European Conference on Logics in Artificial Intelligence (JELIA06)

Introducing ***Attempt*** in a modal logic of intentional action

Emiliano Lorini[1], Andreas Herzig[2], Cristiano Castelfranchi[1]

[1]*Institute of Cognitive Science and Technologies-CNR (Rome, Italy)*

[2]*Institut de Recherche en Informatique de Toulouse (France)*

Outline

- Attempt in Philosophy
- Logic of Intention and Attempt

Outline

- Attempt in Philosophy
- Logic of Intention and Attempt

General definitions

Crucial concept for relating intentional agency with intentional performance (Hornsby 1980, Prichard 1949, McCann 1972).

•

•

.

•

General definitions

Crucial concept for relating intentional agency with intentional performance (Hornsby 1980, Prichard 1949, McCann 1972).

- "The movement of a person's arm is the product of a series of external causes; but some event (the ATTEMPT), and presumably one of those that took place within the brain, was caused by the agent and not by any other events" (Chisholm, 1966).
-
-

General definitions

Crucial concept for relating intentional agency with intentional performance (Hornsby 1980, Prichard 1949, McCann 1972).

- "The movement of a person's arm is the product of a series of external causes; but some event (the ATTEMPT), and presumably one of those that took place within the brain, was caused by the agent and not by any other events" (Chisholm, 1966).

- "If an agent at an instant in time realizes that that instant is an instant at which he intends to perform action x, then logically necessarily he begins trying to do x" (O'Shaughnessy, 1973).

-

General definitions

Crucial concept for relating intentional agency with intentional performance (Hornsby 1980, Prichard 1949, McCann 1972).

- "The movement of a person's arm is the product of a series of external causes; but some event (the ATTEMPT), and presumably one of those that took place within the brain, was caused by the agent and not by any other events" (Chisholm, 1966).

- "If an agent at an instant in time realizes that that instant is an instant at which he intends to perform action x, then logically necessarily he begins trying to do x" (O'Shaughnessy, 1973).

- In New Volitional Theory of Action: "Attempt to do α is the agent's mental act of exterting himself to do α" (Ginet, 1990).

Basic vs Complex Action (1)

According to (Goldman 1970, Danto 1965) we can say that the agent has performed the (intentional) BASIC ACTION of α-ing if and only if:

- the agent intended to perform α;
- the agent successfully performed α and α was not performed BY performing (and intending to perform) some other action β different from α (that is α's execution was directly controlled by the agent).

Basic vs Complex Action (1)

According to (Goldman 1970, Danto 1965) we can say that the agent has performed the (intentional) BASIC ACTION of α-ing if and only if:

- the agent intended to perform α;
- the agent successfully performed α and α was not performed BY performing (and intending to perform) some other action β different from α (that is α's execution was directly controlled by the agent).

Basic Actions are normally conceived as bodily movements of a human or a robot: raising the arm, moving the leg, turning the sensor etc...

Basic vs Complex Action (2)

COMPLEX ACTIONS are actions that an agent intentionally does BY doing (and intending to do) some more elementary action (BY expresses *counterfactual dependence*, Goldman 1970).

-

-

Basic vs Complex Action (2)

COMPLEX ACTIONS are actions that an agent intentionally does BY doing (and intending to do) some more elementary action (BY expresses *counterfactual dependence*, Goldman 1970).

Making the definition more precise...

-
-

Basic vs Complex Action (2)

COMPLEX ACTIONS are actions that an agent intentionally does BY doing (and intending to do) some more elementary action (BY expresses *counterfactual dependence*, Goldman 1970).

Making the definition more precise...

An agent *i*'complex action β is an action that agent *i* intentionally does:

- by doing (and intending to do) some more elementary action α and

- by relying on some external natural event, some other agent *j*'s action or some state of affairs.

Turning on the light

- *The agent intends to do* β = *"turn on the light in the room" (the agent intends to bring it about that* P = *"the light is on in the room").*
-
  -
  -

Turning on the light

- *The agent intends to do β = "turn on the light in the room" (the agent intends to bring it about that P = "the light is on in the room").*

- *β is done by doing*
  - *the intentional action α of "flipping the switch" and*
  - *by relying on the event "the electric circuit will bring it about that the light is on".*

Attempt and Basic Action (1)

exploiting ATTEMPT for defining BASIC ACT...

-
  -
  -

Attempt and Basic Action (1)

exploiting ATTEMPT for defining BASIC ACT...

- We say that the agent has performed the (intentional) BASIC ACTION α if and only if:
  - the agent intended to perform α (or the agent intended to try to perform α);
  - the agent successfully performed α simply BY *attempting* to perform it.

Attempt and Basic Action (1)

exploiting ATTEMPT for defining BASIC ACT...

- We say that the agent has performed the (intentional) BASIC ACTION α if and only if:
  - the agent intended to perform α (or the agent intended to try to perform α);
  - the agent successfully performed α simply BY *attempting* to perform it.

The agent has raised the arm above the head (having the intention to do it) simply by trying to raise the arm above the head.

Attempt and Basic Action (2)

But a BASIC ACTION α is *successfully* performed by the agent only if the execution preconditions of α hold.

-
  -
  -

Attempt and Basic Action (2)

But a BASIC ACTION α is *successfully* performed by the agent only if the execution preconditions of α hold.

