

Logics for Agency and Multi-Agent Systems

What agents can plan: ATL and strategic STIT

Jan Broersen
Andreas Herzig
Nicolas Troquard

ESLLI'07, Dublin, Friday, August 17th, 2007

Looking back

- Monday: introduction to modal logic
- Tuesday: **What agents can do**
 - Logic of ability: Coalition Logic
- Wednesday: **What agents do, what agents know they (can) do**
 - Logic of agency: STIT
- Thursday: **What agents want**
 - Logic of intention: Cohen and Levesque
- Friday: **What agents can plan**
 - Logic of plans: ATL, Strategic STIT, Epistemic strategic STIT

Logics for Agency and Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

2

Looking forward

- Alternating time temporal logic (ATL)
 - Semantics in terms of ATSS
 - CTL as a fragment
 - CL as a fragment
- Horty's logic for strategic STIT ability
- Strategic STIT: a proposal
- Adding epistemic modalities: Conformant strategic STIT
- Deontic ATL (time permitting)
- Future directions and conclusions

Logics for Agency and Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

3

Explaining the STIT-view on action

- STIT, CL and ATL share the same 'view' on **action**
- Actions** are seen as a relation between **agents** and **effects**: an **agent** performing an action **forces / ensures** the possible worlds to be among those that satisfy the **effect**.
- PDL has a different view: actions are seen as a relation between action names and effects.
- In both views, we can specify action preconditions using material implication.
- In PDL it is unclear how to introduce agency.

Logics for Agency and Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

4

From one step choice to strategic choice

- We will now assume **doing** involves a **series** of steps \Rightarrow **strategies**
- A bit misleading, since also in the CL context one can speak of strategies.
- Even more misleading, strategies might not be the right name, maybe 'tactics' or 'conditional plans' is better.
- In the one step case, we have independence of agents. Here we only have it with respect to single steps (see the axioms later on).

Logics for Agency and Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

5

The lame and the blind

- Andreas had the example with **the two-way switch**.
- A similar example for the strategic case is **'the lame and the blind'**
- The **lame** can toggle or skip:
 $\diamond[Lame]X \text{ toggled} \wedge \diamond[Lame]X \text{ L-skipped}$
- The **blind** can observe or skip:
 $\diamond[Blind]X \text{ observed} \wedge \diamond[Blind]X \text{ B-skipped}$
- Neither** can see to it that the light is on:
 $\neg \diamond[Lame]X \text{ light} \wedge \neg \diamond[Lame]X \text{ light}$
- But **jointly** they can, **strategically**:
 $\diamond[Lame, Blind]_{XX} \text{ light}$
- Note the **XX** is essential: they would not be able to do it in one step.

Logics for Agency and Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

6

Alternating Time Temporal Logic (ATL)

Features of ATL

- ATL [Alur & Henzinger & Kupferman 1997] Gains more and more popularity as a modeling language for multi-agent systems.
- Can be interpreted on **concurrent game structures** (CGSS / MGMs) and effectivity function semantics. 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.

More features of ATL

- In dynamic logic **action refinement** (as in HTN-planning) is hard to express. In ATL not.
- ATL is decidable (though EXPTIME-complete), and there is a model checker (Mocha).
- Ongoing research on the combination of ATL with **epistemic logic** (ATEL, vd Hoek & Wooldridge: Studia Logica 2003, Jamroga & Agotnes 2005, etc.).
- There are approaches to combining ATL with obligations [Jamroga & Van der Hoek & Wooldridge: DEON'04, Broersen: DEON'06].

ATL syntax (informally)

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 $[[A]] \eta$ ('A do not have a strategy ensuring $\neg \eta$ ' / 'whatever strategy A take, η is a possible outcome' / 'A cannot prevent η from happening')

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 ')

ATL syntax (formally)

$\varphi, \psi, \dots ::= p \mid \neg \varphi \mid \varphi \wedge \psi \mid \langle [A] \rangle \eta \mid [[A]] \eta$
 $\eta, \theta, \dots ::= \varphi U^e \psi$

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

Common Semantic Ground: (ATSS)

On **Alternating Transition Systems** we can interpret:

- Coalition logic,
- STIT-logics,
- Strategic STIT Logics,
- Alternating Time Temporal Logic (ATL),
- Computation Tree Logic (CTL),
- CTL*,
- ATL*,
- and many more.

