‘Can’, ‘knowing that you can’, ‘doing’ and ‘knowing that you do’

Jan Broersen
Information and Computing Sciences
Utrecht University

Disclaimer: this is work in progress...

Overview

- Introduction
- Alternating Transition Systems
- ‘can / ability’: Coalition logic
- ‘knowing that it is possible to do’
- ‘doing’: STIT-logic
- ‘knowing that you do’
- ‘knowing that you can’ (or: ‘knowing how to’)
- Going fully strategic: ATL / ATEL / ATL-STIT / E-ATL-STIT
- Summary results
- Conclusions

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Notions central in this talk

- The notions of the title
- ‘Knowing’ is usually applied to conditions: we say that ‘we know that Bush is a liar’ to describe our knowledge about one of the conditions applying to Bush.
- In what sense can we apply the qualification ‘knowing’ to actions?
- E.g.:
  I am giving a talk and know it
  I know how to give a talk
  I know Bush is destroying the world
  Bush is knowingly destroying the world (he has bad intentions)
  Bush is unknowingly destroying the world (he is only stupid)
  Bush is saving the world but I believe he is destroying it (I am stupid)

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Common Semantic Ground: Alternating Transition Systems (ATSs)

- We will discuss: Coalition Logic, STIT-logic, Strategic STIT Logic, Alternating Time Temporal Logic (ATL)
- All of these can be interpreted on ATSs
- But we can also interpret: Computation Tree Logic (CTL), CTL*, ATL*, and, of course, many more.
- In particular, we discuss epistemic extenions. To this end we extend ATS with an epistemic equivalence relation.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

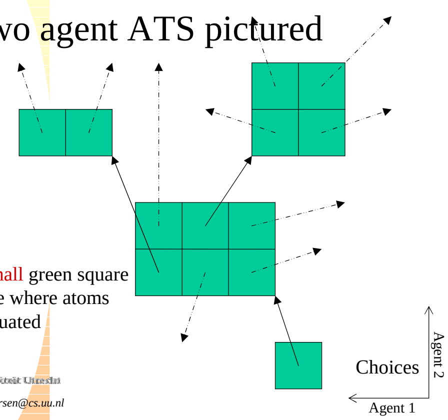
Characteristics of Alternating Transition Systems

1. Consist of states and collective choices (actions)
2. Deterministic past
3. Non-deterministic future
4. Relative to a fixed and finite set of agents E
5. A strategy is a mapping from states and agents to choices
6. Non-determinism of actions (choices) of agents is *only* due to lack of knowledge about what other agents choose in the same state
7. from 5. it follows: system actions (actions performed by all agents simultaneously) are deterministic and serial.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

A two agent ATS pictured

Each small green square is a state where atoms are evaluated



CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Alternating Transition Systems, formally

Coalition Logic is interpreted on game structures.
We interpret it on Alternating Transition Systems.

Models: $\mathcal{ATS} = (S, \mathcal{C}, \pi)$

(1) S are states.

(2) $\mathcal{C} : A \times S \mapsto 2^{2^S}$ yields for each agent and state a set of choices / actions. Furthermore, the function $\mathcal{R}X(s) = \{\bigcap_{a \in A} C \mid C \in \mathcal{C}(a, s)\}$ yields a non-empty set of singleton sets for each s.

(3) π valuates atomic propositions.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

‘can’: Coalition Logic (Pauly, ...)

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Can: Coalition logic

- Syntax: modalities ⟨[A]⟩Xϕ (and duals [⟨A⟩]Xϕ) for every group of agents A.
- In Pauly’s original syntax these same operators are written as just [A]ϕ and ⟨A⟩ϕ. We explicate the three constituing modalities: ⟨[A]⟩Xϕ means A *can* cooperate to *ensure* ϕ in the *next* state.
- Axiomatization shows that CL relates normal and non-normal phenomena:

$(\bot)$	$\neg\langle[A]\rangle X\bot$
$(\top)$	$\langle[A]\rangle X\top$
(N)	$\neg\langle[\emptyset]\rangle X\neg\varphi \rightarrow \langle[E]\rangle X\varphi$
(M)	$\langle[A]\rangle X(\varphi \wedge \psi) \rightarrow \langle[A]\rangle X\varphi$
(S)	$\langle[A_1]\rangle X\varphi \wedge \langle[A_2]\rangle X\psi \rightarrow \langle[A_1 \cup A_2]\rangle X(\varphi \wedge \psi)$ if $A_1 \cap A_2 = \emptyset$
(RE)	from $\varphi \equiv \psi$ infer $\langle[A]\rangle X\varphi \equiv \langle[A]\rangle X\psi$

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

CL semantics: strategies

A strategy α_a for an agent a, is a function $\alpha_a : S \mapsto 2^S$, such that $\alpha_a(s) \in C(a, s)$, assigning choices of the agent a to states.

Strategy functions α_a for individual agents a are straightforwardly combined to (partial) strategy functions $\alpha_A : S \times E \mapsto 2^S$ for sets of agents $A = \{a_1, a_2, \ldots, a_k\} \subseteq E$. Notation: $\alpha_A = \alpha^1_{a_1} \mid \alpha^2_{a_2} \mid \ldots \mid \alpha^k_{a_k}$

Furthermore, for $A \cap B = \emptyset$, we use $\alpha_A \mid \beta_B$ to denote the joined partial strategy function build from α_A and β_B.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

CL semantics: a follow-up function

Clearly, for system strategies $\alpha_E : S \times E \mapsto 2^S$ with $E = \{a_1, a_2, \ldots, a_n\}$, we have $\alpha_E = \alpha^1_{a_1} \mid \alpha^2_{a_2} \mid \ldots \mid \alpha^n_{a_n}$

Given a system strategy α_E, the follow-up function $F_{\alpha_E} : S \mapsto S$ is the intersection of all choices for indiv. agents: $F_{\alpha_E}(s) = \bigcap_{a \in E} \alpha_E(s, a)$.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

CL semantics: truth definitions

We have the standard definitions for the propositional logic operators, plus the following definition for the CL-operator:

$\mathcal{M}, s \models \langle[A]\rangle X\varphi \Leftrightarrow$
$\exists\beta_A$ such that $\forall\gamma_{\overline{A}}$ it holds that $\mathcal{M}, F_{\beta_A \mid \gamma_{\overline{A}}}(s) \models \varphi$

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

CKI master-colloquium 2006

'knowing that you can': Epistemic Coalition Logic?

Universiteit Utrecht

broersen@cs.uu.nl

Epistemic Coalition Logic

- We have to add epistemic indistinghuishability relations to ATSs. We get EATSs.
- Initial attempt: add S5 indistinghuishability relations over states.
- Observation: first unravel, or, only use trees.
- In this context (to be precise, in ATEL [vd Hoek & Wooldridge: Studia Logica 2003]), a problem emerged that, in this context, is usually refered to simply as 'the problem of uniform strategies'.

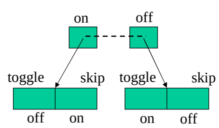
CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

The Problem of Uniform Strategies

Consider the following EATS that applies to a blind person entering a room. He does not know whether or not the light is on, but he knows there is a switch that controls the light.



In ATEL it is perceived as counter-intuitive that in both top-states it holds that: $K_{blind} \langle[blind]\rangle X\ On$ while the agent does not know a uniform strategy...

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

Analysis of the problem

- Actually the formula is not counter-intuitive; it expresses exactly what is the case!
- If in addition we want to express that the agent does not know what to do to ensure that the light is on, we need to strengthen the language.
- Idea: make the strategies explicit in the object language. There are several attempts in the literature (towards dynamic logic versions of ATL)
- Our solution: break up the original operator $\langle[A]\rangle X$ into $\Diamond_A [A]X$ which enables us to express that in the example: $\neg \Diamond_{blind} K_{blind} [blind]\ X\ On$ and $\neg K_{blind} [blind]\ X\ On$
- We can also express that things might be such that *actually*: $[blind]\ X\ LightOn$, at the same time!
- Having a uniform strategy is simply expressed as: $\Diamond_{blind} K_{blind} [blind]\ X\ On$
- $[A]X$ is actually known as a so called (C)STIT operator in the philosophical literature!