- We say that the agent has performed the (intentional) BASIC ACTION α if and only if:
  - the agent intended to perform α (or the agent intended to try to perform α);
  - the agent *attempted* to perform α and the execution preconditions of α did hold.

Attempt and Basic Action (3)

- We say that the agent has *failed* to perform the (intentional) BASIC ACTION α if and only if:
  - the agent intended to perform α (or the agent intended to try to perform α);
  - the agent *attempted* to perform α and the execution preconditions of α did not hold.

Example: Raising my arm

Preconditions for "raising my arm above the head":

PRECOND 1: my arm is not paralyzed. PRECOND 2: my arm is not tied (or is not blocked).

1.

2.

3.

Example: Raising my arm

Preconditions for "raising my arm above the head":

PRECOND 1: my arm is not paralyzed. PRECOND 2: my arm is not tied (or is not blocked).

1. If both precond hold and I attempt to raise my arm above the head, I *successfully* raise my arm above the head.

2.

3.

Example: Raising my arm

Preconditions for "raising my arm above the head":

PRECOND 1: my arm is not paralyzed. PRECOND 2: my arm is not tied (or is not blocked).

1. If both precond hold and I attempt to raise my arm above the head, I *successfully* raise my arm above the head.

2. If I attempt to raise my arm when my arm is paralyzed then my attempt *completely fails*: the external world is unaffected by the attempt (and nobody perceives it).

3.

Example: Raising my arm

Preconditions for "raising my arm above the head":

PRECOND 1: my arm is not paralyzed. PRECOND 2: my arm is not tied (or is not blocked).

1. If both precond hold and I attempt to raise my arm above the head, I *successfully* raise my arm above the head.

2. If I attempt to raise my arm when my arm is paralyzed then my attempt *completely fails*: the external world is unaffected by the attempt (and nobody perceives it).

3. If my arm is simply tied with a rope and I attempt to raise it, I move it of few centimeters: *failed but "partial" execution of the basic action*.

Outline

- Attempt in Philosophy

- Logic of Intention and Attempt

$\mathcal{LIA}$: A logic of Intention and Attempt

- Multi-modal logic of time, attempts, goals and beliefs.
-
-

$\mathcal{LIA}$: A logic of Intention and Attempt

- Multi-modal logic of time, attempts, goals and beliefs.

- Based on a combination of an enhanced version of linear temporal logic with actions and Cohen and Levesque's logic of goal and intention.

-

$\mathcal{LIA}$: A logic of Intention and Attempt

- Multi-modal logic of time, attempts, goals and beliefs.

- Based on a combination of an enhanced version of linear temporal logic with actions and Cohen and Levesque's logic of goal and intention.

- The notion of *atomic (basic) action* is substituted with the more primitive notion *attempt*.

The language

$$\varphi := p|\top|\neg\varphi|\varphi \wedge \psi|\, [[i,\alpha]]\,\varphi|G\varphi|X\varphi|\varphi Until\psi|Bel_i\varphi|Goal_i\varphi$$

The language

$$\varphi := p|\top|\neg\varphi|\varphi \wedge \psi|\,[[i,\alpha]]\,\varphi|G\varphi|X\varphi|\varphi Until\psi|Bel_i\varphi|Goal_i\varphi$$

Temporal formulas

HENCEFORTH, NEXT, UNTIL

The language

$$\varphi := p|\top|\neg\varphi|\varphi \wedge \psi|[[i,\alpha]]\ \varphi|G\varphi|X\varphi|\varphi Until \psi|Bel_i\varphi|Goal_i\varphi$$

Attempt formulas

"if agent *i* attempts to do α then φ holds after α's occurrence".

The language

$$\varphi := p|\top|\neg\varphi|\varphi \wedge \psi|\,[[i,\alpha]]\,\varphi|G\varphi|X\varphi|\varphi Until\psi|Bel_i\varphi|Goal_i\varphi$$

Belief and Goal formulas

"agent *i* believes that/wants that φ holds".

Proof system

0a. All tautologies of propositional calculus

1a. $G(\varphi \longrightarrow \psi) \longrightarrow (G\varphi \longrightarrow G\psi)$

2a. $X\neg\varphi \longleftrightarrow \neg X\varphi$

3a. $X(\varphi \longrightarrow \psi) \longrightarrow (X\varphi \longrightarrow X\psi)$

4a. $G\varphi \longrightarrow \varphi \wedge XG\varphi$

5a. $G(\varphi \longrightarrow X\varphi) \longrightarrow (\varphi \longrightarrow G\varphi)$

6a. $\varphi Until\psi \longrightarrow F\psi$

7a. $\varphi Until\psi \longleftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$

Inference Rules:

R1. $\frac{\vdash\varphi \; \vdash\varphi\longrightarrow\psi}{\vdash\psi}$ (*modus ponens*)

R2. $\frac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)

R3. $\frac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)

R4. $\frac{\vdash\varphi}{\vdash Bel\varphi}$ (*Bel-necessitation*)

R5. $\frac{\vdash\varphi}{\vdash Goal\varphi}$ (*Goal-necessitation*)

1b. $Bel_i\varphi \wedge Bel_i(\varphi \longrightarrow \psi) \longrightarrow Bel_i\psi$

2b. $\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$

3b. $Bel_i\varphi \longrightarrow Bel_iBel_i\varphi$

4b. $\neg Bel_i\varphi \longrightarrow Bel_i\neg Bel_i\varphi$

5b. $Goal_i\varphi \wedge Goal_i(\varphi \longrightarrow \psi) \longrightarrow Goal_i\psi$

