reconsideration here amounts, e.g., to

$$R - \{r_1, r_4\}$$

Recovering from violations: reconsidering weak intentions

Recovering from violations: reconsidering weak intentions

$$F = \{a, \mathbf{I}d\}$$

$$R = \{r_1 : a \rightarrow_{\mathbf{I}} b,$$

$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_3 : \mathbf{I}d \Rightarrow_{\mathbf{I}} \neg c, \}$$

$$\succ = \emptyset$$

Recovering from violations: reconsidering weak intentions

$$F = \{a, \mathbf{I}d\}$$
$$R = \{r_1 : a \rightarrow_{\mathbf{I}} b,$$
$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$
$$r_3 : \mathbf{I}d \Rightarrow_{\mathbf{I}} \neg c, \}$$
$$\succ = \emptyset$$

Recovering from violations: reconsidering weak intentions

$$F = \{a, \mathbf{I}d\}$$
$$R = \{r_1 : a \to_{\mathbf{I}} b,$$
$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$
$$r_3 : \mathbf{I}d \Rightarrow_{\mathbf{I}} \neg c, \}$$
$$\succ = \emptyset$$

Intention reconsideration here amounts to

$$D^{-}_{\mathbf{I}p_1,\dots\mathbf{I}p_n} = \begin{cases} D & \text{if } \mathbf{I}p_1,\dots,\mathbf{I}p_n \text{ not provable} \\ (F, R^{\mathbf{O}} R^{\mathbf{I}'}, \succ') & \text{otherwise} \end{cases}$$

where

$$R^{\mathbf{I}'} = R^{\mathbf{I}} \cup \{s : \mathbf{I}p_1,\dots,\mathbf{I}p_{i-1},\mathbf{I}p_{i+1},\dots,\mathbf{I}p_n \rightsquigarrow_{\mathbf{I}} \ \sim p_i |$$
$$1 \le i \le n\}$$
$$\succ' = \ \succ - \{r \succ s \mid r \in R^{\mathbf{I}'} - R^{\mathbf{I}}\}.$$

Recovering from violations: reconsidering weak intentions

$$F = \{a, \mathbf{I}d\}$$
$$R = \{r_1 : a \rightarrow_{\mathbf{I}} b,$$
$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$
$$r_3 : \mathbf{I}d \Rightarrow_{\mathbf{I}} \neg c, \; r_4 : \rightsquigarrow_{\mathbf{I}} c\}$$
$$\succ = \emptyset$$

Intention reconsideration here amounts to

$$D^{-}_{\mathbf{I}p_1,\dots\mathbf{I}p_n} = \begin{cases} D & \text{if } \mathbf{I}p_1, \dots, \mathbf{I}p_n \text{ not provable} \\ (F, R^{\mathbf{O}} R^{\mathbf{I}'}, \succ') & \text{otherwise} \end{cases}$$

where

$$R^{\mathbf{I}'} = R^{\mathbf{I}} \cup \{s : \mathbf{I}p_1, \dots, \mathbf{I}p_{i-1}, \mathbf{I}p_{i+1}, \dots, \mathbf{I}p_n \rightsquigarrow_{\mathbf{I}} \; \sim p_i | \; 1 \leq i \leq n\}$$
$$\succ' = \; \succ - \{r \succ s \mid r \in R^{\mathbf{I}'} - R^{\mathbf{I}}\}.$$

Is everything okay?

Is everything okay?

- A similar procedure can be devised to derive new intentions, i.e., by adding new defeasible rules (AGM revision)

$$
\begin{aligned}
F &= \{a\} \\
R &= \{r_1 : a \rightarrow_{\mathbf{I}} b, \\
&\qquad r_2 : a \Rightarrow_{\mathbf{O}} c\} \\
\succ &= \emptyset
\end{aligned}
$$

Is everything okay?

- A similar procedure can be devised to derive new intentions, i.e., by adding new defeasible rules (AGM revision)

$$
\begin{aligned}
F &= \{a\} \\
R &= \{r_1 : a \rightarrow_{\mathbf{I}} b, \\
&\qquad r_2 : a \Rightarrow_{\mathbf{O}} c, \; r_3 : \Rightarrow_{\mathbf{I}} c\} \\
\succ &= \emptyset
\end{aligned}
$$

Is everything okay?

