Modal Logic and BDI Logic: an Overview

Stefania Costantini, University of L'Aquila

Part of this material is a re-elaboration of material kindly provided by colleagues.
In particular, I thank Alessandro Agostini, Fausto Giunchiglia, Rui Zhang, Vincenzo Maltese,
Iyad Rahwan, Adina Magda Florea

Introduction

- We want to model situations like this one:
- 1. "Fausto is always happy" (necessarily happy)
- 2. "Fausto is happy under certain circumstances" (possibly happy)

- In Propositional Logic/Classical Logic we could have: HappyFausto

- In modal logic we have:
- 1. $\square$ HappyFausto
- 2. $\Diamond$ HappyFausto

- As we will see, this is captured through the notion of "possible words" and of "accessibility relation"

Syntax

- We extend propositional first-order logic (PL) with two logical *modal* operators:

  □ (box) and ◊ (diamond)

  □P : “Box P” or “necessarily P” or “P is necessary true”

  ◊P : “Diamond P” or “possibly P” or “P is possible”

  Note that we define □P = ¬◊¬P, i.e. ◊ is a primitive symbol

Syntax

- The grammar of well-formed formulas (wwf) of underlying logic language L is extended as follows:

falso

vero

<Atomic Formula> ::= A | B | ... | P | Q | ... | ⊥ | ⊤ |

<wff> ::= <Atomic Formula> | ¬<wff> | <wff>∧ <wff> | <wff>∨ <wff> |

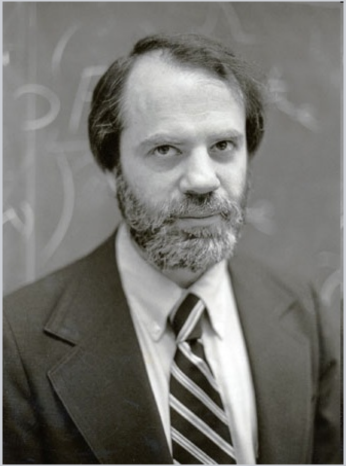
<wff> ▯ <wff> | <wff> ▯ <wff> | □ <wff> | ◊ <wff>

Different interpretations

Philosophy	□P : "P is necessary" ◊P : "P is possible"
Epistemic	□aP : "Agent a believes P " or "Agent a knows P"
Temporal logics	□P : "P is always true" ◊P : "P is sometimes true"

Modal Logic: Semantics

- Semantics is given in terms of Kripke Models (also known as Kripke Sructures possible worlds structures)
- Due to American logician Saul Kripke, City University of NY (result dating back to 1959)

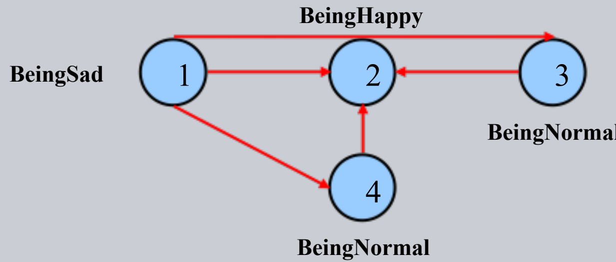


Semantics: Kripke Models

- A Kripke Model is a triple M = <W, R, I> where:
- W is a non empty set of worlds
- R ⊆ W x W is a binary relation called the accessibility relation
- I is an interpretation function I: L → pow(W) such that to each proposition P we associate a set of possible worlds I(P) in which P holds

Semantics: Kripke Models

- Consider the following situation:



- M = <W, R, I>

  W = {1, 2, 3, 4}

  R = {<1, 2>, <1, 3>, <1, 4>, <3, 2>, <4, 2>}

  I(BeingHappy) = {2}   I(BeingSad) = {1}   I(BeingNormal) = {3, 4}

- Each w ∊ W is said to be a world, point, state, event, situation, class … according to the problem we model

Truth relation (true in a world)

- Given a Kripke Model $M = \langle W, R, I \rangle$, a proposition P and a possible world $w \in W$, we say that "w satisfies P in M" or that "P is satisfied by w in M" or "P is true in M via w", in symbols:

  $M, w \vDash P$ in the following cases:

  1. P atomic $\quad w \in I(P)$
  2. $P = \neg Q \quad M, w \nvDash Q$
  3. $P = Q \wedge T \quad M, w \vDash Q$ and $M, w \vDash T$
  4. $P = Q \vee T \quad M, w \vDash Q$ or $M, w \vDash T$
  5. $P = Q \supset T \quad M, w \nvDash Q$ or $M, w \vDash T$
  6. $P = \Box Q \quad$ for every $w' \in W$ such that wRw' then $M, w' \vDash Q$
  7. $P = \Diamond Q \quad$ for some $w' \in W$ such that wRw' then $M, w' \vDash Q$

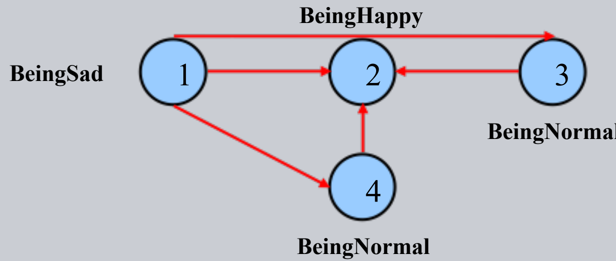
  NOTE: wRw' can be read as "w' is accessible from w via R" or that "w' is reachable from w via R"

Semantics: Kripke Models

- Each $w \in W$ is said to be a world, point, state, event, situation, class … according to the problem we model
- For "world" we mean a Propositional Logic model (a set of atoms). Focusing on this definition, we can see a Kripke Model as a set of different PL models related by an "evolutionary" relation R; in such a way we are able to represent formally - for example - the evolution of a model in time.
- In a Kripke model, $\langle W, R \rangle$ is called frame and is a relational structure.

Semantics: Kripke Model

- Consider the following situation:



- $M = \langle W, R, I \rangle$

  $W = \{1, 2, 3, 4\}$

  $R = \{\langle 1, 2\rangle, \langle 1, 3\rangle, \langle 1, 4\rangle, \langle 3, 2\rangle, \langle 4, 2\rangle\}$

  $I(\text{BeingHappy}) = \{2\}$ $I(\text{BeingSad}) = \{1\}$ $I(\text{BeingNormal}) = \{3, 4\}$

$M, 2 \vDash \text{BeingHappy}$ $M, 2 \vDash \neg\text{BeingSad}$

$M, 4 \vDash \neg\text{BeingHappy}$ $M, 1 \vDash \Diamond\text{BeingHappy}$ $M, 1 \vDash \neg\Diamond\text{BeingSad}$

Satisfiability and Validity

- Satisfiability

A proposition $P \in L$ is satisfiable in a Kripke model $M = \langle W, R, I \rangle$ if

$M, w \models P$ for all worlds $w \in W$. // P vale in tutti i mondi del grafo

We can then write $M \models P$ // il modello implica P indipendentemente dal mondo

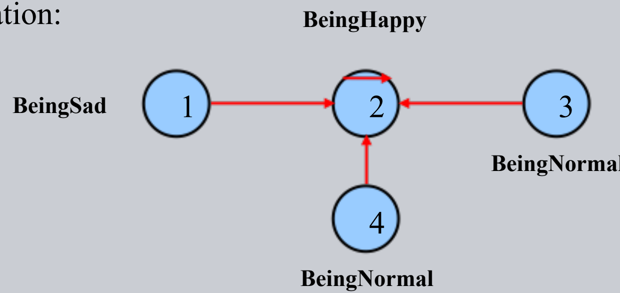
- Validity

A proposition $P \in L$ is valid if P is satisfiable for all models M (and by varying the frame $\langle W, R \rangle$).