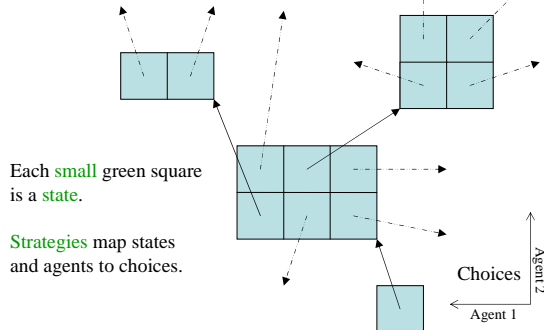
Do we give yet another semantics for ATL?

- **No.** The semantics in terms of ATSs was the first one for ATL.
- We **also** give semantics in terms of ATS, but only reformulate the 'tools' for defining truth on an ATS.
- The new notations enable us to define, in addition, a strategic version of STIT.
- Why no game models? Extra features of the models, like labels, obscure what the models represent in relation to what the language can say...

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 (earlier: AGT)
5. A **strategy** is a **total** function from states and agents to choices
6. Non-determinism of actions (choices) of agents is *only* due other agents making choices at the same state
7. from 6. it follows: **system actions** (actions performed by all agents simultaneously) are **deterministic**.
8. from 5. it follows: **system actions** are **serial**.

A two agent ATS



Alternating Transition Systems, formally

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

- (1) S are states.
- (2) $\mathcal{C} : E \times S \mapsto 2^{\mathcal{C}}$ yields for each agent and state a set of choices / actions. Furthermore, the function $R_X(s) = \{\bigcap_{a \in E} C \mid C \in \mathcal{C}(a, s)\}$ yields a non-empty set of singleton sets for each s .
- (3) π evaluates atomic propositions.

Strategies

A strategy α_a for an agent a , is a total function $\alpha_a : S \mapsto 2^S$, such that $\alpha_a(s) \in \mathcal{C}(a, s)$, assigning choices of the agent a to states. (note that we do not use a mapping from sequences to states)

Strategy functions α_a for individual agents a are straightforwardly extended to strategy functions $\alpha_A : S \mapsto 2^S$ for groups $A \subseteq E$ by $\alpha_A(s) = \bigcap_{a \in A} \alpha_a(s)$.

Furthermore, for $A \cap B = \emptyset$, we use $\alpha_A \parallel \beta_B$ to denote the joint strategy function built from intersecting the choices in α_A and β_B .

Restricting strategies to sub-groups

For $B \subseteq A$ the strategy $\alpha_A \upharpoonright_B$ denotes the strategy function that is the restriction of the strategy function α_A to the domain of agents B : $\alpha_A \upharpoonright_B(s) = \alpha_A(s) \cup \alpha_A^1(s) \cup \alpha_A^2(s) \dots$ where the $\alpha_A^j(s)$ are the *alternatives* for $\alpha_A(s)$ where *only* agents in $A \setminus B$ take *other* choices than they do in $\alpha_A(s)$

Algebraic properties:

$$\alpha_A \upharpoonright_A = \alpha_A$$

$$(\alpha_A \parallel \beta_B) \upharpoonright_A = \alpha_A \text{ and } (\alpha_A \parallel \beta_B) \upharpoonright_B = \beta_B$$

$$(\alpha_A \upharpoonright_C \parallel \beta_B \upharpoonright_C) = (\alpha_A \parallel \beta_B) \upharpoonright_C$$

Semantics: a follow-up function

Clearly, system strategies $\alpha_E : S \mapsto 2^S$ can be seen as the 'histories' of an ATS.

Thus, for a system strategy α_E , the *unique* follow-up state of a state s is given by $\alpha_E(s)$.

$(\alpha_E)^n(s)$ denotes n applications of the follow-up function, starting from s

Reformulating the CL semantics

We can now easily give a truth definition for the **Coalition Logic** modality.

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

Other notions of ability then in CL might be thought of.

Why is this equivalent to the neighborhood / effectivity function semantics we saw Tuesday?

ATL semantics

We can now **also** easily give a truth definition for the ATL modality.