6b. $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$

7b. $Goal_i\varphi \longrightarrow Bel_iGoal_i\varphi$

8b. $\neg Goal_i\varphi \longrightarrow Bel_i\neg Goal_i\varphi$

9b. $Bel_i\varphi \longrightarrow Goal_i\varphi$

10b. $Bel_i\,[[j,\alpha]]\,\psi \wedge \neg Bel_i\,[[j,\alpha]]\,\bot \longrightarrow [[j,\alpha]]\,Bel_i\psi$

11b. $[[j,\alpha]]\,Bel_i\psi \wedge \neg\,[[i,\alpha]]\,\bot \longrightarrow Bel_i\,[[j,\alpha]]\,\psi$

12b. $Bel_i(GBel_i\psi \longleftrightarrow Bel_iG\psi)$

13b. $Goal_i\,\langle\langle i,\alpha\rangle\rangle\,\top \longleftrightarrow \langle\langle i,\alpha\rangle\rangle\,\top$

14b. $[[i,\alpha]]\,\varphi \wedge [[i,\alpha]]\,(\varphi \longrightarrow \psi) \longrightarrow [[i,\alpha]]\,\psi$

15b. $X\varphi \longrightarrow [[i,\alpha]]\,\varphi$

Proof system: Time

Standard proof system of LTL (Goldblatt, 1990; Gabbay et al., 1980)

0a. All tautologies of propositional calculus

1a. $G(\varphi \longrightarrow \psi) \longrightarrow (G\varphi \longrightarrow G\psi)$

2a. $X\neg\varphi \longleftrightarrow \neg X\varphi$

3a. $X(\varphi \longrightarrow \psi) \longrightarrow (X\varphi \longrightarrow X\psi)$

4a. $G\varphi \longrightarrow \varphi \wedge XG\varphi$

5a. $G(\varphi \longrightarrow X\varphi) \longrightarrow (\varphi \longrightarrow G\varphi)$

6a. $\varphi Until\psi \longrightarrow F\psi$

7a. $\varphi Until\psi \longleftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$

Inference Rules:

R1. $\frac{\vdash\varphi \;\vdash\varphi\longrightarrow\psi}{\vdash\psi}$ (*modus ponens*)

R2. $\frac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)

R3. $\frac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)

R4. $\frac{\vdash\varphi}{\vdash Bel\varphi}$ (*Bel-necessitation*)

R5. $\frac{\vdash\varphi}{\vdash Goal\varphi}$ (*Goal-necessitation*)

1b. $Bel_i\varphi \wedge Bel_i(\varphi \longrightarrow \psi) \longrightarrow Bel_i\psi$

2b. $\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$

3b. $Bel_i\varphi \longrightarrow Bel_iBel_i\varphi$

4b. $\neg Bel_i\varphi \longrightarrow Bel_i\neg Bel_i\varphi$

5b. $Goal_i\varphi \wedge Goal_i(\varphi \longrightarrow \psi) \longrightarrow Goal_i\psi$

6b. $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$

7b. $Goal_i\varphi \longrightarrow Bel_iGoal_i\varphi$

8b. $\neg Goal_i\varphi \longrightarrow Bel_i\neg Goal_i\varphi$

9b. $Bel_i\varphi \longrightarrow Goal_i\varphi$

10b. $Bel_i\,[[j,\alpha]]\,\psi \wedge \neg Bel_i\,[[j,\alpha]]\,\bot \longrightarrow [[j,\alpha]]\,Bel_i\psi$

11b. $[[j,\alpha]]\,Bel_i\psi \wedge \neg\,[[i,\alpha]]\,\bot \longrightarrow Bel_i\,[[j,\alpha]]\,\psi$

12b. $Bel_i(GBel_i\psi \longleftrightarrow Bel_iG\psi)$

13b. $Goal_i\,\langle\langle i,\alpha\rangle\rangle\,\top \longleftrightarrow \langle\langle i,\alpha\rangle\rangle\,\top$

14b. $[[i,\alpha]]\,\varphi \wedge [[i,\alpha]]\,(\varphi \longrightarrow \psi) \longrightarrow [[i,\alpha]]\,\psi$

15b. $X\varphi \longrightarrow [[i,\alpha]]\,\varphi$

Proof system: Bel and Goal

KD45 logic for Bel and Goal + Positive and Negative Introspection of Goals + Inclusion Bel/Goal (Cohen & Levesque, 1990)