- A similar procedure can be devised to derive new intentions, i.e., by adding new defeasible rules (AGM revision)

$$
\begin{aligned}
F &= \{a\} \\
R &= \{r_1 : a \rightarrow_{\mathbf{I}} b, \\
&\qquad r_2 : a \Rightarrow_{\mathbf{O}} c,\ r_3 : \Rightarrow_{\mathbf{I}} c\} \\
\succ &= \emptyset
\end{aligned}
$$

- However, all the above operations apply only to the last rule of the reasoning chains supporting “illegal” intentions

BIO-DL with paths

BIO-DL with paths

$$
\begin{aligned}
F &= \{\mathbf{O}\,GoToSpain, \mathbf{I}Sarsuela, Hungry\}\\
R &= \{r_1 : \mathbf{I}Sarsuela \rightarrow_{\mathbf{I}} GoToBarcelona,\\
&\qquad r_2 : \mathbf{I}GoToBarcelona \rightarrow_{\mathbf{I}} GoToSpain,\\
&\qquad r_3 : \rightsquigarrow_{\mathbf{O}} \neg EatModerately,\\
&\qquad r_4 : \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} EatModerately,\\
&\qquad r_5 : Hungry \Rightarrow_{\mathbf{I}} \neg EatModerately,\\
&\qquad r_6 : \Rightarrow_{\mathbf{I}} \neg EatModerately,\\
&\qquad r_7 : \mathbf{I}\neg EatModerately, \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} Abstinence\}\\
\succ &= \{r_3 \succ r_4\}
\end{aligned}
$$

BIO-DL with paths

$$F = \{\mathbf{O}\,GoToSpain, \mathbf{I}Sarsuela, Hungry\}$$

$$\begin{aligned} R = \{ & r_1 : \mathbf{I}Sarsuela \rightarrow_{\mathbf{I}} GoToBarcelona, \\ & r_2 : \mathbf{I}GoToBarcelona \rightarrow_{\mathbf{I}} GoToSpain, \\ & r_3 : \rightsquigarrow_{\mathbf{O}} \neg EatModerately, \\ & r_4 : \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} EatModerately, \\ & r_5 : Hungry \Rightarrow_{\mathbf{I}} \neg EatModerately, \\ & r_6 : \Rightarrow_{\mathbf{I}} \neg EatModerately, \\ & r_7 : \mathbf{I}\neg EatModerately, \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} Abstinence\} \end{aligned}$$

$$\succ = \{r_3 \succ r_4\}$$

$+\Delta^{\mathbf{I}}[\mathbf{I}Sarsuela][r_1]GoToBarcelona \quad +\Delta^{\mathbf{I}}[\mathbf{I}Sarsuela][r_1][r_2]GoToSpain$

BIO-DL with paths

$$F = \{\mathbf{O}\, GoToSpain, \mathbf{I} Sarsuela, Hungry\}$$

$$\begin{aligned} R = \{ & r_1 : \mathbf{I} Sarsuela \to_{\mathbf{I}} GoToBarcelona, \\ & r_2 : \mathbf{I} GoToBarcelona \to_{\mathbf{I}} GoToSpain, \\ & r_3 : \rightsquigarrow_{\mathbf{O}} \neg EatModerately, \\ & r_4 : \mathbf{I} Sarsuela \Rightarrow_{\mathbf{O}} EatModerately, \\ & r_5 : Hungry \Rightarrow_{\mathbf{I}} \neg EatModerately, \\ & r_6 : \Rightarrow_{\mathbf{I}} \neg EatModerately, \\ & r_7 : \mathbf{I} \neg EatModerately, \mathbf{I} Sarsuela \Rightarrow_{\mathbf{O}} Abstinence\} \end{aligned}$$

$$\succ = \{r_3 \succ r_4\}$$

$+\Delta^{\mathbf{I}}[\mathbf{I} Sarsuela][r_1] GoToBarcelona$ $+\Delta^{\mathbf{I}}[\mathbf{I} Sarsuela][r_1][r_2] GoToSpain$
$+\partial^{\mathbf{O}}[\mathbf{I} Sarsuela][r_4] EatModerately$ $+\partial^{\mathbf{I}}[Hungry][r_5]\neg EatModerately$
$+\partial^{\mathbf{I}}[r_6]\neg EatModerately$