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

$\exists \beta_A$ such that $\forall \gamma_{\bar{A}}$ it holds that $\exists n \geq 1$ such that

- (1) $\mathcal{M}, (\beta_A \parallel \gamma_{\bar{A}})^n(s) \models \psi$ and
- (2) $\forall i$ with $1 \leq i < n$ we have $\mathcal{M}, (\beta_A \parallel \gamma_{\bar{A}})^i(s) \models \varphi$

CL and CTL as subsets of ATL

- By substituting \perp for ψ , we immediately see that CL is **embedded** in ATL
- CTL is also embedded, through the definitions:

$$E(\varphi U^e \psi) \equiv_{def} \langle [E] \rangle (\varphi U^e \psi)$$

$$A(\varphi U^e \psi) \equiv_{def} \langle [\emptyset] \rangle (\varphi U^e \psi)$$

ATL: axiomatization

- (T) an axiomatization for propositional logic
 (\perp) $\neg \langle [A] \rangle X \perp$
 (T) $\langle [A] \rangle X \top$
 (N) $\neg \langle [\emptyset] \rangle \varphi \rightarrow \langle [A] \rangle X \varphi$
 (S) $\langle [A_1] \rangle X \varphi \wedge \langle [A_2] \rangle X \psi \rightarrow \langle [A_1 \cup A_2] \rangle X (\varphi \wedge \psi)$ for $A_1 \cap A_2 = \emptyset$
 (FP_G) $\langle [A] \rangle G \varphi \leftrightarrow \varphi \wedge \langle [A] \rangle X \langle [A] \rangle G \varphi$
 (GFP_G) $\langle [\emptyset] \rangle G (\theta \rightarrow (\varphi \wedge \langle [A] \rangle X \theta)) \rightarrow \langle [\emptyset] \rangle G (\theta \rightarrow \langle [A] \rangle G \varphi)$
 (FP_F) $\langle [A] \rangle \psi U \varphi \leftrightarrow \varphi \vee (\psi \wedge \langle [A] \rangle X \langle [A] \rangle \psi U \varphi)$
 (LFP_F) $\langle [\emptyset] \rangle G ((\varphi \vee (\psi \wedge \langle [A] \rangle X \theta)) \rightarrow \theta) \rightarrow \langle [\emptyset] \rangle G (\langle [A] \rangle \psi U \varphi \rightarrow \theta)$

Modus Ponens
 $\langle [A] \rangle$ Monotonicity
 $\langle [\emptyset] \rangle G$ Necessitation

Strategic ability STIT

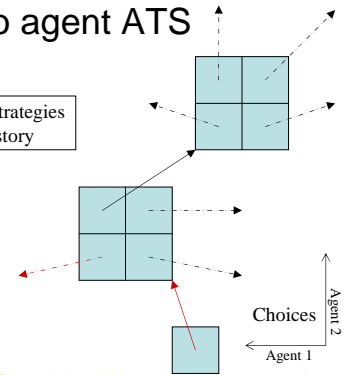
- CL and ATL were developed independently. But there was yet another independent development, in the philosophical literature [Agency and Deontic Logic, John Horty 2001]
- Horty discusses whether STIT can be generalized to the strategic case.
- He manages to define a notion of **strategic stit ability**, which is very close to ATL. [Embedding Alternating-time Temporal Logic in Strategic STIT Logic of Agency, Jan Broersen, Andreas Herzig and Nicolas Troquard, Journal of Logic and Computation, 2006.]

Strategic STIT?

- Classical STIT: an agent sees to it that ϕ if he selects a **one step choice** that ensures ϕ
- Strategic STIT: an agent strategically sees to it that ϕ if he performs a **strategy** that ensures ϕ
- **Problem:** which history an agent performs is not determined by a history-state pair \Rightarrow one history may belong to several strategies.
- **Horty's solution:** quantify over histories \Rightarrow strategic ability
- Our solution: take strategies as the objects (worlds) on which to evaluate formulas.

A two agent ATS

Agent 1 has at least two strategies consistent with the red history



Differences between Horty's proposal and ATL

Similarities: independence of agents, fused operators, temporal structures.

Differences: Horty has no discrete time, Horty assumes trees, ATL has no 'fields', ATL has no past modalities, ATL does not evaluate with respect to moment-history pairs.