0a. All tautologies of propositional calculus	1b. $Bel_i\varphi \wedge Bel_i(\varphi \longrightarrow \psi) \longrightarrow Bel_i\psi$
1a. $G(\varphi \longrightarrow \psi) \longrightarrow (G\varphi \longrightarrow G\psi)$	2b. $\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$
2a. $X\neg\varphi \longleftrightarrow \neg X\varphi$	3b. $Bel_i\varphi \longrightarrow Bel_iBel_i\varphi$
3a. $X(\varphi \longrightarrow \psi) \longrightarrow (X\varphi \longrightarrow X\psi)$	4b. $\neg Bel_i\varphi \longrightarrow Bel_i\neg Bel_i\varphi$
4a. $G\varphi \longrightarrow \varphi \wedge XG\varphi$	5b. $Goal_i\varphi \wedge Goal_i(\varphi \longrightarrow \psi) \longrightarrow Goal_i\psi$
5a. $G(\varphi \longrightarrow X\varphi) \longrightarrow (\varphi \longrightarrow G\varphi)$	6b. $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$
6a. $\varphi Until\psi \longrightarrow F\psi$	7b. $Goal_i\varphi \longrightarrow Bel_iGoal_i\varphi$
7a. $\varphi Until\psi \longleftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$	8b. $\neg Goal_i\varphi \longrightarrow Bel_i\neg Goal_i\varphi$
Inference Rules:	9b. $Bel_i\varphi \longrightarrow Goal_i\varphi$
R1. $\frac{\vdash\varphi \ \vdash\varphi\longrightarrow\psi}{\vdash\psi}$ (*modus ponens*)	10b. $Bel_i\,[[j,\alpha]]\,\psi \wedge \neg Bel_i\,[[j,\alpha]]\,\bot \longrightarrow [[j,\alpha]]\,Bel_i\psi$
R2. $\frac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)	11b. $[[j,\alpha]]\,Bel_i\psi \wedge \neg\,[[i,\alpha]]\,\bot \longrightarrow Bel_i\,[[j,\alpha]]\,\psi$
R3. $\frac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)	12b. $Bel_i(GBel_i\psi \longleftrightarrow Bel_iG\psi)$
R4. $\frac{\vdash\varphi}{\vdash Bel\varphi}$ (*Bel-necessitation*)	13b. $Goal_i\,\langle\langle i,\alpha\rangle\rangle\,\top \longleftrightarrow \langle\langle i,\alpha\rangle\rangle\,\top$
R5. $\frac{\vdash\varphi}{\vdash Goal\varphi}$ (*Goal-necessitation*)	14b. $[[i,\alpha]]\,\varphi \wedge [[i,\alpha]]\,(\varphi \longrightarrow \psi) \longrightarrow [[i,\alpha]]\,\psi$
	15b. $X\varphi \longrightarrow [[i,\alpha]]\,\varphi$

Proof system: NL & NF for Bel

No Learning and No Forgetting for Bel (Herzig & Longin, 2004)

0a. All tautologies of propositional calculus

1a. $G(\varphi \longrightarrow \psi) \longrightarrow (G\varphi \longrightarrow G\psi)$

2a. $X\neg\varphi \longleftrightarrow \neg X\varphi$

3a. $X(\varphi \longrightarrow \psi) \longrightarrow (X\varphi \longrightarrow X\psi)$

4a. $G\varphi \longrightarrow \varphi \wedge XG\varphi$

5a. $G(\varphi \longrightarrow X\varphi) \longrightarrow (\varphi \longrightarrow G\varphi)$

6a. $\varphi Until\psi \longrightarrow F\psi$

7a. $\varphi Until\psi \longleftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$

Inference Rules:

R1. $\frac{\vdash\varphi \ \vdash\varphi\longrightarrow\psi}{\vdash\psi}$ (*modus ponens*)

R2. $\frac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)

R3. $\frac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)

R4. $\frac{\vdash\varphi}{\vdash Bel\varphi}$ (*Bel-necessitation*)

R5. $\frac{\vdash\varphi}{\vdash Goal\varphi}$ (*Goal-necessitation*)

1b. $Bel_i\varphi \wedge Bel_i(\varphi \longrightarrow \psi) \longrightarrow Bel_i\psi$

2b. $\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$

3b. $Bel_i\varphi \longrightarrow Bel_iBel_i\varphi$

4b. $\neg Bel_i\varphi \longrightarrow Bel_i\neg Bel_i\varphi$

5b. $Goal_i\varphi \wedge Goal_i(\varphi \longrightarrow \psi) \longrightarrow Goal_i\psi$

6b. $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$

7b. $Goal_i\varphi \longrightarrow Bel_iGoal_i\varphi$

8b. $\neg Goal_i\varphi \longrightarrow Bel_i\neg Goal_i\varphi$

9b. $Bel_i\varphi \longrightarrow Goal_i\varphi$

10b. $Bel_i\,[[j,\alpha]]\,\psi \wedge \neg Bel_i\,[[j,\alpha]]\,\bot \longrightarrow [[j,\alpha]]\,Bel_i\psi$

11b. $[[j,\alpha]]\,Bel_i\psi \wedge \neg\,[[i,\alpha]]\,\bot \longrightarrow Bel_i\,[[j,\alpha]]\,\psi$

12b. $Bel_i(GBel_i\psi \longleftrightarrow Bel_iG\psi)$

13b. $Goal_i\,\langle\langle i,\alpha\rangle\rangle\,\top \longleftrightarrow \langle\langle i,\alpha\rangle\rangle\,\top$

14b. $[[i,\alpha]]\,\varphi \wedge [[i,\alpha]]\,(\varphi \longrightarrow \psi) \longrightarrow [[i,\alpha]]\,\psi$

15b. $X\varphi \longrightarrow [[i,\alpha]]\,\varphi$

Proof system: Time/Attempt

Interaction Time/Attempt (Broersen, 2003)

0a. All tautologies of propositional calculus

1a. $G(\varphi \longrightarrow \psi) \longrightarrow (G\varphi \longrightarrow G\psi)$

2a. $X\neg\varphi \longleftrightarrow \neg X\varphi$

3a. $X(\varphi \longrightarrow \psi) \longrightarrow (X\varphi \longrightarrow X\psi)$

