Logic and Artificial Intelligence

Lecture 20

Eric Pacuit

Currently Visiting the Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

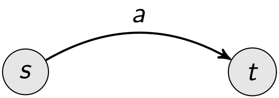
November 14, 2011

Actions

1. Actions as *transitions between states, or situations*:

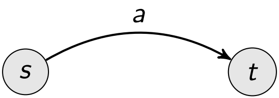
Actions

1. Actions as *transitions between states, or situations*:

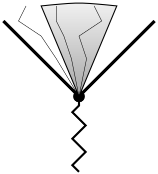


Actions

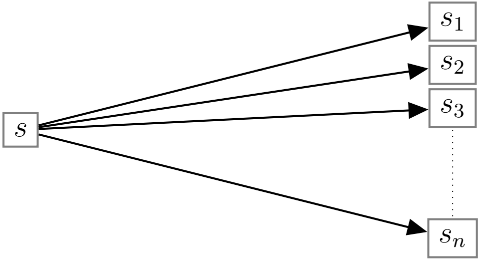
1. Actions as *transitions between states, or situations*:



2. Actions *restrict* the set of possible future histories.



J. van Benthem, H. van Ditmarsch, J. van Eijck and J. Jaspers. *Chapter 6: Propositional Dynamic Logic*. Logic in Action Online Course Project, 2011.



s
s_1
s_2
s_3
s_n

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha ; \beta \mid \alpha^* \mid \varphi ?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathrm{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha ; \beta \mid \alpha^* \mid \varphi ?$$

where $a \in \mathrm{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathrm{P}\}, V\rangle$ where for each $a \in \mathrm{P}$, $R_a \subseteq W \times W$ and $V : \mathrm{At} \to \wp(W)$

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathrm{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathrm{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathrm{P}\}, V\rangle$ where for each $a \in \mathrm{P}$, $R_a \subseteq W \times W$ and $V : \mathrm{At} \to \wp(W)$

$[\alpha]\varphi$ means "after doing α, φ will be true"

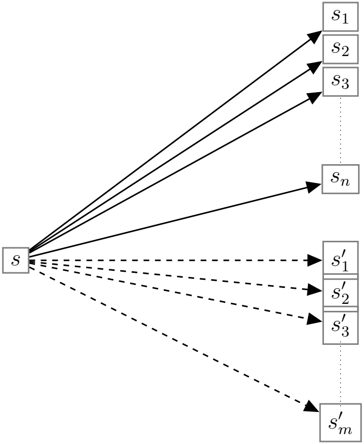
$\langle\alpha\rangle\varphi$ means "after doing α, φ may be true"

$$\mathcal{M}, w \models [\alpha]\varphi \text{ iff for each } v, \text{ if } wR_\alpha v \text{ then } \mathcal{M}, v \models \varphi$$

$$\mathcal{M}, w \models \langle\alpha\rangle\varphi \text{ iff there is a } v \text{ such that } wR_\alpha v \text{ and } \mathcal{M}, v \models \varphi$$

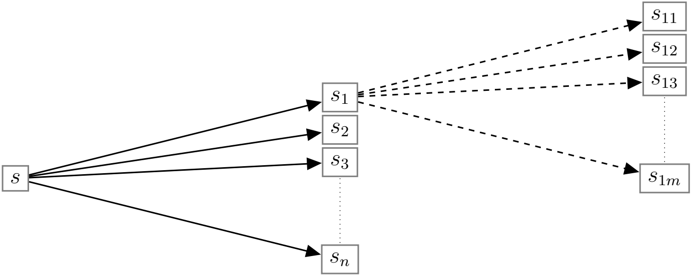
Union

$$R_{\alpha\cup\beta} := R_\alpha \cup R_\beta$$



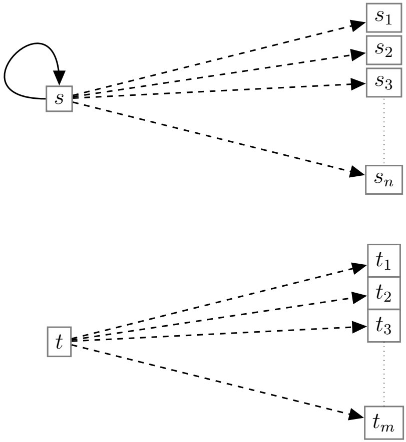
Sequence

$$R_{\alpha;\beta} := R_\alpha \circ R_\beta$$



Test

$$R_{\varphi?} = \{(w,w) \mid \mathcal{M}, w \models \varphi\}$$



Iteration

$$R_{\alpha^*} := \cup_{n \geq 0} R_\alpha^n$$

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$
8. Modus Ponens and Necessitation (for each program α)

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha;\beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$ (Fixed-Point Axiom)
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$ (Induction Axiom)
8. Modus Ponens and Necessitation (for each program α)

Actions and Ability

An early approach to interpret PDL as logic of actions was put forward by Krister Segerberg.

Segerberg adds an “agency” program to the PDL language δA where A is a formula.

K. Segerberg. *Bringing it about*. JPL, 1989.

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

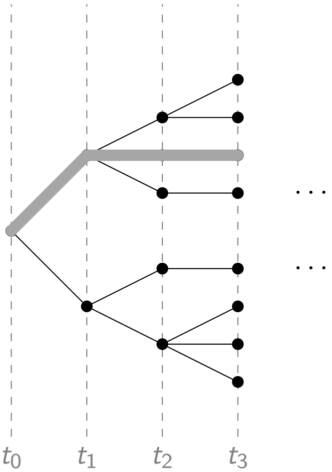
3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

1. $[\delta A]A$
2. $[\delta A]B \rightarrow ([\delta B]C \rightarrow [\delta A]C)$

Actions and Agency in Branching Time

Alternative accounts of agency do not include explicit description of the actions:



STIT

- Each node represents a choice point for the agent.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)
- The key modality is $[i\ stit]\varphi$ which is intended to mean that the agent i can "see to it that φ is true".
  - $[i\ stit]\varphi$ is true at a history moment pair provided the agent can choose a (set of) branch(es) such that every future history-moment pair satisfies φ

STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[i\ \textit{stit}]\varphi$: "the agent has the ability to bring about φ".