BIO-DL with paths

$$F = \{\mathbf{O}\,GoToSpain, \mathbf{I}Sarsuela, Hungry\}$$

$$\begin{aligned} R = \{ & r_1 : \mathbf{I}Sarsuela \rightarrow_{\mathbf{I}} GoToBarcelona, \\ & r_2 : \mathbf{I}GoToBarcelona \rightarrow_{\mathbf{I}} GoToSpain, \\ & r_3 : \leadsto_{\mathbf{O}} \neg EatModerately, \\ & r_4 : \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} EatModerately, \\ & r_5 : Hungry \Rightarrow_{\mathbf{I}} \neg EatModerately, \\ & r_6 : \Rightarrow_{\mathbf{I}} \neg EatModerately, \\ & r_7 : \mathbf{I}\neg EatModerately, \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} Abstinence\} \end{aligned}$$

$$\succ = \{r_3 \succ r_4\}$$

$+\Delta^{\mathbf{I}}[\mathbf{I}Sarsuela][r_1]GoToBarcelona$ $\quad +\Delta^{\mathbf{I}}[\mathbf{I}Sarsuela][r_1][r_2]GoToSpain$

$+\partial^{\mathbf{O}}[\mathbf{I}Sarsuela][r_4]EatModerately$ $\quad +\partial^{\mathbf{I}}[Hungry][r_5]\neg EatModerately$

$+\partial^{\mathbf{I}}[r_6]\neg EatModerately$

$+\partial^{\mathbf{O}}[Hungry][r_5, \mathbf{I}Sarsuela][r_7]Abstinence$ $\quad +\partial^{\mathbf{O}}[r_6, \mathbf{I}Sarsuela]Abstinence$

BIO-DL with paths

$$F = \{\mathbf{O}\,GoToSpain, \mathbf{I}Sarsuela, Hungry\}$$

$$\begin{aligned}
R = \{ & r_1 : \mathbf{I}Sarsuela \rightarrow_{\mathbf{I}} GoToBarcelona, \\
& r_2 : \mathbf{I}GoToBarcelona \rightarrow_{\mathbf{I}} GoToSpain, \\
& r_3 : \rightsquigarrow_{\mathbf{O}} \neg EatModerately, \\
& r_4 : \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} EatModerately, \\
& r_5 : Hungry \Rightarrow_{\mathbf{I}} \neg EatModerately, \\
& r_6 : \Rightarrow_{\mathbf{I}} \neg EatModerately, \\
& r_7 : \mathbf{I}\neg EatModerately, \mathbf{I}Sarsuela \Rightarrow_{\mathbf{O}} Abstinence\}
\end{aligned}$$

$$\succ = \{r_3 \succ r_4\}$$

$+\Delta^{\mathbf{I}}[\mathbf{I}Sarsuela][r_1]GoToBarcelona$

$+\Delta^{\mathbf{I}}[\mathbf{I}Sarsuela][r_1][r_2]GoToSpain$

$+\partial^{\mathbf{O}}[\mathbf{I}Sarsuela][r_4]EatModerately$

$+\partial^{\mathbf{I}}[Hungry][r_5]\neg EatModerately$

$+\partial^{\mathbf{I}}[r_6]\neg EatModerately$

$+\partial^{\mathbf{O}}[Hungry][r_5, \mathbf{I}Sarsuela][r_7]Abstinence$

$+\partial^{\mathbf{O}}[r_6, \mathbf{I}Sarsuela]Abstinence$

Reconsidering intentions: strong intentions

Reconsidering intentions: strong intentions

Definition (Rule Removal with paths)

Let $D = (F, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$ be an agent theory.
For each $r \in R^{\mathbf{O}}_{\mathrm{sd}}$ such that the paths $\mathcal{L}_1, \dots \mathcal{L}_n$ such that

$$D \vdash +\Delta^{\mathbf{I}}\mathcal{L}_1 p, \dots, D \vdash +\Delta^{\mathbf{I}}\mathcal{L}_n p$$

and

$$D \vdash +\partial^{\mathbf{O}}\mathcal{Y}\neg p$$

D_{-X} is such that

1. $X = \{w_1, \dots, w_m\}$ is the smallest set of strict rules in $R^{\mathbf{I}}$ such that, for each $k \in \{1, \dots, n\}$, there is at least a $w_j \in X$ that occurs in $\mathcal{L}_k$,
2. $R^{\mathbf{I}}_{-X} = R^{\mathbf{I}} - X$, and
3. $F_{-X} = F$, $R^{\mathbf{O}}_{-X} = R^{\mathbf{O}}$, and $\succ_{-X} = \succ$.