We can write $\models P$

Satisfiability

- Consider the following situation:



- M = <W, R, I>

  W = {1, 2, 3, 4}

  R = {<1, 2>, <2, 2>, <3, 2>, <4, 2>}

  I(BeingHappy) = {2} I(BeingSad) = {1} I(BeingNormal) = {3, 4}

M, w ⊨ □BeingHappy for all w ∈ W, therefore □BeingHappy is satisfiable in M.

Axiomatic theory of Knowledge

- The modal operator □*i* becomes *Ki*

- Worlds accessible from *w* according to accessibility relation *Ri* are those indistinguishable to agent *i* from world *w*
- *Ki* means “agent *i* knows that”
- Start with the simple axioms:
- **(Classical)** All propositional tautologies are valid
- **(Modus Ponens)** if □ and □□□ are valid, infer that □ is valid

Axiomatic theory of Knowledge (More Axioms)

- **(K)** From $(Ki\varphi \wedge Ki(\varphi \to \psi))$ infer $Ki\psi$
- Means that the agent knows all the consequences of his knowledge
- This is also known as logical omniscience

- **(Necessitation)** From φ, infer that $Ki\varphi$
- Means that the agent knows all propositional tautologies

Axiomatic theory of Knowledge (More Axioms)

D "UN AGENTE NON SA"

D "UN AGENTE SA"

- **Axiom (D)** $\neg K_i(\varphi \wedge \neg\varphi)$
- This is called the axiom of consistency
- **Axiom (T)** $(K_i\varphi) \rightarrow \varphi$
- This is called the veridity axiom
- Means that if an agent cannot know something that is not true.

Axiomatic theory of Knowledge (More Axioms)

- **Axiom (4)** $Ki\ \varphi \rightarrow Ki\ Ki\ \varphi$
- Called the positive introspection axiom

- **Axiom (5)** $\neg Ki\ \varphi \rightarrow Ki\ \neg Ki\ \varphi$
- Called the negative introspection axiom

Axiomatic theory of Knowledge (Overview of Axioms, Modal Logic S5)

Name	Axiom	Accessibility Relation
Axiom K	$(K_i(\varphi) \wedge K_i(\varphi \rightarrow \psi)) \rightarrow K_i(\psi)$	NA
Axiom D	$\neg K_i(p \wedge \neg p)$	Serial
Axiom T	$K_i\varphi \rightarrow \varphi$	Reflexive
Axiom 4	$K_i\varphi \rightarrow K_iK_i\varphi$	Transitive
Axiom 5	$\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$	Euclidean

Table 13.1: Axioms and corresponding constraints on the accessibility relation.

5. Logics of knowledge and belief

- Used to model "modes of truth" of cognitive agents
- Distributed modalities
- Cognitive agents ➔ characterise an intelligent agent using symbolic representations and **mentalistic notions**:
- knowledge - John knows humans are mortal
- beliefs - John took his umbrella because he believed it was going to rain
- desires, goals - John wants to possess a PhD
- intentions - John intends to work hard in order to have a PhD
- commitments - John will not stop working until getting his PhD

Logics of knowledge and belief

- How to represent knowledge and beliefs of agents?
- FOPL (First-Order Predicate Logic) augmented with two modal operators **K** and **B**

  K(a,□) - **a** knows □

  B(a,□) - **a** believes □

  with □□LFOPL, a□A, set of agents

- Associate with each agent a set of possible worlds
- Kripke model Ma of agent *a* for a formula □
- Ma =<W, R, I>

  with R □ W X W

  and I - interpretation of the formula on a Kripke frame <W,R> which makes the formula true for agent *a*

Logics of knowledge and belief: Observations

- An agent *knows* a propositions in a given world if the proposition holds in all worlds accessible to the agent from the given world

$$Ma \models_W \mathbf{K}\varphi \quad \text{iff} \quad \forall w'\colon\ R(w,w') \rightarrow Ma \models_{W'} \varphi$$

- An agent *believes* a propositions in a given world if the proposition holds in all worlds accessible to the agent from the given world

$$Ma \models_W \mathbf{B}\varphi \quad \text{iff} \quad \forall w'\colon\ R(w,w') \rightarrow Ma \models_{W'} \varphi$$