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

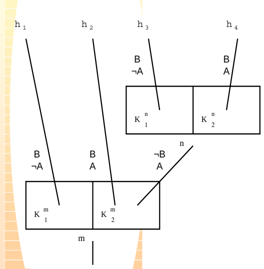
'doing': STIT theory (Chellas, Belnap, Horty, ...)

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

(C)STIT-theory (I)



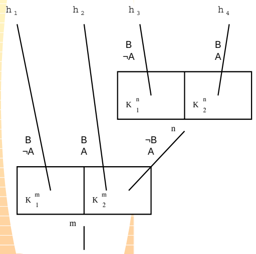
- Actions are not state transformers, but only history selectors and agent-effect binders.
- Crucial: evaluation is w.r.t. moment-history pairs
  ⇒ a two-dimensional modal logic

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

(C)STIT-theory (II)



$M, m, h_3 \vDash A$

$\square$ = `historical necessity'

F = `Future'

P = `Past'

$M, m, h_3 \vDash \square(A \vee B)$

$M, m, h_3 \vDash F\ \neg A$

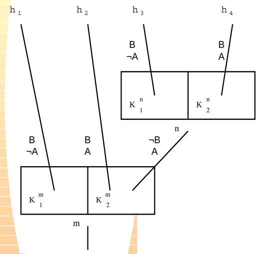
$M, n, h_3 \vDash P\ A$

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

(C)STIT-theory (III)



$m,h \vDash [\alpha \text{ stit: } A]$
iff
$\forall$ (m,h') for which
Act(α,m,h) = Act(α,m,h') it holds that
m,h' ⊨ A

$m,h_3 \vDash [\alpha \text{ stit: } A]$

$m,h_3 \nvDash [\alpha \text{ stit: } B]$

$m,h_3 \nvDash [\alpha \text{ stit: } \neg B]$

$m,h_3 \vDash [\alpha \text{ stit: } F\ B]$

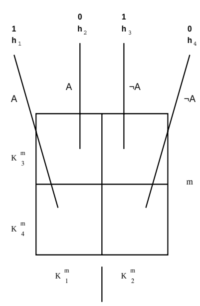
CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

Multi-agency in STIT

- Groups are introduced by requiring independence of choices
- Stit is generalized according to the 'sure thing principle': seeing to it is ensuring a condition no matter what the other agents do; indeed, as in Coalition Logic.



CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

Interpreting a CL-STIT logic on ATSs

- On ATSs we could interpret STIT logics fairly easy by evaluating with respect to history / state pairs of ATSs.
- However, we will evaluate with respect to strategy / state pairs for reasons that will become clear later.
- This results in the same STIT-logics.
- This is easy to see: all history / state pairs belonging to same choice evaluate to the same value. Than we might just as well evaluate with respect to choices directly.
- We introduce modal operators for quantification over strategies. Stit's historical necesity $\square\ \phi$ then coincides with the special case $\square_{\alpha_E}\ \phi$ where α_E is a system strategy.
- We introduce STIT operators of the from [A]X ϕ saying that ϕ is true next, given the present choice of A, no matter what the other agents do.

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

Strategies: restrictions to agent sub-domains

We need one more definition regarding *partial* strategy functions $\alpha_A : S \times E \mapsto 2^S$, where $A \subseteq E$.

For $B \subseteq A$ the strategy $\alpha_A \upharpoonright_B$ denotes the (partial) strategy function that is the restriction of the (partial) strategy function α_A to the domain of agents B

Note that $\alpha_A \upharpoonright_A = \alpha_A$.

Note that $(\alpha_A \,|\, \beta_B)\upharpoonright_A = \alpha_A$ and $(\alpha_A \,|\, \beta_B)\upharpoonright_B = \beta_B$.

