Logics of agency
Applications of STIT

EASSS 2018

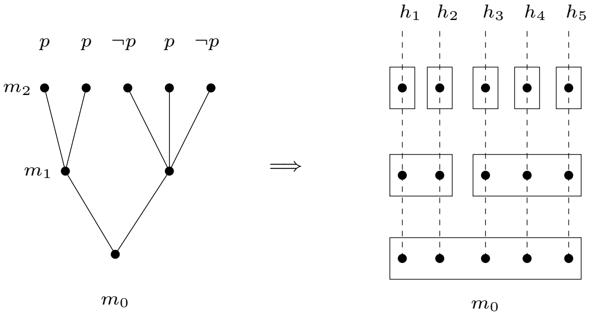
Overview of this chapter

- STIT and obligations: “oughts to do”
- Powers in concurrent games and in STIT
- STIT and knowledge “knowing how to play”

Outline

BT structures (reminder)

BT structure $\langle Mom, < \rangle$:



- History = maximally $<$-ordered set of moments
- $Hist$ = set of all histories
- H_m = set of histories passing through the moment m
- Explode moments into indexes (moment/history pairs)
  - $m_0/h_3 \not\models \mathbf{F}p$
  - $m_0/h_1 \models \mathbf{F}p$

$BT + AC$ models (reminder)

A $BT + AC$ model is a tuple $\mathcal{M} = \langle Mom, <, Choice, v \rangle$, where:

- $\langle Mom, < \rangle$ is a BT structure;
- $Choice$: Agt $\times$ $Mom \rightarrow \mathcal{P}(\mathcal{P}(Hist))$;
  - $Choice$: Agt $\times$ $Mom \rightarrow \mathcal{P}(\mathcal{P}(Hist))$
    - $Choice(a, m)$ = repertoire of choices for agent a at moment m
    - $Choice$ is a function mapping each agent and each moment m into a partition of H_m
  - $Choice(a, m) : Hist \rightarrow \mathcal{P}(Hist)$
    - For $h \in H_m$: $Choice(a, m)(h)$ = the particular choice of a at index m/h.
  - Independence of agents/choices: Let h, m.
    For all collections of $X_a \in Choice(a, m)(h)$, $\bigcap_{a \in \text{Agt}} X_a \neq \emptyset$.
  - No choice between undivided histories: if $\exists m' > m$ s.t. $h, h' \in H_{m'}$ then $h' \in Choice(a, m)(h)$.
- v is a valuation function v : Prop $\rightarrow \mathcal{P}(Mom \times Hist)$.

Language (reminder)

- "Chellas" stit:
  $M, m/h \models [G\, cstit : \varphi]$ iff $M, m/h' \models \varphi$ for all $h' \in Choice(G, m)(h)$
- historical necessity:
  $M, m/h \models \Box\varphi$ iff $M, m/h' \models \varphi$ for all $h' \in H_m$

Utilitarian deontic models (Horty 2001)

A utilitarian deontic model is a tuple $\mathcal{M} = \langle Mom, <, Choice, Value, v \rangle$, where:

- $\langle Mom, <, Choice, v \rangle$ is a $BT + AC$ model;
- $Value$ maps each history $h \in Hist$ to a real value $Value : Hist \longrightarrow \mathbb{R}$.

$Value(h) \leq Value(h')$ means that h' is at least as desirable as h.

Truth value of ought statements

Let $\mathcal{M} = \langle Mom, <, Choice, Value, v \rangle$.

$\mathcal{M}, m/h \models \bigcirc\varphi \iff \exists h' \in H_m :$

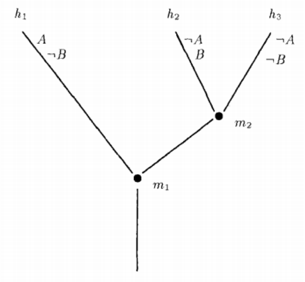
$$\begin{cases} (1) & \mathcal{M}, m/h' \models \varphi \\ (2) & \forall h'' \in H_m : \text{ if } Value(h') \leq Value(h'') \text{ then } \mathcal{M}, m/h'' \models \varphi \end{cases}$$

(φ is true for some history, and φ is true for all histories at least as desirable.)