None of the differences are really essential.

Can we do all the same we did before, strategically?

- Can we define a notion of strategic STIT?
- Can we add epistemic modalities?
- Can we preserve decidability?
- Can we find complete axiomatizations?

CL-STIT semantics: truth definition

Take the following semantics for the CL-STIT-operators for historical possibility \diamond and 'seeing to it that next' $[A]X\phi$:

$$\mathcal{M}, \alpha_E, s \models \diamond\phi \Leftrightarrow \exists \beta_E \text{ such that } \mathcal{M}, \beta_E, s \models \phi$$

$$\mathcal{M}, \alpha_E, s \models [A]X\phi \Leftrightarrow \forall \beta_{\bar{A}} \text{ if } \gamma_E = \alpha_E \upharpoonright_A \parallel \beta_{\bar{A}} \text{ then } \mathcal{M}, \gamma_E, \gamma_E(s) \models \phi$$

Properties

- The CL operator is definable: $\langle [A] \rangle X\phi \equiv_{\text{def}} \diamond [A]X\phi$ (easy to see that for this operator evaluation with respect to α_E is no longer necessary)
- Historical necessity is not definable, but: $\square [E]X\phi \leftrightarrow [\emptyset]X\phi$
- Note that this version of STIT takes effect in the next state: $[A]X\phi \rightarrow [E]X\phi$ while, for Chellas STIT, we have the instantaneous: $[A]\phi \rightarrow \phi$
- Our instantaneous versions, where historical necessity and CL are definable \Rightarrow G-STIT / NCL presented by Nicolas on Wednesday

Semantics for strategic STIT

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

$\mathcal{M}, \alpha_A, s \models [B] \phi U^{ee} \psi \Leftrightarrow$
 $\forall \beta_{A \cap B}$ it holds that $\exists n \geq 1$ such that
 if $\alpha_E = \alpha_A \upharpoonright_{A \cap B} \parallel \beta_{A \cap B}$ then
 (1) $\mathcal{M}, \alpha_A, (\alpha_E)^i(s) \models \psi$ and
 (2) $\forall i$ with $1 \leq i < n$ we have $\mathcal{M}, \alpha_A, (\alpha_E)^i(s) \models \varphi$

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

Intersection A and B empty \Rightarrow property is historically necessary

Comments

- Strategies have become possible worlds.
- Several validities.
- ATL is definable as:

$$\langle [A] \rangle \eta \equiv_{def} \diamond_A [A] \eta$$

- So now we have a logic (semantically defined) that embeds CL, ATL, CTL, and has a notion of strategic STIT.

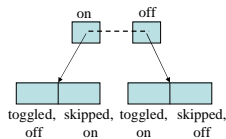
Philosophical questions

- What can we make of the fact that we always evaluate with respect to an actual strategy?
- Does this mean the agents have committed to this strategy?
- Answer 1: **no**. There are also versions of CTL* where this is the case.
- Answer 2: **no**. It might be seen as an underlying assumption of determinism of the world. If you do not like it, you should add a believe operator.

Ability under uncertainty: adding epistemics

Uniform Strategies

A blind person enters 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 epistemic CL we can only say $K_{\text{blind}} \langle [\text{blind}] \rangle X \text{On}$.

Problem: we cannot talk about **uniform strategies**.

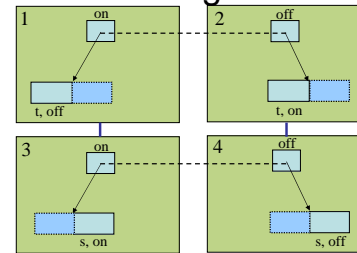
Analysis of the problem

- Solution: break down the original operator $\langle [A] \rangle X$ into its parts $\diamond [A] X$ as we did for CL-STIT and X-STIT.
- Not having a **uniform strategy** is then simply expressed as:
 $\neg \diamond K_{\text{blind}} [\text{blind}] X \text{On}$
- Conformantly (that is uniformly) seeing to it is expressed as: $K_{\text{doctor}} [\text{doctor}] X \text{patient_cured}$
- Add an epistemic uncertainty relation between history-state pairs.

How to picture histories as possible worlds?