- The difference between **B** and **K** is given by their properties (sometimes the difference is neglected)

Properties of knowledge

(A1) Distribution axiom:

K(a, □) □ **K**(a, □ □ □) □ **K**(a, □)

"The agent ought to be able to reason with its knowledge"

□(□ □□) □ (□□ □ □□) (Axiom of distribution of modality)

K(a,□ □□) □ (K(a,□) □ K(a,□))

(A2) Knowledge axiom: **K**(a, □) □ □

"The agent can not know something that is false"

□ □ □ □ (T)

K(a, □) □ □

Properties of knowledge

(A3) Positive introspection axiom

K(a, □) □ **K**(a, **K**(a, □))

□X □ □□X (S4)

K(a, □) □ K(a, K(a, □))

(A4) Negative introspection axiom

□K(a, □) □ K(a, □K(a, □))

□X □ □□ X (S5)

Inference rules for knowledge

(R1) Epistemic necessitation

|- □ □ **K**(a, □)

modal rule of necessity |- □ □ □□

(R2) Logical omniscience

□ □ □ and **K**(a, □) □ **K**(a, □)

Properties of belief

Distribution axiom: **B**(a, □) □ **B**(a, □ □ □) □ **B**(a, □)

YES

Knowledge axiom: **B**(a, □) □ □ NO

Positive introspection axiom

B(a, □) □ B(a, B(a, □))

YES

Negative introspection axiom

□B(a, □) □ B(a, □B(a, □)) problematic

Inference rules for belief

(R1) Epistemic necessitation

|- □ □ **B**(a, □) *problematic*

modal rule of necessity |- □ □ □□

(R2) Logical omniscience

□ □ □ and **B**(a, □) □ **B**(a, □)

usually NO

Some more axioms for beliefs

Knowing what you believe

$$\mathbf{B}(a, \square) \;\; \square \;\; \mathbf{K}(a, \mathbf{B}(a, \square))$$

Believing what you know

$$\mathbf{K}(a, \square) \;\; \square \;\; \mathbf{B}(a, \square)$$

Have confidence in the belief of another agent

$$\mathbf{B}(a1, \mathbf{B}(a2, \square)) \;\; \square \;\; \mathbf{B}(a1, \square)$$

Using modal reasoning in practice

At the Pub

- Three Logicians enter a pub.
- The barman asks: “Three beers, as usual?”
- The first logician says: “I do not know!”
- The second logician then says “I also do not know!”
- The third logician says “Yes!”

Q: How did they reason?

The King's Three Wise Men Puzzle

- The King called the three wisest men in the country.
- He painted a spot on each of their foreheads and told them that at least one of them has a white spot on his forehead. Then he asked them in turn: “Do you know if you have a white spot on your forehead?”
- The first wise man said: “I do not know whether I have a white spot”.
- The second man then says “I also do not know whether I have a white spot”.
- The third man says then “I know I have a white spot on my forehead”.

Q: How did the third wise man reason?

Two-wise men problem - Genesereth, Nilsson

Variation of the King's Wise Men Puzzle

- The King called the two wisest men in the country.
- He painted a spot on one of their foreheads and told them that at least one of them has a white spot on his forehead.
- The first wise man said: "I do not know whether I have a white spot"
- The second wise man said then "I know I have a white spot on my forehead".

Q: How did the second wise man reason?

(1) We call the first wiseman "B" and the second wiseman "A"

(2) Ther spot is on "B"'s forehead

(3) A and B know that each can see the other's forehead. Thus:

(1a') If A does not have a white spot, B will know that A does not have a white spot

(1a'') A knows (1a')

(1b') If B does not have a white spot, A will know that B does not have a white spot

(1b'') B knows (1b')

(2) A and B each know that at least one of them have a white spot, and they each know that the other knows that. In particular

(2a) A knows that B knows that either A or B has a white spot (and so does B w.r.t. A)

(3) B says that he does not know whether he has a white spot, and A thereby knows that B does not know