Example: "Reparational" oughts

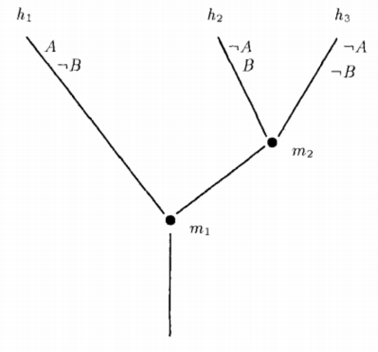
Obligations rising from violations of previous obligations. Example ((Thomason 1984), (Horty 2001)):

- It ought to be the case at the moment m_1 that a will soon board a plane to visit his aunt.
- A: "a the agent will board the plane";
- B: "a will call his aunt to say that he is not coming".
- At the moment m_1, three histories unfold.
- In h_1, agent a boards the plane.
- In h_2, a does not board the plane and calls his aunt to tell her that he will not be visiting.
- In h_3, a does not board the plane and does not call his aunt to tell her that he will not be visiting.



(Picture from (Horty 2001))

Example: “Reparational” oughts (ctd)



(Picture from (Horty 2001))

- $Value(h_1) = 10$
- $Value(h_2) = Value(h_2) = 4$
- $Value(h_3) = Value(h_3) = 0$
- $m_1/\_ \models \bigcirc A \land \neg \bigcirc B$
- $m_2/\_ \models \bigcirc B$

More theses from (Belnap et al. 2001)

Definition (Restricted complement thesis)

A variety of constructions concerned with agents and agency—including deontic statements, imperatives, and statements of intentions, among others—must take agentives as their complements.

Definition (Stit normal form thesis)

In investigations of those constructions that take agentives as complements, nothing but confusion is lost if the complements are taken to be all and only stit sentences.

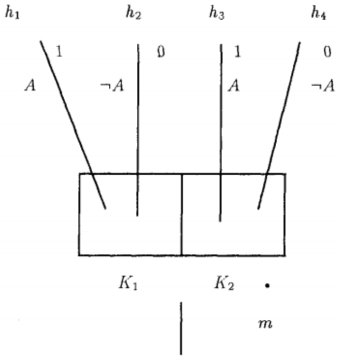
Ought to do: Meinong-Chisholm reduction

Roderick Chisholm suggests:

> *"S ought to bring it about that p" can be defined as "It ought to be that S brings it about that p."* (Chisholm 1964, p. 150)

Agent a ought to see to it that φ:

$$\bigcirc[a\, cstit : \varphi]$$



(Picture from (Horty 2001))

- $\bigcirc A$
- $\neg \bigcirc [a\, cstit\colon A]$

Logical principles of utilitarian deontic models

Valid principles:

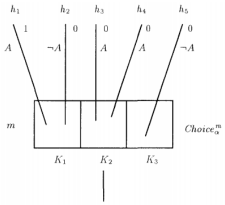
- $\bigcirc$ is a normal modal operator
- $\bigcirc\varphi \rightarrow \Diamond\varphi$
- $\bigcirc\varphi \rightarrow \Box \bigcirc \varphi$
- $\neg(\bigcirc[a\, cstit\colon \varphi] \wedge \bigcirc[b\, cstit\colon \neg\varphi])$
- $\bigcirc[a\, cstit\colon \varphi] \rightarrow \bigcirc\varphi$

However,

- $\bigcirc\varphi \rightarrow \bigcirc[a\, cstit\colon \varphi]$ is not valid
- $\bigcirc\varphi \wedge \Diamond[a\, cstit\colon \varphi] \rightarrow \bigcirc[a\, cstit\colon \varphi]$ is not valid either! (next two slides)