4a. $G\varphi \longrightarrow \varphi \wedge XG\varphi$

5a. $G(\varphi \longrightarrow X\varphi) \longrightarrow (\varphi \longrightarrow G\varphi)$

6a. $\varphi Until\psi \longrightarrow F\psi$

7a. $\varphi Until\psi \longleftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$

Inference Rules:

R1. $\frac{\vdash\varphi \vdash\varphi\longrightarrow\psi}{\vdash\psi}$ (*modus ponens*)

R2. $\frac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)

R3. $\frac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)

R4. $\frac{\vdash\varphi}{\vdash Bel\varphi}$ (*Bel-necessitation*)

R5. $\frac{\vdash\varphi}{\vdash Goal\varphi}$ (*Goal-necessitation*)

1b. $Bel_i\varphi \wedge Bel_i(\varphi \longrightarrow \psi) \longrightarrow Bel_i\psi$

2b. $\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$

3b. $Bel_i\varphi \longrightarrow Bel_iBel_i\varphi$

4b. $\neg Bel_i\varphi \longrightarrow Bel_i\neg Bel_i\varphi$

5b. $Goal_i\varphi \wedge Goal_i(\varphi \longrightarrow \psi) \longrightarrow Goal_i\psi$

6b. $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$

7b. $Goal_i\varphi \longrightarrow Bel_iGoal_i\varphi$

8b. $\neg Goal_i\varphi \longrightarrow Bel_i\neg Goal_i\varphi$

9b. $Bel_i\varphi \longrightarrow Goal_i\varphi$

10b. $Bel_i\,[[j,\alpha]]\,\psi \wedge \neg Bel_i\,[[j,\alpha]]\,\bot \longrightarrow [[j,\alpha]]\,Bel_i\psi$

11b. $[[j,\alpha]]\,Bel_i\psi \wedge \neg\,[[i,\alpha]]\,\bot \longrightarrow Bel_i\,[[j,\alpha]]\,\psi$

12b. $Bel_i(GBel_i\psi \longleftrightarrow Bel_iG\psi)$

13b. $Goal_i\,\langle\langle i,\alpha\rangle\rangle\,\top \longleftrightarrow \langle\langle i,\alpha\rangle\rangle\,\top$

14b. $[[i,\alpha]]\,\varphi \wedge [[i,\alpha]]\,(\varphi \longrightarrow \psi) \longrightarrow [[i,\alpha]]\,\psi$

15b. $X\varphi \longrightarrow [[i,\alpha]]\,\varphi$

Proof system: Mind/World

Interaction Goal/Attempt

0a. All tautologies of propositional calculus	1b. $Bel_i\varphi \wedge Bel_i(\varphi \longrightarrow \psi) \longrightarrow Bel_i\psi$
1a. $G(\varphi \longrightarrow \psi) \longrightarrow (G\varphi \longrightarrow G\psi)$	2b.$\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$
2a. $X\neg\varphi \longleftrightarrow \neg X\varphi$	3b. $Bel_i\varphi \longrightarrow Bel_iBel_i\varphi$
3a. $X(\varphi \longrightarrow \psi) \longrightarrow (X\varphi \longrightarrow X\psi)$	4b. $\neg Bel_i\varphi \longrightarrow Bel_i\neg Bel_i\varphi$
4a. $G\varphi \longrightarrow \varphi \wedge XG\varphi$	5b. $Goal_i\varphi \wedge Goal_i(\varphi \longrightarrow \psi) \longrightarrow Goal_i\psi$
5a. $G(\varphi \longrightarrow X\varphi) \longrightarrow (\varphi \longrightarrow G\varphi)$	6b.$\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$
6a. $\varphi Until\psi \longrightarrow F\psi$	7b. $Goal_i\varphi \longrightarrow Bel_iGoal_i\varphi$
7a. $\varphi Until\psi \longleftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$	8b. $\neg Goal_i\varphi \longrightarrow Bel_i\neg Goal_i\varphi$
Inference Rules:	9b. $Bel_i\varphi \longrightarrow Goal_i\varphi$
R1. $\frac{\vdash\varphi \ \vdash\varphi\longrightarrow\psi}{\vdash\psi}$ (*modus ponens*)	10b. $Bel_i\,[[j,\alpha]]\,\psi \wedge \neg Bel_i\,[[j,\alpha]]\,\bot \longrightarrow [[j,\alpha]]\,Bel_i\psi$
R2. $\frac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)	11b. $[[j,\alpha]]\,Bel_i\psi \wedge \neg\,[[i,\alpha]]\,\bot \longrightarrow Bel_i\,[[j,\alpha]]\,\psi$
R3. $\frac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)	12b. $Bel_i(GBel_i\psi \longleftrightarrow Bel_iG\psi)$
R4. $\frac{\vdash\varphi}{\vdash Bel\varphi}$ (*Bel-necessitation*)	13b. $Goal_i\,\langle\langle i,\alpha\rangle\rangle\,\top \longleftrightarrow \langle\langle i,\alpha\rangle\rangle\,\top$
R5. $\frac{\vdash\varphi}{\vdash Goal\varphi}$ (*Goal-necessitation*)	14b. $[[i,\alpha]]\,\varphi \wedge [[i,\alpha]]\,(\varphi \longrightarrow \psi) \longrightarrow [[i,\alpha]]\,\psi$
	15b. $X\varphi \longrightarrow [[i,\alpha]]\,\varphi$