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

Truth definition CL-STIT on ATSs

$\mathcal{M}, \alpha_A, s \vDash \Diamond_B \varphi \Leftrightarrow \exists \beta_B$ such that $\mathcal{M}, \beta_B, s \vDash \varphi$

$\mathcal{M}, \alpha_A, s \vDash [B]X\varphi \Leftrightarrow \forall \beta_{\overline{A \cap B}}$ it holds that $\mathcal{M}, \alpha_A, F_{\alpha_E}(s) \vDash \varphi$ where α_E is defined as: $\alpha_E = \alpha_A \upharpoonright_{A \cap B} | \beta_{\overline{A \cap B}}$

The first truth condition keeps the state fixed, the second the strategy.

The crucial idea is to take the intersection of groups A and B in the strategy definition above.

Indeed, ⟨[A]⟩Xϕ can now be defined as $\Diamond_A$ [A]Xϕ

CKI master-colloquium 2006

Universiteit Utrecht

broersen@cs.uu.nl

Validities of CL-STIT

- The following are valid (there is no prove of completeness yet)

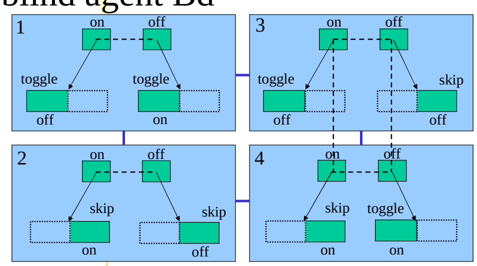
$\models [A]\Box_A\varphi \leftrightarrow \Box_A[A]\varphi$
For every A the logic of $[A]\varphi$ is KD
For every A the logic of $\Box_A\varphi$ is $S5$
$\models \Box_A\varphi \rightarrow \Box_B\Box_A\varphi$
$\models \Diamond_A\varphi \rightarrow \Box_B\Diamond_A\varphi$
$\models [A]X\varphi \rightarrow [B]X\varphi$ for $A \subseteq B$
$\models \langle A\rangle X\varphi \rightarrow \langle B\rangle X\varphi$ for $A \supseteq B$
$\models [A]X\varphi \wedge [B]X\psi \rightarrow [A \cup B]X(\varphi \wedge \psi)$
$\models \langle A \cup B\rangle X(\varphi \wedge \psi) \rightarrow \langle A\rangle X\varphi \vee \langle B\rangle X\psi$

Back to the blind: how to approach his problem semantically?

- Epistemic indistinguishability relations are now defined to relate strategy / state pairs (this is crucial!)
- Drawback: it gets rather difficult to picture the models...
- In the next picture, there are three relations:
  1. arrows are the actions and interpret X and [A]
  2. blue lines relate strategies (equiv. rel.) and interpret $\Diamond_A$
  3. *doted* lines relate epistemically indistinguishable pairs (equiv. rel.) and interpret K_A

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

A model for the situation of the blind agent Bd



4 (blue) strategies, 8 strategy-state pairs
1 & 2: Bd decides to toggle, skip, respectively
3 & 4: Bd asks a second, sighted agent to flip a coin and to execute 3 if heads and 4 if tails.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

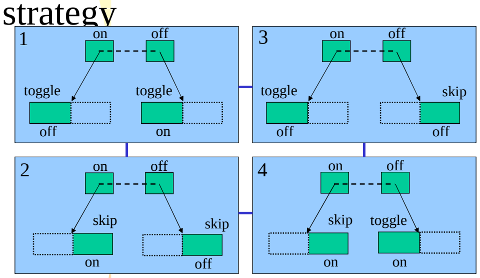
Formulas true in the model

- Everywhere it holds that K_{blind} $\Diamond_{blind}$ [blind] X On
- Everywhere it holds that ¬ $\Diamond_{blind}$ K_{blind} [blind] X On
- Everywhere it holds that ¬ K_{blind} [blind] X On

- Depending on the actual strategy / state pair, for the blind agent it may hold that
  On or ¬ On and
  [blind] X On or ¬[blind] X On

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

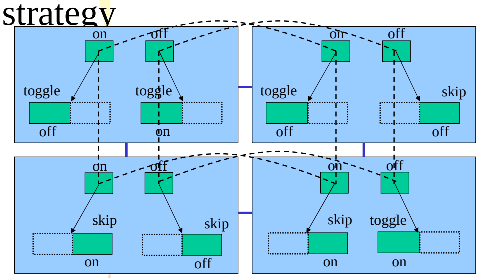
Knowing to have a uniform strategy



Now the agent Bd can simply ask the second, sighted (and obedient) agent to implement one of the strategies.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Unknowingly performing a strategy



In the lower-right square: the agent is now able to see, follows a strategy of making light, but does not know it (presumably he is drunk).