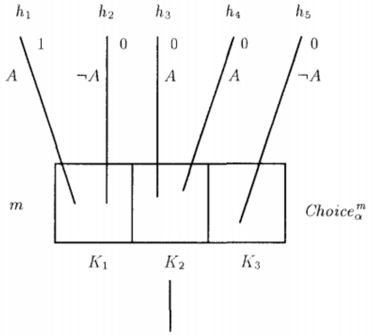
Karen buys a horse

- Karen, wishes to buy a horse, but she has only \$10,000 to spend and the horse she wants is selling for \$15,000;
- We imagine that Karen offers \$10,000 for the horse at the moment m (choice K_1);
- It is up to the owner of the horse to decide whether to accept the offer. The history h_1 represents a scenario in which the owner accepts Karen's offer, h_2 a scenario in which the offer is rejected;
- A is the statement that Karen will become less wealthy by the amount of \$10,000;
- The unique best history is h_1, in which the offer is accepted, and, as a consequence, Karen buys the horse and becomes less wealthy by \$10,000;
- Since Karen is less wealthy by \$10,000 in the unique best history, we must conclude that it ought to be that she is less wealthy by \$10,000;
- Of course, Karen also has the ability (throwing away the money) to see to it that she is less wealthy by \$10,000 (choice K_2);
- But we would not wish to conclude that Karen ought to see to it that she is less wealthy by \$10,000.



(Picture from (Horty 2001))

Karen buys a horse (ctd)

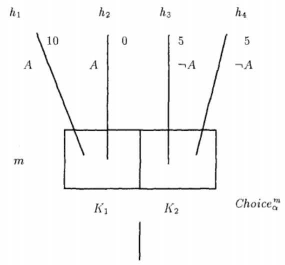


(Picture from (Horty 2001))

- $\bigcirc A$
- $\Diamond[a\, cstit\colon A]$
- $\neg \bigcirc [a\, cstit\colon A]$

Criticism of the utilitarian deontic model (risk seeking)

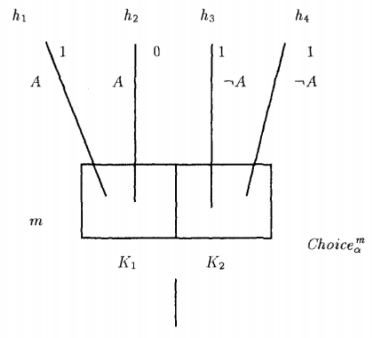
- An agent a is faced with two options at the moment m: to gamble the sum of five dollars (K_1), or to refrain from gambling (K_2).
- If a gambles, there is a history in which he wins ten dollars, and another in which he loses his stake;
- If a does not gamble, he preserves his original stake;
- the utility associated with each history at m is entirely determined by the sum of money that a possesses at the end;
- The letter A stands for the proposition that a gambles;
- $\bigcirc[a\, cstit: A]$ holds at m.



(Picture from (Horty 2001))

Criticism of the utilitarian deontic model (missing obligations)

If we change the utilities:



(Picture from (Horty 2001))

- Then $\neg \bigcirc [a\, cstit: A]$. Good.
- But we should expect that it is wise not to gamble here.
- However, $\neg \bigcirc [a\, cstit: \neg A]$.

A first solution with ordered choices

Definition (weak preference over choices)

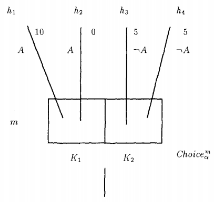
$X \leq Y$ iff $\forall h \in X, \forall h' \in Y : Value(h) \leq Value(h')$.

A “ought to stit” operator with ordered choices:

$m/h \models [a\, ostit\colon A]$ iff $\exists K \in Choice(a, m)$ such that

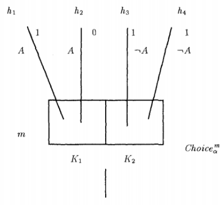
1. $\{m\} \times K \subseteq ||A||$ and
2. $\forall K' \in Choice(a, m) : K' \leq K$.

Gambling again



(Picture from (Horty 2001))

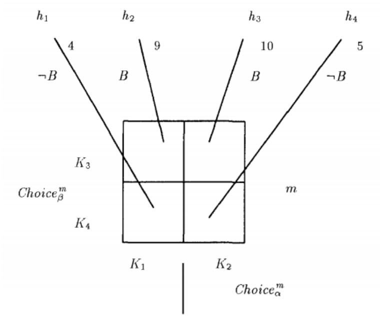
$\neg[a\,ostit: A] \wedge \neg[a\,ostit: \neg A]$



(Picture from (Horty 2001))

$[a\,ostit: \neg A]$

Further problem with multiagency



(Picture from (Horty 2001))

K_2 seems preferable for agent a, but it is not the case that $K_1 \leq K_2$.