Two-wise men problem - Genesereth, Nilsson

(1) A and B know that each can see the other's forehead. (2) A and B each know that at least one of them have a white spot, and they each know that the other knows that. In particular

(2a) A knows that B knows that either A or B has a white spot

(3) B says that he does not know whether he has a white spot, and A thereby knows that B does not know

Proof of why A answers YES from B saying “I don’t’ know”

1. **K**A(□WA □ **K**B(□ WA) Axiom 1
2. **K**A(**K**B(WA □ WB)) Axiom 2
3. **K**A(□**K**B(WB)) Axiom 3

Proof

4. □WA □ **K**B(□WA) 1, A2 A2: **K**(a, □) □ □
5. **K**B(□WA □ WB) 2, A2

6. **K**B(□WA) □ **K**B(WB) 5, A1 A1: K(a,□ □□) □ (K(a,□) □ K(a,□))
7. □WA □ **K**B(WB) 4, 6

8. □**K**B(WB) □ WA contrapositive of 7
9. **K**A(WA) answer Yes!!! 3, 8, R2

R2: □ □ □ and **K**(a, □) *infer* **K**(a, □)

6. Temporal logic

- The time may be linear or branching; the branching can be in the past, in the future of both
- Time is viewed as a set of moments with a strict partial order, $<$, which denotes temporal precedence.
- Every moment is associated with a possible state of the world, identified by the propositions that hold at that moment

Modal operators of temporal logic (linear) LTL

p **U** q - **p** is true until **q** becomes true - until

Xp - **p** is true in the next moment - next

Pp - **p** was true in a past moment - past

Fp - **p** will finally (eventually) be true in the future - eventually

Gp - **p** will always be true in the future – always

$$\mathbf{F}p \equiv \text{true } \mathbf{U}\, p$$

$$\mathbf{G}p \equiv \neg \mathbf{F} \neg p$$

F – one time point

G – each time point

Branching time logic - CTL

- Temporal structure with a branching time future and a single past - **time tree T**
- **CTL – Computational Tree Logic**
- In a branching logic of time, a path at a given moment is any maximal set of moments containing the given moment and all the moments in the future along some particular branch of **T (so, a path is actually a subtree)**

* **Situation** - a world **w** at a particular time point ***t***, ***wt***

- ***State formulas*** - evaluated at a specific *time point* in a time tree
- ***Path formulas*** - evaluated over a *specific path* in a time tree

Branching time logic - CTL

CTL Modal operators over both state and path formulas

From Temporal logic (linear)

Fp - **p** will sometime be true in the future - eventually

Gp - **p** will always be true in the future - always

Xp - **p** is true in the next moment - next

p **U** q - **p** is true until **q** becomes true - until

(**p** holds on a path ***s*** starting in the current moment ***t*** until **q** comes true)

F – one time point
G – each time point

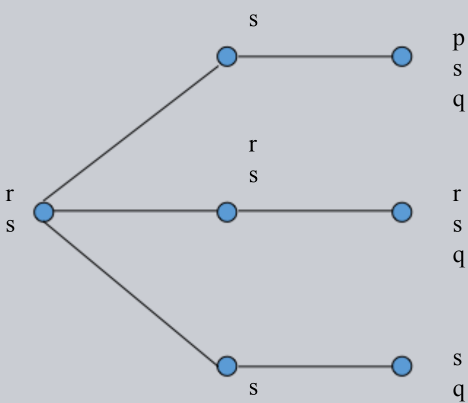
Modal operators over path formulas (branching)

Ap - at a particular time moment, **p** is true in all paths emanating from that point - inevitable p

Ep - at a particular time moment, **p** is true in some path emanating from that point - optional p

A – all path
E – some path

- **s** is true in each time point (G) and in all path (A)
- **r** is true in each time point (G) in some path (E)
- **p** will eventually (F) be true in some path (E)
- **q** will eventually (F) be true in all path (A)



AGs

EGr

EFp

AFq

F - eventually
G - always
A - inevitable
E - optional

r - Alice is in Italy
p -Alice visits Paris
s – Paris is the capital of France
q - It is spring time