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, \mathit{Choice}, V \rangle$ where

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let $Hist$ be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \to \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \to \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \to \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \to \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$
- V : At $\to \wp(T \times Hist)$ is a valuation function assigning to each atomic proposition

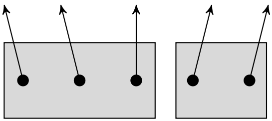
STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let $Hist$ be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \to \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \to \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$
- V : At $\to \wp(T \times Hist)$ is a valuation function assigning to each atomic proposition

Many Agents

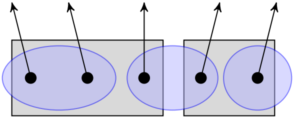
The previous model assumes there is *one* agent that “controls” the transition system.



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

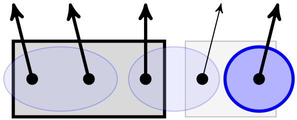
What if there is more than one agent?



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

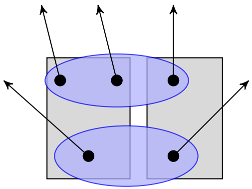
What if there is more than one agent? *Independence of agents*



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

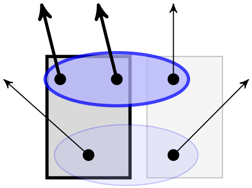
What if there is more than one agent? *Independence of agents*



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent? *Independence of agents*



STIT Language

$$\varphi \ = \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i \ stit]\varphi \mid [i \ dstit : \ \varphi] \mid \Box\varphi$$

STIT Language

$$\varphi \ = \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i \ stit]\varphi \mid [i \ dstit: \ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$

STIT Language

$$\varphi \ = \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i \ stit]\varphi \mid [i \ dstit : \ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$

STIT Language

$$\varphi \; = \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i \; stit]\varphi \mid [i \; dstit : \; \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i\ stit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$

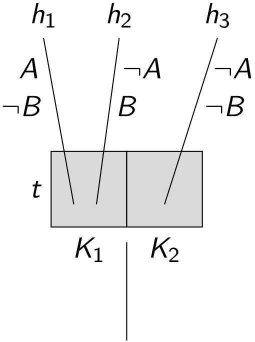
STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i \; stit]\varphi \mid [i \; dstit : \; \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i \; stit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$
- $\mathcal{M}, t/h \models [i \; dstit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$ and there is a $h'' \in H_t$ such that $\mathcal{M}, t/h \models \neg\varphi$

STIT: Example

The following are false: $A \to \Diamond[stit]A$ and $\Diamond[stit](A \vee B) \to \Diamond[stit]A \vee \Diamond[stit]B$.



J. Horty. *Agency and Deontic Logic*. 2001.

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ \mathit{stit}]$: $[i\ \mathit{stit}](\varphi \to \psi) \to ([i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\psi)$, $[i\ \mathit{stit}]\varphi \to \varphi$, $[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}][i\ \mathit{stit}]\varphi$, $\neg[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\neg[i\ \mathit{stit}]\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ stit]$: $[i\ stit](\varphi \to \psi) \to ([i\ stit]\varphi \to [i\ stit]\psi)$, $[i\ stit]\varphi \to \varphi$, $[i\ stit]\varphi \to [i\ stit][i\ stit]\varphi$, $\neg[i\ stit]\varphi \to [i\ stit]\neg[i\ stit]\varphi$
- $\Box\varphi \to [i\ stit]\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ \mathit{stit}]$: $[i\ \mathit{stit}](\varphi \to \psi) \to ([i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\psi)$, $[i\ \mathit{stit}]\varphi \to \varphi$, $[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}][i\ \mathit{stit}]\varphi$, $\neg[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\neg[i\ \mathit{stit}]\varphi$
- $\Box\varphi \to [i\ \mathit{stit}]\varphi$
- $(\bigwedge_{i \in \mathcal{A}} \Diamond[i\ \mathit{stit}]\varphi_i) \to \Diamond(\bigwedge_{i \in \mathcal{A}}[i\ \mathit{stit}]\varphi_i)$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ stit]$: $[i\ stit](\varphi \to \psi) \to ([i\ stit]\varphi \to [i\ stit]\psi)$, $[i\ stit]\varphi \to \varphi$, $[i\ stit]\varphi \to [i\ stit][i\ stit]\varphi$, $\neg[i\ stit]\varphi \to [i\ stit]\neg[i\ stit]\varphi$
- $\Box\varphi \to [i\ stit]\varphi$
- $(\bigwedge_{i \in \mathcal{A}} \Diamond[i\ stit]\varphi_i) \to \Diamond(\bigwedge_{i \in \mathcal{A}}[i\ stit]\varphi_i)$
- Modus Ponens and Necessitation for $\Box$

M. Xu. *Axioms for deliberative STIT*. Journal of Philosophical Logic, Volume 27, pp. 505 - 552, 1998.

P. Balbiani, A. Herzig and N. Troquard. *Alternative axiomatics and complexity of deliberative STIT theories*. Journal of Philosophical Logic, 37:4, pp. 387 - 406, 2008.

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ \mathit{stit}]\varphi$: the agent can "see to it that" φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ stit]\varphi$: the agent can "see to it that" φ is true
- $\Diamond[i\ stit]\varphi$: the agent has the ability to bring about φ