States: choices of others

We define the "strategic contexts" agent a might face.

$$State(a, m) = Choice(\mathsf{Agt} \setminus \{a\}, m) \ .$$

When there are two players (e.g., previous example):

$$State(a, m) = Choice(b, m) \ ,$$

and

$$State(b, m) = Choice(a, m) \ .$$

Choice dominance

Definition (weak choice dominance)

Let $K, K' \in Choice(a, m)$. $K \preceq_a K'$ iff $K \cap S \leq K' \cap S$ for every $S \in States(a, m)$.

On the previous example: $K_1 \prec_a K_2$.

Optimal choice: a second solution

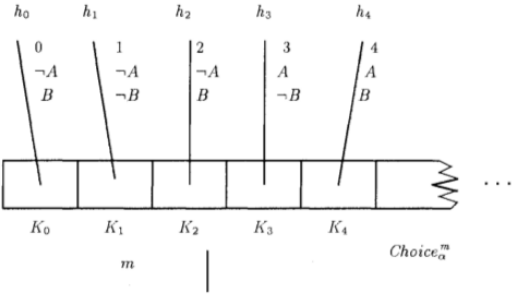
Define:

$$Optimal(a,m) = \{K \in Choice(a,m) \mid \nexists K' \in Choice(a,m), K \prec_a K'\} \ .$$

When there is a finite number of choices, this revision of $[a\,ostit\colon A]$ of works well:

$m/h \models [a\,ostit\colon A]$ iff $\{m\} \times K \subseteq ||A||$ for every $K \in Optimal(a,m)$.

Further problem with infinite repertoires of choices



(Picture from (Horty 2001))

We'd like to have $[a\, ostit\colon A]$ and $\neg[a\, ostit\colon \neg A]$.
But $Optimal(a, m) = \emptyset$...

The “ought to stit” operator

We revise $[a\, ostit\colon A]$ further into:

$m/h \models [a\, ostit\colon A]$ iff for every $K \in Choice(a, m)$, if $\{m\} \times K \not\subseteq ||A||$, then there is $K' \in Choice(a, m)$ such that:

1. $K \prec_a K'$, and
2. $\{m\} \times K' \subseteq ||A||$, and
3. $\{m\} \times K'' \subseteq ||A||$ for each $K'' \in Choice(a, m)$ such that $K' \preceq_a K''$.

This is obligation to do. Noted $\odot[a\, cstit\colon A]$ in (Horty 2001, p. 77).

Outline

Concurrent games and Coalition Logic

Definition (Concurrent game models)

A *concurrent game model* is a tuple $G = (S, \{\Sigma_a | a \in \mathsf{Agt}\}, o, V)$ where:

- S is a nonempty set of states,
- Σ_a is a nonempty set of choices for every agent $a \in \mathsf{Agt}$,
- $o : S \times \prod_{a \in \mathsf{Agt}} \Sigma_a \longrightarrow S$ is an outcome function,
- $V : S \longrightarrow \mathcal{P}(\mathsf{Prop})$ is a valuation function.

$$G, s \models \langle\!\langle C \rangle\!\rangle \varphi \text{ iff } \exists \sigma_C \in \Sigma_C, \forall \sigma_{\overline{C}} \in \Sigma_{\mathsf{Agt} \setminus C} \text{ s.t. } G, o((\sigma_C, \sigma_{\overline{C}})) \models \varphi$$

where for every coalition $C \subseteq \mathsf{Agt}$ we note $\Sigma_C = \Pi_{a \in C} \Sigma_a$.

(Pauly 2001,2002), (Alur, Henzinger, Kupferman 2002),
(Goranko, Jamroga 2004), (Goranko, Jamroga, Turrini 2010,2013)

Axiomatics of Coalition Logic

- Propositional Logic
- $\langle\![C]\!\rangle \top$
- $\neg \langle\![C]\!\rangle \bot$
- $\neg \langle\![\emptyset]\!\rangle \neg\varphi \to \langle\![\mathsf{Agt}]\!\rangle \varphi$
- $\langle\![C]\!\rangle (\varphi \wedge \psi) \to \langle\![C]\!\rangle \varphi$
- $\langle\![C_1]\!\rangle \varphi \wedge \langle\![C_2]\!\rangle \psi \to \langle\![C_1 \cup C_2]\!\rangle (\varphi \wedge \psi)$, when $C_1 \cap C_2 = \emptyset$
- if $\vdash \varphi \leftrightarrow \psi$ then $\vdash \langle\![C]\!\rangle \varphi \leftrightarrow \langle\![C]\!\rangle \psi$

Power in STIT?