- \Rightarrow make a 'world' for each history state pair
- Picture next slide:
 1. 'black arrows' interpret $[A]X$
 2. 'blue lines' interpret \diamond
 3. 'dotted lines' interpret K_A

An epistemic model for the blind agent



4 history-state pairs

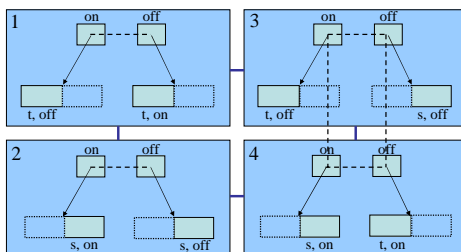
- 1 & 2: Blind toggles without knowing the state
- 3 & 4: Blind skips without knowing the state

Formulas true in the model

- Everywhere it holds that $K_{\text{blind}} \diamond [\text{blind}] X \text{On}$
- Everywhere it holds that $\neg \diamond K_{\text{blind}} [\text{blind}] X \text{On}$
- Depending on the actual history / state pair, for the blind agent it may hold that On or $\neg \text{On}$ and $[\text{blind}] X \text{On}$ or $\neg [\text{blind}] X \text{On}$

Knowing how: the strategic case

A model for the blind agent



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

Philosophical issues

- What is the meaning of an agent knowing a strategy? Agents cannot know the future!
- Opinion: not a problem of strategic STIT, but a problem of how we use it.
- **Knowledge** is a less relevant concept for agents anyway \Rightarrow we should actually consider **belief** in all the examples we had so far.

Open questions (1/2)

- Why do so many researchers assume that for modeling 'knowing how' we need the names of choices in the object language (like in PDL)? Is this because they are computer scientists?
- Interpreting NCL on ATs?
- Formal properties for the strategic case?

Logics for Agency and
Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

43

Open questions (2/2)

- Can we add an account of belief revision?
- Can we add an account of preference revision? (intention revision following from that)
- How to account for intentions in the strategic case?
- How do we account for deontic modalities?
- Logics for games: we have actions, strategies, knowledge, but still no preferences (there are many different ways to do this)

Logics for Agency and
Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

44

Intention

- Intentions are in between desires and actions.
- They are needed / used to prevent the agent from continually reconsidering his actions \Rightarrow stabilization of deliberation
- Since strategies are possible worlds, it is easy (semantically) to define a notion of strategic intention.
- Reasoning about action refinement.

Logics for Agency and
Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

45

Deontic ATL

If time permits

Temporal Aspects of Obligation

Claim: Temporal aspects of obligation (and, in general, motivational attitudes) are often neglected, and not well understood.

- There is a huge difference between having to achieve something **now**, **eventually**, **before something else**, **always** or **repeatedly** (to name a few possibilities). In SDL we can only say ' $O\phi$ '.
- $O\phi \wedge O\neg\phi$ can become consistent if we are able to express the temporal aspects of ϕ .
- Conflicts like the famous Chisholm-set may get consistent.

Logics for Agency and
Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

47

Intuitive properties not formalized in temporal deontic logics

- **First example:** if I am obliged to achieve ρ in the future, and if I do not achieve it now, then, next I am still obliged to achieve ρ in the future.
- **Second example:** presently I should not perform actions (choices) that cut me off from complying to obligations concerning the not-immediate future.

Logics for Agency and
Multi-Agent Systems

Jan Broersen, Andreas Herzig,
Nicolas Troquard

48

Can agents violate achievement obligations?

Examples motivating the need for 'deadlines'.

1. Can an agent violate an obligation to stop smoking some time in the future?
2. Is the obligation to return a library book some time in the future violated by burning the book?
Suggestion: study this as a problem of the interaction with ability.

Specific problem description

What can we say about: the *semantics*, and the *logic* of an operator $O_A(p \leq \delta : \xi_A)$ for 'agents *A* are obliged to meet condition *p* before condition δ otherwise they suffer negative condition ξ_A '

- Formulas ξ_A express conditions which are thought to be negative for *A*, independent of time.
- To comply to $O_A(p \leq \delta : \xi_A)$ it is sufficient that *A* satisfy *p* once, at a time of their choosing, provided it is before or (ultimately) at δ .
- conditions δ are not necessarily deadlines. We have an obligation concerning the order of conditions.
- The condition δ is not guaranteed to occur!
- The condition δ in the operator $O_A(p \leq \delta : \xi_A)$ constrains the reference time of the obligation.