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

CKI master-colloquium 2006

Going fully strategic: series of choices

Universiteit Utrecht

broersen@cs.uu.nl

CKI master-colloquium 2006

Features of Alternating time Temporal Logic (ATL)

- ATL (Alur & Henzinger & Kupferman 1997) gains more and more popularity as a modeling language for multi-agent systems.
- Can be interpreted on multi-player game structures (MGMs). But, we prefer alternating transition systems (ATSs) for the semantics.
- Relation with planning: strategies appearing in the semantics are like conditional plans.
- Has notions of agency / ability / control (unlike e.g. CTL), enables deliberate versions, game-theoretic notions, etc.
- In dynamic logic action refinement (as in HTN-planning) is hard to express. In ATL not.
- ATL is decidable, there is a tableaux system, and there is a model checker (Mocha).
- Ongoing research on the combination of ATL with epistemic logic (ATEL, vd Hoek & Wooldridge: Studia Logica 2003, etc.).
- There are approaches to combining ATL with obligations [Jamroga & Van der Hoek & Wooldridge: DEON'04, Broersen: DEON'06].

Universiteit Utrecht

broersen@cs.uu.nl

CKI master-colloquium 2006

ATL syntax (informal)

We slightly adapt the standard ATL syntax:

ATL has *operators for quantification over strategies*

$\langle[A]\rangle\ \eta$ ('A have a strategy that ensures η') and

$[\langle A\rangle]\ \eta$ ('A do not have a strategy ensuring $\neg\eta$, so whatever strategy they take, η is a possible outcome'),

and it has *linear time* operators (at positions η)

$\phi\ U\ \psi$ ('Until'), $\phi\ U_w\ \psi$ ('weak Until'),

$G\ \phi$ ('Globally ϕ'), $X\ \phi$ ('neXt ϕ'),

$F\ \phi$ ('at some Future point ϕ')

Universiteit Utrecht

broersen@cs.uu.nl

CKI master-colloquium 2006

ATL syntax (formally)

$$\begin{array}{lcl}\varphi,\psi,\ldots & := & p \mid \neg\varphi \mid \varphi\wedge\psi \mid \langle[A]\rangle\eta \mid [\langle A\rangle]\eta \\ \eta,\theta,\ldots & := & \varphi U^{ee}\psi\end{array}$$

$$\begin{array}{l}\langle[A]\rangle X\varphi \equiv_{def} \langle[A]\rangle(\bot U^{ee}\varphi) \\ \langle[A]\rangle F\varphi \equiv_{def} \varphi \vee \langle[A]\rangle(\top U^{ee}\varphi) \\ \langle[A]\rangle G\varphi \equiv_{def} \neg[\langle A\rangle]F\neg\varphi \\ \langle[A]\rangle(\varphi U^{e}\psi) \equiv_{def} \varphi \wedge \langle[A]\rangle(\varphi U^{ee}\psi) \\ \langle[A]\rangle(\varphi U\psi) \equiv_{def} \langle[A]\rangle(\varphi U^{e}(\varphi\wedge\psi)) \\ \langle[A]\rangle(\varphi U_w\psi) \equiv_{def} \neg[\langle A\rangle](\neg\psi U\neg\varphi) \\ [\langle A\rangle]X\varphi \equiv_{def} [\langle A\rangle](\bot U^{ee}\varphi) \\ etc.\end{array}$$

Universiteit Utrecht

broersen@cs.uu.nl

CKI master-colloquium 2006

ATL semantics

Given a system strategy α_E, the follow up function $F_{\alpha_E} : S \mapsto S$ is the intersection of all choices for indiv. agents: $F_{\alpha_E}(s) = \bigcap_{a\in E} \alpha_E(s,a)$.