$\Diamond[a\, astit\colon \varphi]$ does not capture any kind of power.

$\Diamond[a\, cstit\colon \varphi]$ does.

How to embed Coalition Logic?

A discrete-deterministic STIT

Hypothesis (discreteness)

Given a moment m_1, there exists a successor moment m_2 such that $m_1 < m_2$ and there is no moment m_3 such that $m_1 < m_3 < m_2$.

$m/h \models \mathbf{X}\varphi$ iff φ is true at the moment immediately after m on h

Hypothesis (determinism)

$\forall m \in Mom,\ \exists m' \in Mom\ (m < m'$ *and* $\forall h \in H_{m'}, Choice(\mathsf{Agt}, m)(h) = H_{m'})$

Translation of Coalition Logic to discrete-deterministic STIT

$$
\begin{array}{lcl}
tr(p) & = & \Box p, \text{ for } p \in \textsf{Prop} \\
tr(\neg\varphi) & = & \neg tr(\varphi) \\
tr(\varphi \vee \psi) & = & tr(\varphi) \vee tr(\psi) \\
tr(\langle\!\![C]\!\!\rangle\varphi) & = & \Diamond[C]\mathbf{X}tr(\varphi)
\end{array}
$$

In STIT terminology

"the coalition C has the power to φ"

can be paraphrased by

"it is historically possible that C sees to it that next φ"

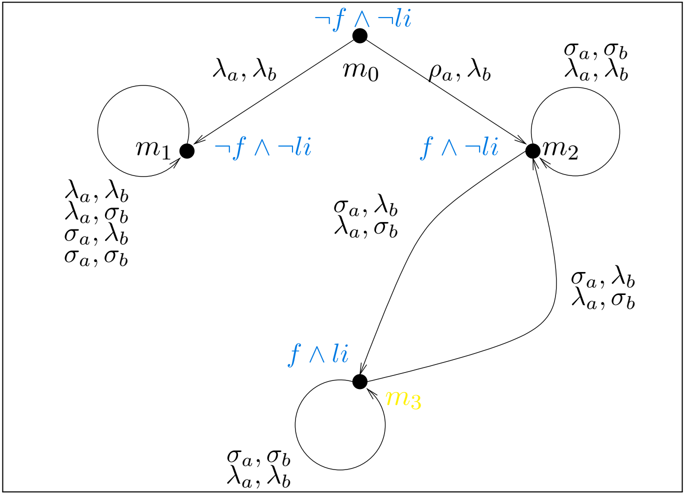
Theorem ((Broersen, Herzig, Troquard 2006))

tr is a correct embedding of CL into discrete-deterministic STIT.

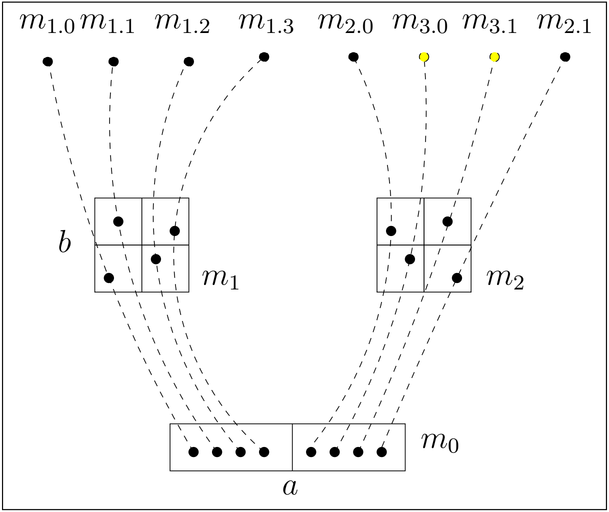
Example: Ann and Bill switch the light

- Four states: m_0, m_1, m_2, m_3
- li = light is on (at m_3)
- f = lamp is functioning (at m_2 and m_3)
- At moment m_0, agent a has a choice between *repairing* a broken lamp (ρ_a) or *remaining passive* (λ_a). Agent b has the vacuous choice of *remaining passive* (λ_b).
- If a chooses not to repair, the system reaches m_1. If a chooses to repair, the system reaches m_2.
- In m_1, m_2 and m_3 both a and b can choose to *toggle* a light switch (τ_a and τ_b) or *not toggle* (λ_a and λ_b).
- If a repairs at m_0 then a and b 'play toggling' between m_2 and m_3