6 BDI logic

Modal operators **Bel**, **Des**, **Int**, **(KW)**

M=<W, T, R, **B**, **D**, **I**> where T = time (under Linear-Time Temporal Logic)

B: belief accessibility relation for operator Bel

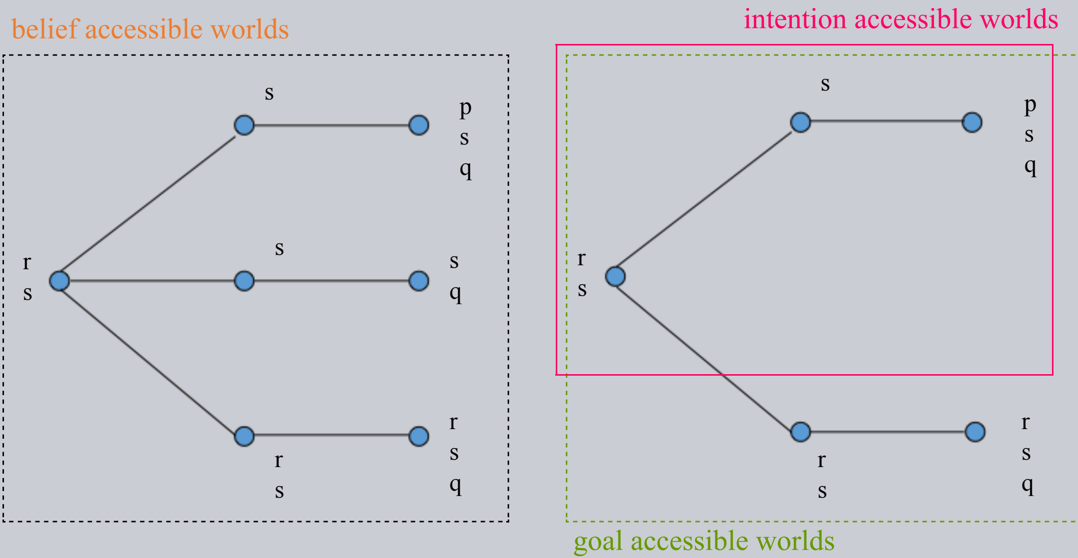
belief accessible worlds; the worlds the agent believes possible

D: desire (goal) accessibility relation for operator Des

- Requires the desires to be consistent; therefore Desires □ Goals
- Each situation has associated a set of goal -accessible worlds - **realism**
- **Strong realism** = for each belief-accessible world **w** at a given time moment **t**, there must be a goal-accessible world that is a sub-world of **w** at time **t**

I : intention accessibility relation for operator Int

- Intentions - similarly represented by sets of intention-accessible worlds. These are the worlds the agent has committed to realize.
- Corresponding to each intention-accessible world at some time **t** there must be a goal-accessible world: however, the agent may choose only a subset of goal-accessible worlds



r - Alice is in Italy p -Alice visits Paris

s – Paris is the capital of France q - it is spring time

Assiomi Bel, Des, Int

$M \models_t x\mathbf{Bel}p$ iff $(\forall t' : (t,t') \in B(x,t) \Rightarrow M \models_{t'} p)$

an agent ***x*** has a belief ***p*** in a given moment ***t*** if and only if ***p*** is true in all belief accessible worlds of the agent in that moment

$M \models_t x\mathbf{Des}p$ iff $(\forall t' : (t,t') \in D(x,t) \Rightarrow M \models_{t'} p)$

an agent ***x*** has a desire ***p*** in a given moment ***t*** if and only if ***p*** is true in all goal accessible worlds of the agent in that moment

$M \models_t x\mathbf{Int}p$ iff $(\forall s : s \in I(x,t) \Rightarrow M \models_{s,t} \mathbf{F}p)$

at each moment ***t***, **I** assigns a set of paths that the agent **x** has selected or preferred, i.e., if the agent has selected ***p*** as an intention, ***p*** will hold eventually in the future