Reconsidering intentions: strong intentions

$$F = \{a, \mathbf{I}b\}$$

$$R = \{r_1 : a \rightarrow_{\mathbf{I}} \neg c,$$

$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_3 : \mathbf{I}b \rightarrow_{\mathbf{I}} d,$$

$$r_4 : \mathbf{I}d, a \rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \emptyset$$

$+\partial^{\mathbf{O}}[\mathbf{I}b][r_2]c$

$+\Delta^{\mathbf{I}}[a][r_1]\neg c \quad +\Delta^{\mathbf{I}}[\mathbf{I}b][r_3, a][r_4]\neg c$

Reconsidering intentions: strong intentions

$$F = \{a, \mathbf{I}b\}$$

$$R = \{r_1 : a \rightarrow_{\mathbf{I}} \neg c,$$

$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_3 : \mathbf{I}b \rightarrow_{\mathbf{I}} d,$$

$$r_4 : \mathbf{I}d, a \rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \emptyset$$

$+\partial^{\mathbf{O}}[\mathbf{I}b][r_2]c$

$+\Delta^{\mathbf{I}}[a][r_1]\neg c \quad +\Delta^{\mathbf{I}}[\mathbf{I}b][r_3, a][r_4]\neg c$

Reconsidering intentions: strong intentions

$$F = \{a, \mathbf{I}b\}$$
$$R = \{r_1 : a \to_{\mathbf{I}} \neg c,$$
$$r_2 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$
$$r_3 : \mathbf{I}b \to_{\mathbf{I}} d,$$
$$r_4 : \mathbf{I}d, a \to_{\mathbf{I}} \neg c\}$$
$$\succ = \emptyset$$

$+\partial^{\mathbf{O}}[\mathbf{I}b][r_2]c$

$+\Delta^{\mathbf{I}}[a][r_1]\neg c \quad +\Delta^{\mathbf{I}}[\mathbf{I}b][r_3, a][r_4]\neg c$

Reconsidering intentions: weak intentions (simplified)

Reconsidering intentions: weak intentions (simplified)

Definition (Contraction with paths)

Let $D = (F, R^{\mathbf{O}}, R^{\mathbf{I}}, \succ)$ be an agent theory.
For each $r \in R^{\mathbf{O}}_{\mathrm{sd}}$ such that the paths $\mathcal{L}_1, \ldots \mathcal{L}_n$ such that

$$D \vdash +\partial^{\mathbf{I}}\mathcal{L}_1 p, \ldots, D \vdash +\partial^{\mathbf{I}}\mathcal{L}_n p$$

and

$$D \vdash +\partial^{\mathbf{O}}\mathcal{Y}\neg p$$

the theory the theory $D_{\flat p} = (F, R^{\mathbf{O}}, R^{\mathbf{I}\prime}, \succ')$ is such that

1. $R^{\mathbf{I}\prime} = R^{\mathbf{I}} \cup \{s :\rightsquigarrow_{\mathbf{I}} \sim q\} \cup \{t :\rightsquigarrow_{\mathbf{I}} \sim x\}$,
2. $\succ' = \succ - [\{r_k \succ s | r_k \in R^{\mathbf{I}}[q], r_k$ occurs in $\mathcal{L}_k \ \forall k \in \{1, \ldots, n\}\} \cup \{w \succ t \mid w$ is rebutted and is such that its head is p or w occurs in $\mathcal{L}_k \ \forall k \in \{1, \ldots, n\}\}]$.