Game model


$\neg f \wedge \neg li$
m_0
λ_a, λ_b
ρ_a, λ_b
σ_a, σ_b
λ_a, λ_b
m_1
$\neg f \wedge \neg li$
$f \wedge \neg li$
m_2
λ_a, λ_b
λ_a, σ_b
σ_a, λ_b
σ_a, σ_b
σ_a, λ_b
λ_a, σ_b
σ_a, λ_b
λ_a, σ_b
$f \wedge li$
m_3
σ_a, σ_b
λ_a, λ_b

Corresponding STIT model



Beyond Coalition Logic

From Alternating-time Temporal Logic to Strategic Chellas stit of ability

ATL (Alur, Henzinger, Kupferman 2002):

- concurrent game models
- $\langle\langle C\rangle\rangle \mathbf{X}\varphi \mid \langle\langle C\rangle\rangle \varphi \mathbf{U}\psi$

Corresponds to ((Broersen, Herzig, Tr. 2006, JLC)):

- Models: discrete-time and deterministic STIT
- $p \rightsquigarrow \Box p$,
- $\langle\langle C\rangle\rangle \mathbf{X}\varphi \rightsquigarrow \Diamond_s[C\, scstit: \mathbf{X}\varphi]$,
- $\langle\langle C\rangle\rangle \varphi \mathbf{U}\psi \rightsquigarrow \Diamond_s[C\, scstit: \varphi\mathbf{U}\psi]$.

Outline

CL models vs. $BT + AC$ models

Coalition Logic

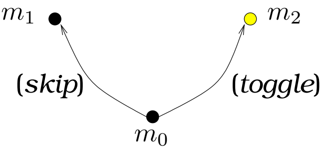
- Concurrent game models
- Neighborhood models (effectivity structures)
- Idea: associate a strategic game (form) to every world

In $BT + AC$ models, indexes represent both

- the current state of affairs of the world, and
- the current choice/commitment of agents

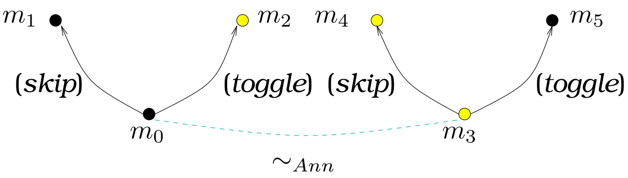
Ann toggles

- At m_0, the light is off: $m_0 \models \neg li$
- Ann can *toggle* or *skip*
- $m_0 \models \langle\!\langle Ann \rangle\!\rangle li$
  at m_0, "Ann is able to achieve li"



Poor blind Ann – a CL account

- As before, the light is off: $m_0 \models \neg li$
- Ann is blind and cannot distinguish a world where the light is on from a world where the light is off
- $m_0 \models K_{Ann}\langle\!\langle Ann \rangle\!\rangle li$
  at m_0, "Ann knows she is able to achieve li"



Adding knowledge

A logical language of action and knowledge should be able to distinguish the following scenarii:

1. the agent a knows it has a particular action/choice in its repertoire that ensures φ, possibly without knowing which choice to make to ensure φ.
2. the agent a 'knows how to' / 'can' / 'has the power to' ensure φ.