F - eventually

G - always

A - inevitable

E - optional

Relationship among Modal Operators

Belief-goal compatibility

If an agent adopts **p** as a goal, then the agent believes that there exists (at least) one path on which **p** will be true as it is an adopted desire but it needs not believe that it will ever reach that point

x**Des**p $\Rightarrow$ (x**Bel** (**E G** p)

Goal-intention compatibility

If an agent adopts **p** as an intention, it should have adopted it as a goal to be achieved

x**Int**p $\Rightarrow$ x**Des**p

F - eventually
G - always
A - inevitable
E - optional

Relationship among Modal Operators

Beliefs about intentions

$$x\mathbf{Int}p \Rightarrow x\mathbf{Bel}(x\mathbf{Int}p))$$

No infinite deferral

The agent should not procrastinate with respect to its intentions; if the agent forms an intention, then sometimes in the future it will give up this intention

$$x\mathbf{Int}p \Rightarrow \mathbf{A}\ \mathbf{F}(\neg x\mathbf{Int}p))$$

F - eventually
G - always
A - inevitable
E - optional

Commitment

- An agent is considered as being committed to its intention but, cf. no infinite deferral, it will give up these intentions eventually - when?
- Different types of agents will have different commitment strategies.

F - eventually
G - always
A - inevitable
E - optional

Blindly committed agent

- o maintains its intentions until it believes it has achieved them

$$x\mathbf{Int}(\mathbf{A}\ \mathbf{F}p)\ \square\mathbf{A}\ (x\mathbf{Int}(\mathbf{A}\ \mathbf{F}p)\ \square\ x\mathbf{Bel}p) \quad \text{(exclusive } \square\text{)}$$

- o an agent can be committed to means (**p** is an action) or to ends (**p** is a formula)
- o defined only for intentions toward actions or conditions that are true for all paths in the agent's intention accessible worlds.

Single-minded committed agent

- maintains its intentions as long as it believes they are still options

$$x\mathbf{Int}(\mathbf{A}\ \mathbf{F}p)\ \square\mathbf{A}\ (x\mathbf{Int}(\mathbf{A}\ \mathbf{F}p)\ \square\ (x\mathbf{Bel}p\ \square\ \square x\mathbf{Bel}(\mathbf{E}\ \mathbf{F}p)))$$

Open-minded committed agent

- maintains its intentions as long as these intentions are still its desires (goals)

$$x\mathbf{Int}(\mathbf{A}\ \mathbf{F}p)\ \square\mathbf{A}\ (x\mathbf{Int}(\mathbf{A}\ \mathbf{F}p)\ \square\ (x\mathbf{Bel}p\ \square\ \square x\mathbf{Des}(\mathbf{E}\ \mathbf{F}p)))$$

F - eventually
G - always
A - inevitable
E - optional

References

- M. P. Singh, A.S. Rao. Formal methods in DAI: Logic-based representation and reasoning. In Multiagent Systems - A Modern Approach to Distributed Artficial Intelligence, G. Weiss (Ed.), The MIT Press, 2001, p.331-355.
- M. Wooldrige. Reasoning about Rational Agents. The MIT Press, 2000, Chapter 2
- A.S. Rao, M.P. Georgeff. Modeling rational agents within a BDI-architecture. In Readings in Agents, M. Huhns & M. Singh (Eds.), Morgan Kaufmann, 1998, p.317-328.
- M.R. Genesereth, N.J. Nilsson. Logical Foundations of Artificial Intelligence. Morgan Kaufmann, 1987, Chapter 9.
- D. Kayser: La représentation des connaissances. Hermès, 1997.
- J.Y. Halpern. Reasoning about knowledge: A survey. In Handbook of Logic in Artificial Intelligence and Logic Programming, Vol.4, D. Gabbay, C.A. Hoare, J.A. Robinson (Eds.), Oxford University Press, 1995, p.1-34.
- A. Florea. Bazele logice ale inteligentei artificiale. UPB, 1995.