Reduction approaches for obligation

Examples of reduction approaches:

- *Anderson* reduces Standard Deontic Logic (SDL) to alethic modal logic + violation constants
 $O_a\phi \equiv_{\text{def}} \Box(\neg\phi \rightarrow \text{Viol}(a))$
- *J-J Meyer* reduces deontic action logic to (a special variant of) dynamic logic + violation constants
 $O_a(\alpha) \equiv_{\text{def}} [\sim\alpha] \text{Viol}(a)$

The advantage is clear: we can do the formal reasoning (tableaux, model checking, etc.) in the logic reduced to.

Define $O_A(p \leq \delta : \xi_A)$ in the logic *ATL* as a *syntactic abbreviation* whose semantics corresponds to the intended interpretation.

Interaction with ability / control

Kant: we will never have a sense of ought (in the sense of his categorical imperatives) about things we cannot do.

Practical obligations: A ought ϕ implies that ...

1. A can achieve ϕ (Kant for 'profane' obligations)
2. A can avoid ϕ (ϕ is not 'settled' / complying is 'deliberate')
3. A ought not to achieve that he can never achieve ϕ (backwards propagation of obligations: planning!)

Reducing $O_A(p \leq \delta : \xi_A)$

- The main idea behind the reduction is:
 $O_A(p \leq \delta : \xi_A)$ holds if and only if it is not the case that the agents *A* have a strategy to achieve δ , to avoid *p* at all moments until the first time δ , and avoid the negative condition ξ_A at the point where δ .

- Characterization in ATL:

$$(\delta \wedge (\neg p \rightarrow \xi_A)) \vee \neg \langle [A] \rangle ((\neg p \wedge \neg \delta) U^e (\delta \wedge \neg p \wedge \neg \xi_A))$$

Consequences of the definition

- The bad strategies are those where *A* do not accomplish *p* before δ . That they are bad is represented by the property that ξ_A is not guaranteed to be avoided at the points where δ .
- We can also imagine a situation where only some of the bad strategies possibly give rise to a ξ_A at the points where δ . In that case we have conditionality of the obligation with respect to certain sets of histories (future research).

Persistence Property

- Validity time: the obligation is discarded by either reaching the deadline, or the achievement:

$$\models_{\text{ATL}} O_A(\rho \leq \delta) \rightarrow \langle A \rangle (O_A(\rho \leq \delta) \cup_w (\rho \vee \delta))$$

The issue of 'settledness'

- $\models_{\text{ATL}} \rho \rightarrow O_A(\rho \leq \delta)$,
- Or, more in particular:
- $\models_{\text{ATL}} \neg \langle A \rangle (\neg \rho \cup \delta) \rightarrow O_A(\rho \leq \delta)$
- Things that are (possibly) settled, are obliged!

The problem is an instance of a *general problem* for 'reduction approaches' to *deontic logic*:

- In Meyer's dynamic deontic logic we have $\langle \sim \alpha \rangle \perp \rightarrow O(\alpha)$
- In Anderson's reduction we have $\Box \phi \rightarrow O\phi$

Eliminated by considering interaction with ability

Define a *deliberate* version:

$$O_A^{dl}(\rho \leq \delta : \xi_A) \equiv_{def} O_A(\rho \leq \delta : \xi_A) \wedge \neg O_A(\rho \leq \delta : \perp)$$

If A have an obligation, they should at least be able / have a strategy to violate it.

This also eliminates:

- $\models_{\text{ATL}} \langle A \rangle G \neg \delta \rightarrow O_A(\rho \leq \delta)$

Other interactions with abilities

A variant with 'ought implies can'

$$O_A^{oc}(\rho \leq \delta : \xi_A) \equiv_{def} O_A(\rho \leq \delta : \xi_A) \wedge \langle A \rangle (\neg \delta U \rho)$$

Ought implies ought not to achieve that cannot (? better suggestions welcome):

$$O_A^{pl}(\rho \leq \delta : \xi_A) \equiv_{def} O_A(\rho \leq \delta : \xi_A) \wedge \langle A \rangle (\langle A \rangle F \rho \vee \xi_A)$$

Yet more variants

We can also define a *stronger* form of obligation, where the complementary agents $E \setminus A$ have the power to impose the negative condition ξ_A on A:

$$O_A^5(\rho \leq \delta : \xi_A) \equiv_{def} \neg \langle E \setminus A \rangle (\neg \rho U (\delta \wedge \neg \xi_A))$$

$$O_A^6(\rho \leq \delta : \xi_A) \equiv_{def} \dots$$

etc.