Axiom of intentional attempt (1)

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha \rangle\rangle \top$

-
-
-

Axiom of intentional attempt (1)

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha \rangle\rangle \top$

Semantic correspondence:

if $R^{att}_{i:\alpha}(w) = \emptyset$ then $\exists\ w'$ such that $w' \in G_i(w)$ and $R^{att}_{i:\alpha}(w') = \emptyset$

-
-
-

Axiom of intentional attempt (1)

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha \rangle\rangle \top$

Semantic correspondence:

if $R^{att}_{i:\alpha}(w) = \emptyset$ then $\exists\ w'$ such that $w' \in G_i(w)$ and $R^{att}_{i:\alpha}(w') = \emptyset$

-

- It establishes that an agent attempts to do some action α if and only if the agent has the goal to attempt to do action α.

-

Axiom of intentional attempt (1)

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha \rangle\rangle \top$

Semantic correspondence:

if $R^{att}_{i:\alpha}(w) = \emptyset$ then $\exists\ w'$ such that $w' \in G_i(w)$ and $R^{att}_{i:\alpha}(w') = \emptyset$

- Never analyzed before in modal logic of intentional action.

- It establishes that an agent attempts to do some action α if and only if the agent has the goal to attempt to do action α.

-

Axiom of intentional attempt (1)

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha \rangle\rangle \top$

Semantic correspondence:

if $R^{att}_{i:\alpha}(w) = \emptyset$ then $\exists\ w'$ such that $w' \in G_i(w)$ and $R^{att}_{i:\alpha}(w') = \emptyset$

- Never analyzed before in modal logic of intentional action.
- It establishes that an agent attempts to do some action α if and only if the agent has the goal to attempt to do action α.
- It relates mental side with the executive and behavioral side.

Axiom of intentional attempt (2)

$Goal_i \langle\langle i, \alpha; \beta \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha; \beta \rangle\rangle \top$

Applicable to sequences of basic actions α;β;... performed by the same agent and not involving perception (epistemic actions).

Axiom of intentional attempt (2)

$Goal_i \langle\langle i, \alpha; \beta \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha; \beta \rangle\rangle \top$

Applicable to sequences of basic actions $\alpha;\beta;\ldots$ performed by the same agent and not involving perception (epistemic actions).

An agent cannot stop in the middle of a pre-planned sequence and revise his pushing intentions (unless he does some epistemic action)

Axiom of intentional attempt (2)

$Goal_i \langle\langle i, \alpha; \beta \rangle\rangle \top \leftrightarrow \langle\langle i, \alpha; \beta \rangle\rangle \top$

Applicable to sequences of basic actions α;β;... performed by the same agent and not involving perception (epistemic actions).

An agent cannot stop in the middle of a pre-planned sequence and revise his pushing intentions (unless he does some epistemic action)

A football playing robot having the goal to perform the sequence turn-right;advance;shoot.

Even if the robot is blocked by another player, he will attempt to execute the three basic actions in sequence

Definition of Action (1)

DEFINITION 1: EXECUTION PRECONDITIONS.

$Pre: ACT \rightarrow PROP.$

.

Definition of Action (1)

DEFINITION 1: EXECUTION PRECONDITIONS.

$Pre: ACT \rightarrow PROP$.

For example we might have: $FreeLeg = Pre(kickBall)$.

Definition of Action (1)

DEFINITION 1: EXECUTION PRECONDITIONS.

$Pre: ACT \rightarrow PROP$.

For example we might have: $FreeLeg = Pre(kickBall)$.

DEFINITION 2: ACTION. $\langle i, \alpha \rangle \varphi =_{def} \langle\langle i, \alpha \rangle\rangle \varphi \wedge Pre(\alpha)$.

Definition of Action (1)

DEFINITION 1: EXECUTION PRECONDITIONS.

$Pre: \ ACT \rightarrow PROP$.

For example we might have: $FreeLeg = Pre(kickBall)$.

DEFINITION 2: ACTION. $\langle i, \alpha \rangle \, \varphi =_{def} \langle\langle i, \alpha \rangle\rangle \, \varphi \wedge Pre(\alpha)$.

Action executions are attempts whose execution preconditions hold.

ACTION FAILURE: $\langle\langle i, \alpha \rangle\rangle \top \wedge \neg Pre(\alpha)$

ACTION SUCCESS: $\langle\langle i, \alpha \rangle\rangle \top \wedge Pre(\alpha) =_{def} \langle i, \alpha \rangle \top$

Definition of Action (2)

- $[[i, \alpha]]\, \varphi \rightarrow [i, \alpha]\, \varphi$.
-
- A consequence of an attempt to perform α is also a consequence of the successful performance of basic action α.
-

Definition of Action (2)

-
- $Pre(\alpha) \rightarrow ([[i, \alpha]]\, \varphi \leftrightarrow [i, \alpha]\, \varphi)$.
-
- if the *execution preconditions* hold then the consequences of the attempt are equivalent to the consequences of the associated basic action.

Formulating Action theories (1)

Execution preconditions for the three actions *loading*, *pulling* and *picking-up*:

$Pre(pull) = freeHand$,

$Pre(load) = freeHand$,

$Pre(pickUp) = freeArm \wedge freeHand$.