Reconsidering intentions: weak intentions (simplified)

$$F = \{\mathbf{I}b\}$$

$$R = \{r_1 : \Rightarrow_{\mathbf{I}} a,$$

$$r_2 : \mathbf{I}a \Rightarrow_{\mathbf{I}} \neg c,$$

$$r_3 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_4 : \mathbf{I}a \Rightarrow_{\mathbf{I}} p,$$

$$r_5 : \Rightarrow_{\mathbf{I}} \neg p,$$

$$r_6 : \mathbf{I}\neg p \Rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \{r_5 \succ r_4\}$$

$$+\partial^{\mathbf{O}}[\mathbf{I}b][r_3]c$$

$$+\partial^{\mathbf{I}}[r_1][r_2]\neg c \qquad +\partial^{\mathbf{I}}[r_1][r_4]p$$

$$-\partial^{\mathbf{I}}[-r_5][-r_6]\neg c$$

Reconsidering intentions: weak intentions (simplified)

$$F = \{\mathbf{I}b\}$$

$$R = \{r_1 :\Rightarrow_{\mathbf{I}} a, \qquad \Longleftarrow r_1' :\leadsto_{\mathbf{I}} \neg a$$

$$r_2 : \mathbf{I}a \Rightarrow_{\mathbf{I}} \neg c,$$

$$r_3 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_4 : \mathbf{I}a \Rightarrow_{\mathbf{I}} p,$$

$$r_5 :\Rightarrow_{\mathbf{I}} \neg p,$$

$$r_6 : \mathbf{I}\neg p \Rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \{r_5 \succ r_4\}$$

$$+\partial^{\mathbf{O}}[\mathbf{I}b][r_3]c$$

$$+\partial^{\mathbf{I}}[r_1][r_2]\neg c \qquad +\partial^{\mathbf{I}}[r_1][r_4]p$$

$$-\partial^{\mathbf{I}}[-r_5][-r_6]\neg c$$

Reconsidering intentions: weak intentions (simplified)

$$F = \{\mathbf{I}b\}$$

$$R = \{r_1 : \Rightarrow_{\mathbf{I}} a, \qquad \Longleftarrow r_1' : \rightsquigarrow_{\mathbf{I}} \neg a$$

$$r_2 : \mathbf{I}a \Rightarrow_{\mathbf{I}} \neg c,$$

$$r_3 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_4 : \mathbf{I}a \Rightarrow_{\mathbf{I}} p,$$

$$r_5 : \Rightarrow_{\mathbf{I}} \neg p,$$

$$r_6 : \mathbf{I}\neg p \Rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \{r_5 \succ r_4\}$$

$$+\partial^{\mathbf{O}}[\mathbf{I}b][r_3]c$$
$$-\partial^{\mathbf{I}}[-r_1][r_2]\neg c$$
$$+\partial^{\mathbf{I}}[r_5][r_6]\neg c$$

Reconsidering intentions: weak intentions (simplified)

$$F = \{\mathbf{I}b\}$$

$$R = \{r_1 : \Rightarrow_{\mathbf{I}} a, \quad \Longleftarrow r_1' : \rightsquigarrow_{\mathbf{I}} \neg a$$

$$r_2 : \mathbf{I}a \Rightarrow_{\mathbf{I}} \neg c,$$

$$r_3 : \mathbf{I}b \Rightarrow_{\mathbf{O}} c,$$

$$r_4 : \mathbf{I}a \Rightarrow_{\mathbf{I}} p,$$

$$r_5 : \Rightarrow_{\mathbf{I}} \neg p, \quad \Longleftarrow r_5' : \rightsquigarrow_{\mathbf{I}} p$$

$$r_6 : \mathbf{I}\neg p \Rightarrow_{\mathbf{I}} \neg c\}$$

$$\succ = \{r_5 \succ r_4\}$$

$$+\partial^{\mathbf{O}}[\mathbf{I}b][r_3]c$$

$$-\partial^{\mathbf{I}}[-r_1][r_2]\neg c$$

$$-\partial^{\mathbf{I}}[-r_5][r_6]\neg c$$

Results and future work

Theorem

Rule-removal with paths and contraction with paths satisfy AGM postulates for contraction

Results and future work

Theorem

Rule-removal with paths and contraction with paths satisfy AGM postulates for contraction

- The role of reparative obligations?

Results and future work

Theorem

Rule-removal with paths and contraction with paths satisfy AGM postulates for contraction

- The role of reparative obligations?
- Complexity?

Results and future work

Theorem

Rule-removal with paths and contraction with paths satisfy AGM postulates for contraction

- The role of reparative obligations?
- Complexity?
- Revise priorities and not rules?

Thanks!