Maintenance obligations

maintenance properties ϕ come with a property ψ functioning as a *relief condition*: if the relief condition occurs, the obligation to maintain ϕ no longer holds.

We can *define* this through:

$$O_A(\varphi \text{ unt } \psi : \xi_A) \equiv_{def} O_A(\psi \leq \neg \varphi : \xi_A)$$

Logical properties (1)

Weakening, strengthening, agglomeration, detachment:

$$\begin{aligned} &\models O_A((\rho \wedge \chi) \leq \delta) \rightarrow O_A(\rho \leq \delta) \\ &\not\models O_A(\rho \leq \delta) \rightarrow O_A(\rho \leq (\delta \wedge \gamma)) \\ &\not\models O_A(\rho \leq \delta) \wedge O_A(\chi \leq \delta) \rightarrow O_A((\rho \wedge \chi) \leq \delta) \\ &\not\models O_A(\rho \leq \delta) \wedge O_A(\rho \leq \gamma) \rightarrow O_A(\rho \leq (\delta \wedge \gamma)) \\ &\not\models O_A(\rho \leq \delta) \wedge O_A(\rho \leq \gamma) \rightarrow O_A(\rho \leq (\delta \vee \gamma)) \\ &\not\models O_A(\rho \leq \delta) \wedge O_A(\delta \leq \gamma) \rightarrow O_A(\rho \leq \gamma) \\ &\models O_A^d(\rho \leq \delta) \wedge O_A^d(\delta \leq \gamma) \rightarrow O_A^d(\rho \leq \gamma) \end{aligned}$$

Logical properties (2)

$$\begin{aligned} &\models \xi_A \rightarrow O_A(\rho \leq \top : \xi_A) \\ &\models O_A(\rho \leq \delta : \top) \\ &\models O_A(\rho \leq \delta : \perp) \rightarrow O_A(\rho \leq \delta : \xi_A) \\ &\models O_A(\rho \leq \delta) \rightarrow [(A)](O_A(\rho \leq \delta)U_w(\rho \vee \delta)) \\ &\models O_A(\rho \leq \delta) \rightarrow O_A(O_A(\rho \leq \delta) \text{ unt } (\rho \vee \delta)) \end{aligned}$$

Logical properties (3)

$$\begin{aligned} &\not\models O_A(\rho \leq \delta) \rightarrow O_A(\delta \leq \rho) \\ &\models O_A(\gamma \leq \gamma) \\ &\models O_A(\top \leq \delta) \quad \not\models O_A(\perp \leq \delta) \quad \not\models \neg O_A(\perp \leq \delta) \\ &\models O_A(\rho \leq \perp) \quad \not\models O_A(\rho \leq \top) \quad \not\models \neg O_A(\rho \leq \top) \\ &\not\models \neg O_A(\perp \leq \top : \xi_A) \quad \not\models O_A(\perp \leq \top : \xi_A) \\ &\models \neg O_A(\perp \leq \top : \perp_A) \\ &\models [(A)]G\neg\delta \rightarrow O_A(\rho \leq \delta) \\ &\not\models \neg(O_A(\rho \leq \delta) \wedge O_A(\neg\rho \leq \delta)) \\ &\models \neg(O_A(\rho \leq \top) \wedge O_A(\neg\rho \leq \top)) \end{aligned}$$

The relation with SDL

Standard Deontic logic (SDL) is simulated.

Theorem:

The logic of $O_A(\rho \leq \top : \perp)$ is standard deontic logic (the modal logic KD)

Conclusions

- By joining (fusing) the intention part with the NCL part, we already have a complete logic for 1-step **action + knowledge + intention** (via Sahlqvist), but not yet for the same combination with strategies (because of the temporal operators).
- On the level of strategies, many formal properties are still unknown.

Thank you for listening