$(F_{\alpha_E})^n(s)$ denotes the unique state that results from state s by taking n steps of the system strategy α_E

Universiteit Utrecht

broersen@cs.uu.nl

CKI master-colloquium 2006

ATL semantics

$\mathcal{M}, s \models \langle[A]\rangle\phi U^{ee}\psi \Leftrightarrow$

$\exists\beta_A$ such that $\forall\gamma_{\overline{A}}$ it holds that $\exists n > 0$ such that

(1) $\mathcal{M}, (F_{\beta_A|\gamma_{\overline{A}}})^n(s) \models \psi$ and

(2) $\forall i$ with $0 < i < n$ we have $\mathcal{M}, (F_{\beta_A|\gamma_{\overline{A}}})^i(s) \models \varphi$

Universiteit Utrecht

broersen@cs.uu.nl

ATEL

Epistemic extension ATEL defined in, [vd Hoek & Wooldridge: Studia Logica 2003]

- Language is extended with an operator K for knowledge
- Semantics is extended by introducing an epistemic indistinguishability (equivalence) relation over the set of worlds S
- Results in an S5 modal logic for the epistemic fragment.
- Was argued to have some counter-intuitive properties.

broersen@cs.uu.nl

The formal semantics in the fully strategic case

$\mathcal{M}, \alpha_A, s \models \Diamond_B \varphi \Leftrightarrow$
$\exists \beta_B$ such that $\mathcal{M}, \beta_B, s \models \varphi$

$\mathcal{M}, \alpha_A, s \models [B]\phi U^{ee} \psi \Leftrightarrow$
$\forall \beta_{\overline{A \cap B}}$ it holds that $\exists n > 0$ such that
(1) $\mathcal{M}, \alpha_A, (F_{\alpha_E})^n(s) \models \psi$ and
(2) $\forall i$ with $0 < i < n$ we have $\mathcal{M}, \alpha_A, (F_{\alpha_E})^i(s) \models \varphi$
where α_E is defined as: $\alpha_E = \alpha_A \restriction_{A \cap B} | \beta_{\overline{A \cap B}}$

The first condition keeps the state fixed, the second the strategy.

The crucial idea was to take the intersection of groups A and B in the strategy definitions above.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Results

- Semantic definition of a fully strategic STIT-operator (no such thing yet in the philosophical literature)
- ATL-STIT is a superset of ATL
- We have an alegant and sound axiomatization suspected to be complete (future work).
- A good solution to the problem of uniform strategies without making strategies explicit in the object language (see also the paper by Andreas Herzig and Nicolas Troquard at AAMAS'06)

More results on the relation between CL/ATL and STIT:

- Embedding of CL in Horty's CSTIT formalism (LCMAS '05)
- Embedding of ATL in Horty's strategic ability CSTIT (JLC 2006)

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

My opinion on George W. Bush

(introducing an operator D for `desire')

Of course: K_{bush} $\Diamond_{bush}$ [bush] F world-destruction,
because: $\Diamond_{bush}$ K_{bush} [bush] F world-destruction.

But although: ¬ D_{bush} [bush] F world-destruction,
and even: D_{bush} [bush] G ¬ world-destruction,
actually: [bush] F world-destruction,
while: ¬ K_{bush} [bush] F world-destruction.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

The End

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl

Back to the blind: how to approach his problem semantically? (1/2)

- We are going to use a STIT-like semantics for the [A] operator.
- However, we are not going to evaluate with respect to history / state pairs as in STIT theory, but with respect to strategy / state pairs (this is crucial!).
- The [A] X ϕ operator is evaluated by keeping the strategy fixed, and evaluating ϕ in the appropriate states.
- The $\Diamond_A$ ϕ operator is evaluated by keeping the state fixed, and evaluating ϕ in alternative strategies; $\Diamond_A$ is a modal strategy quantifier.
- So again we have a two-dimensional modal structure; one dimension of states, and one dimension of strategies.

CKI master-colloquium 2006
Universiteit Utrecht
broersen@cs.uu.nl