Formulating Action theories (2)

For each propositional atom $p \in \Pi$ and basic action $\alpha \in ACT$:

- a propositional formula $\gamma^{+}(\alpha, p)$ describing the *positive effect preconditions* of the attempt to do α with respect to p;
-

Formulating Action theories (2)

For each propositional atom $p \in \Pi$ and basic action $\alpha \in ACT$:

- a propositional formula $\gamma^+(\alpha, p)$ describing the *positive effect preconditions* of the attempt to do α with respect to p;
- a prop. formula $\gamma^-(\alpha, p)$ describing the *negative effect preconditions* of the attempt to do α with respect to p.

Formulating Action theories (3)

$$\gamma^{+}(load, loadedGun) = freeHand \wedge holdsGun$$
$$\gamma^{+}(pull, wounded) = holdsGun \wedge loadedGun \wedge pointedGun \wedge freeHand$$
$$\gamma^{+}(pull, pulledTrigger) = holdsGun \wedge freeHand$$
$$\gamma^{+}(pull, scared) = holdsGun \wedge pointedGun$$
$$\gamma^{+}(pickUp, holdsGun) = gunOnTable \wedge freeArm \wedge freeHand$$

We suppose that for each act α and possible effect p: $\gamma^{-}(\alpha, p) = \bot$

Fitting Global Assumptions (effect laws):

$$holdsGun \wedge pointedGun \to [[pull]]\, scared$$
$$holdsGun \wedge loadedGun \wedge pointedGun \wedge freeHand \to [[pull]]\, wounded$$
$$freeHand \wedge holdsGun \to [[load]]\, loadedGun$$
$$holdsGun \wedge freeHand \to [[pull]]\, pulledTrigger$$
$$gunOnTable \wedge freeArm \wedge freeHand \to [[pickUp]]\, holdsGun$$

Formulating Action theories (4)

We suppose **Completeness** of effect laws.

Global assumptions:

$\neg\gamma^{+}(\alpha, p) \wedge \neg p \rightarrow [[\alpha]]\, \neg p$

$\neg\gamma^{-}(\alpha, p) \wedge p \rightarrow [[\alpha]]\, p$

Formulating Action theories (4)

We suppose **Completeness** of effect laws.

Global assumptions:

$\neg\gamma^{+}(\alpha, p) \wedge \neg p \rightarrow [[\alpha]]\,\neg p$

$\neg\gamma^{-}(\alpha, p) \wedge p \rightarrow [[\alpha]]\,p$

Given the effect law $holdsGun \wedge pointedGun \rightarrow [[pull]]\,scared$ for the action *pulling*, we suppose that:

$\neg(holdsGun \wedge pointedGun) \wedge \neg scared \rightarrow [[pull]]\,\neg scared.$

Formulating Action theories (5)

We suppose **Consistency**.

$\gamma^{+}(\alpha, p) \rightarrow \neg\gamma^{-}(\alpha, p)$.

Formulating Action theories (5)

We suppose **Consistency**.

$\gamma^{+}(\alpha, p) \rightarrow \neg\gamma^{-}(\alpha, p)$.

Formulating Action theories (6)

Suppose that $\gamma^{-}(\alpha, p)$, $\gamma^{+}(\alpha, p)$ are given and that the *completeness assumption* and *consistency assumption* are made then the following equivalences holds:

$$[[i, \alpha]]\, p \leftrightarrow \neg Goal_i \,\langle\langle i, \alpha \rangle\rangle \top \vee \gamma^{+}(\alpha, p) \vee (p \wedge \neg \gamma^{-}(\alpha, p))$$

-

-

Formulating Action theories (6)

Suppose that $\gamma^{-}(\alpha, p)$, $\gamma^{+}(\alpha, p)$ are given and that the *completeness assumption* and *consistency assumption* are made then the following equivalences holds:

$$[[i, \alpha]]\, p \leftrightarrow \neg Goal_i \,\langle\langle i, \alpha\rangle\rangle \top \vee \gamma^{+}(\alpha, p) \vee (p \wedge \neg\gamma^{-}(\alpha, p))$$

- Every planning task can in principle be reduced to the task of finding the correct sequence of attempts for reaching a given result.

- No need to verify whether execution preconditions hold.

Stable vs successful effects (1)

A STABLE POSITIVE EFFECT of an attempt to do some action α is a result that an attempt to perform α can produce even if the execution preconditions of action α do not hold.

Stable vs successful effects (1)

A STABLE POSITIVE EFFECT of an attempt to do some action α is a result that an attempt to perform α can produce even if the execution preconditions of action α do not hold.

$Pre(pull) = freeHand$.

$holdsGun \wedge pointedGun \rightarrow [[pull]]\, scared$.

Scared is a *STABLE POSITIVE EFFECT* of the attempt to *pull*!!!

I can scare you simply by pointing a gun toward you and attempting to pull the trigger!!!

Stable vs successful effects (2)

A SUCCESSFUL POSITIVE EFFECT of an attempt to do some action α is a result that an attempt to perform α may produce only if the execution preconditions of action α hold.

Stable vs successful effects (2)

A <u>SUCCESSFUL POSITIVE EFFECT</u> of an attempt to do some action α is a result that an attempt to perform α may produce only if the execution preconditions of action α hold.

$Pre(pull) = freeHand$.