Two readings of "having a strategy"

- $tr(K_C \langle\!\langle C \rangle\!\rangle \varphi) = K_C \Diamond [C] \mathbf{X} \varphi$ **(de dicto)**
  Group C knows (K) there is (∃) a choice s.t. for all (∀) possible outcomes φ
  - Alternating-time *Epistemic* Temporal Logic ATEL (Wooldridge, van der Hoek 2002)
- We might want: $\Diamond K_C [C] \mathbf{X} \varphi$ **(de re)**
  There is a choice (∃), s.t. group C knows (K) that for all (∀) possible outcomes φ
  - ATEL does not deal with *de re* strategies (Jamroga 2003), (Schobbens 2004)
  - Several corrections (Schobbens 2004), (Jamroga, van der Hoek 2004), (Jamroga, Ågotnes 2006, 2007)
  - First semantics with STIT (Herzig, Troquard 2006)

Epistemic STIT

Language.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{X}\varphi \mid [C]\varphi \mid K_a\varphi$$

$BT + AC + K$-models are tuples $\mathcal{M} = (Mom, <, Choice, \sim, V)$ where:

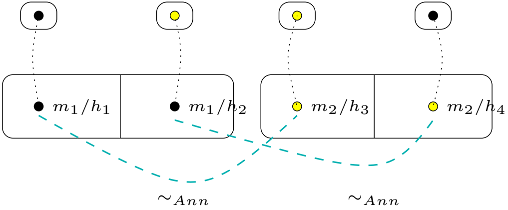
- $(Mom, <, Choice, V)$ is an $BT + AC$-model.
- $\sim \subseteq (Mom \times Hist) \times (Mom \times Hist)$ is a collection of equivalence relations $\sim_i$ (one for every agent $i \in \mathsf{Agt}$) over indexes.

Extra operators:

- $\mathcal{M}, m/h \models K_i\varphi$ iff for all $m'/h' \sim_i m/h$, $\mathcal{M}, m'/h' \models \varphi$

Every K_i is a standard epistemic modality. (Hintikka 1962)

Poor blind Ann again



Epistemic relations are over indexes instead of moments.

- $m_i/h_j \models K_{Ann}\Diamond[Ann]\mathbf{X}\varphi$
  Ann knows she has an action that leads to a lighten moment.
- $m_i/h_j \not\models \Diamond K_{Ann}[Ann]\mathbf{X}\varphi$
  Ann does **not** know how to achieve it.

References I

Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman.
Alternating-time temporal logic.
J. ACM, 49(5):672–713, 2002.

Jan Broersen, Andreas Herzig, and Nicolas Troquard.
Embedding Alternating-time Temporal Logic in Strategic STIT Logic of Agency.
J. Log. Comput., 16(5):559–578, 2006.

Jan Broersen, Andreas Herzig, and Nicolas Troquard.
From Coalition Logic to STIT.
Electr. Notes Theor. Comput. Sci., 157(4):23–35, 2006.

Jan Broersen, Andreas Herzig, and Nicolas Troquard.
A normal simulation of coalition logic and an epistemic extension.
In *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2007), Brussels, Belgium, June 25-27, 2007*, pages 92–101, 2007.

Jan Broersen, Andreas Herzig, and Nicolas Troquard.
What groups do, can do, and know they can do: an analysis in normal modal logics.
Journal of Applied Non-Classical Logics, 19(3):261–290, 2009.

Valentin Goranko and Wojciech Jamroga.
Comparing semantics of logics for multi-agent systems.
Synthese, 139(2):241–280, Mar 2004.

Valentin Goranko, Wojciech Jamroga, and Paolo Turrini.
Strategic games and truly playable effectivity functions.
Autonomous Agents and Multi-Agent Systems, 26(2):288–314, 2013.

J. F. Horty.
Agency and Deontic Logic.
Oxford University Press, Oxford, 2001.

References II

Andreas Herzig and Nicolas Troquard.
Knowing how to play: uniform choices in logics of agency.
In *5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006), Hakodate, Japan, May 8-12, 2006*, pages 209–216, 2006.

Wojciech Jamroga and Thomas Ågotnes.
Constructive knowledge: what agents can achieve under imperfect information.
Journal of Applied Non-Classical Logics, 17(4):423–475, 2007.

Wojciech Jamroga and Wiebe van der Hoek.
Agents that know how to play.
Fundam. Inform., 63(2-3):185–219, 2004.

Marc Pauly.
A modal logic for coalitional power in games.
J. Log. Comput., 12(1):149–166, 2002.

Pierre-Yves Schobbens.
Alternating-time logic with imperfect recall.
Electr. Notes Theor. Comput. Sci., 85(2):82–93, 2004.

Wiebe van der Hoek and Michael Wooldridge.
Tractable multiagent planning for epistemic goals.
In *Proc. First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2002*, pages 1167–1174, 2002.