$holdsGun \wedge loadedGun \wedge pointedGun \wedge freeHand \rightarrow [[pull]]\, wounded$

Wounded is a <u>SUCCESSFUL POSITIVE EFFECT</u> of the attempt to *pull*!!!

I can wound you only if after pointing the gun toward you and attempting to pull the trigger, I correctly execute the pulling movement and the gun is loaded!!!

Stable vs successful effects (3)

- *p* is a *successful positive effect* of the attempt to perform the basic action α if and only if

  $$\models_{LIA} \gamma^{+}(\alpha, p) \rightarrow Pre(\alpha).$$

-

Stable vs successful effects (3)

- *p* is a *successful positive effect* of the attempt to perform the basic action α if and only if

  $\models_{LIA} \gamma^{+}(\alpha, p) \rightarrow Pre(\alpha)$.

- *p* is a *stable positive effect* of the attempt to perform the basic action α if and only if

  there is a model $M \in LIA$ such that $\gamma^{+}(\alpha, p) \wedge \neg Pre(\alpha)$ is *satisfiable* in M.

Stable vs successful effects (4)

- $\neg p$ is a *successful negative effect* of a basic action α if and only if $\models_{LIA} \gamma^{-}(\alpha, p) \rightarrow Pre(\alpha)$

-

Stable vs successful effects (4)

- $\neg p$ is a *successful negative effect* of a basic action α if and only if $\models_{LIA} \gamma^{-}(\alpha, p) \rightarrow Pre(\alpha)$

- $\neg p$ is a *stable negative effect* of a basic action α if and only if

  it exists a model $M \in LIA$ such that $\gamma^{-}(\alpha, p) \wedge \neg Pre(\alpha)$

Stable vs successful effects (5)

- *p* is a *intrinsic effect* of some basic action α if and only if $\models_{LIA} \gamma^{+}(\alpha, p) \leftrightarrow Pre(\alpha)$.

Stable vs successful effects (5)

- *p* is a *intrinsic effect* of some basic action α if and only if

  $\models_{LIA} \gamma^{+}(\alpha, p) \leftrightarrow Pre(\alpha)$.

"The *intrinsic effect* of some (basic) action α is the state of affairs that it is guaranteed to hold when α is attempted and the execution preconditions of action α hold" (Von wright, 1963 and Stoutland, 1968)

The intrinsic effect of the (basic) action of *raising the arm* is *raised arm*.

Some theorems (1)

Attempt awareness

- $\langle\langle i, \alpha \rangle\rangle \top \leftrightarrow Bel_i \langle\langle i, \alpha \rangle\rangle \top$
- $[[i, \alpha]] \bot \leftrightarrow Bel_i [[i, \alpha]] \bot$

Some theorems (2)

DEFINITION 3: INTENTION IN ACT

(Searle 1983, Bratman 1987).

$PDI_i(\alpha) =_{def} Goal_i \langle i, \alpha \rangle \top$

-
-
-
-

Some theorems (2)

DEFINITION 3: INTENTION IN ACT

(Searle 1983, Bratman 1987).

$PDI_i(\alpha) =_{def} Goal_i \langle i, \alpha \rangle \top$

- $PDI_i(\alpha) \rightarrow \langle\langle i, \alpha \rangle\rangle \top$;
- $PDI_i(\alpha) \rightarrow Goal_i \langle\langle i, \alpha \rangle\rangle \top$;
- $PDI_i(\alpha) \rightarrow Goal_i Pre(\alpha)$;
- $PDI_i(\alpha) \rightarrow \neg Bel_i \neg Pre(\alpha)$.

Additional properties

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \rightarrow PDI_i(\alpha)$ is not valid

Additional properties

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \rightarrow PDI_i(\alpha)$ is not valid

“Brett promises to pay Belton fifty dollars if Belton *attempts* to solve a certain chess problem within five minutes .Brett assures Belton that he need not actually solve the problem for getting the fifty dollars."

Additional properties

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \rightarrow PDI_i(\alpha)$ is not valid

“Brett promises to pay Belton fifty dollars if Belton *attempts* to solve a certain chess problem within five minutes .Brett assures Belton that he need not actually solve the problem for getting the fifty dollars."

According to (Mele, 1992) Belton is motivated to attempt to solve problem even if he does not intend to solve the problem.

Additional properties

$Goal_i \langle\langle i, \alpha \rangle\rangle \top \rightarrow PDI_i(\alpha)$ is not valid

“Brett promises to pay Belton fifty dollars if Belton *attempts* to solve a certain chess problem within five minutes .Brett assures Belton that he need not actually solve the problem for getting the fifty dollars."

According to (Mele, 1992) Belton is motivated to attempt to solve problem even if he does not intend to solve the problem.

It is possible "intending to try/attempt to do A" without "intending to do A" (see Bratman, 1987; McCann, 1986).

Other issues

- Definition of Future-directed Intentions (FDI).

- Generation of Intentions (based on Practical Inference) and Persistence of Intentions.

- The doxastic version of "Trying" and "Attempt".

  "Trying to do A is, roughly, doing one thing which one thinks likely in the circumstances to grow into a doing of A" (Sellars, 1967).

$$Doubt_i\varphi =_{def} \neg Bel_i\varphi \wedge \neg Bel_i\neg\varphi.$$

$$Trying_i(\alpha) =_{def} Goal_i \langle i, \alpha \rangle \top \wedge Doubt_i \langle i, \alpha \rangle \top$$

Thanks!!!