

Logics of individual and
collective intentionality
(Lecture V: Group action)

Andreas Herzig, **Emiliano Lorini**

ESSLLI'09

Bordeaux, 20-25 July 2009

Introduction

Joint and group action

⇒ Examples of group action

- ▶ Bill and Bob are painting a house together.
- ▶ Brazil soccer team can win against Italy soccer team.
- ▶ Ann and Mary could have avoided the accident (if they were more cautious).
- ▶ Bill and Bob have the intention to write a paper together, and they start to write it.

Logics of joint and group action

- ▶ Logics for social software: Coalition Logic (Pauly 2001, 2002)
- ▶ Logics for multi-agent systems: Alternating-time Temporal Logic (Alur and Henzinger, 2002; van der Hoek and Wooldridge, 2003), Coalition Logic of Propositional Control (van der Hoek and Wooldridge, 2005)
- ▶ Philosophy of action: “Seeing To It That” (Belnap et al. 2001)

Introduction

Coalition Logic

STIT logic with agents and groups

Towards intentional STIT: from uniform strategies to joint intentions

Coalition Logic

Coalition logic (Pauly 2001, 2002)

- ▶ Social software: logics for modelling procedures involving the interactions between multiple agents
- ▶ E.g., voting procedure

- ▶ It enables to express what a coalition of agents can ensure by doing a joint action

Coalition logic (CL): language

- ▶ $AGT = \{1, \dots, n\}$: a countable set of agents;
- ▶ ATM : a countable set of atomic propositions.

Coalition logic (CL): language (cont.)

The language \mathcal{L}^{CL} of CL with agents and groups is defined by the following BNF:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \langle\!\langle J \rangle\!\rangle\varphi$$

where p ranges over ATM and $J \subseteq AGT$.

\Rightarrow We write $\langle\!\langle i \rangle\!\rangle\varphi$ instead of $\langle\!\langle \{i\} \rangle\!\rangle\varphi$

- ▶ $\langle\!\langle J \rangle\!\rangle\varphi$: the coalition J can ensure φ at the next time point by acting together, whatever the others agents do.
 - ▶ \exists a collective choice of J s.t. \forall next state φ holds.
- ▶ $\langle\!\langle \emptyset \rangle\!\rangle\varphi$: φ is necessarily true at the next time point.

Example: coordinated attack

- ▶ Two agents i and j are trying to move an attack against a common enemy. The enemy will be defeated iff i and j move a coordinated attack (both i and j attack the enemy).
 - ▶ i has two actions available: *attack* and *skip* (do nothing).
 - ▶ j has two actions available: *attack* and *skip* (do nothing).

In formulas:

$$\langle\langle\{i, j\}\rangle\rangle defeatEnemy \wedge \neg\langle\langle i \rangle\rangle defeatEnemy \wedge \neg\langle\langle j \rangle\rangle defeatEnemy$$

Effectivity functions

- ▶ S : a set of states
- ▶ $e : 2^{AGT} \longrightarrow 2^{2^S}$: effectivity function
 - ▶ $X \in e(J)$ iff is a set of possible outcomes for which J is effective (or J can force the world to be in some state of X at the next step)

Example: coordinated attack (cont.)

- ▶ $S = \{s_{\text{Defeat}}, s_{\text{Undefeat}}\}$
- ▶ $e(i) = e(j) = \{\{s_{\text{Defeat}}, s_{\text{Undefeat}}\}\}$
- ▶ $e(\{i, j\}) = \{\{s_{\text{Defeat}}, s_{\text{Undefeat}}\}, \{s_{\text{Defeat}}\}, \{s_{\text{Undefeat}}\}\}$

Playable effectivity function

e is **playable** iff:

1. $\emptyset \notin e(J)$;
2. $S \in e(J)$;
3. $S \setminus X \notin e(\emptyset)$ then $X \in e(AGT)$
(**AGT-maximality**);
4. if $X_1 \in e(J)$ then $X_1 \cup X_2 \in e(J)$
(**Outcome monotonicity**);
5. if $J \cap I = \emptyset$ then if $X_1 \in e(J)$ and $X_2 \in e(I)$ then
 $X_1 \cap X_2 \in e(J \cup I)$
(**Superadditivity**).

Remark

If e is **playable** then:

- ▶ e is **coalition monotonic**: if $J \subseteq I$ then $e(J) \subseteq e(I)$
- ▶ e is **regular**: if $X \in e(J)$ then $S \setminus X \notin e(AGT \setminus J)$

A CL-model is a tuple $M = ((S, E), V)$ where:

- ▶ S is a set of states;
- ▶ $E : S \longrightarrow (2^{AGT} \longrightarrow 2^{2^S})$ associates an effectivity function E_s to every state s in S ;
- ▶ V is a valuation function, that is, $V : S \longrightarrow 2^{ATM}$.

- ▶ Truth conditions for Boolean constructions are entirely standard.
- ▶ $M, s \models \langle J \rangle \varphi$ iff $\{s' \mid M, s' \models \varphi\} \in E_s(J)$.

Validity, satisfiability are defined as usual

A complete axiomatization of CL

- (RE) If $\varphi \leftrightarrow \psi$ then $\langle J \rangle \varphi \leftrightarrow \langle J \rangle \psi$
- (M) $\langle J \rangle (\varphi \wedge \psi) \rightarrow (\langle J \rangle \varphi \wedge \langle J \rangle \psi)$
- (\perp) $\neg \langle J \rangle \perp$
- (\top) $\langle J \rangle \top$
- (AGT) $\neg \langle \emptyset \rangle \varphi \rightarrow \langle AGT \rangle \neg \varphi$
- (S) $(\langle J \rangle \varphi \wedge \langle I \rangle \psi) \rightarrow \langle J \cup I \rangle (\varphi \wedge \psi)$ if $J \cap I = \emptyset$

$\langle\emptyset\rangle$ and $\langle AGT\rangle$ are normal modalities

- ▶ $\vdash (\langle\emptyset\rangle\varphi \wedge \langle\emptyset\rangle\psi) \rightarrow \langle\emptyset\rangle(\varphi \wedge \psi)$
- ▶ $\vdash (\langle AGT\rangle\varphi \wedge \langle AGT\rangle\psi) \rightarrow \langle AGT\rangle(\varphi \wedge \psi)$
- ▶ **If $\vdash \varphi$ then $\vdash \langle J\rangle\varphi$**
- ▶ **proof:**
 1. $\vdash \varphi$ Hypothesis
 2. $\vdash \varphi \leftrightarrow \top$ from 1
 3. $\vdash \langle J\rangle\varphi \leftrightarrow \langle J\rangle\top$ from 2 by Rule RE
 4. $\vdash \langle J\rangle\top$ Axiom \top
 5. $\vdash \langle J\rangle\varphi$ from 3 and 4

Some CL validities (cont.)

$\langle\emptyset\rangle$ and $\langle AGT \rangle$ are inter-definable

- ▶ $\vdash \langle\emptyset\rangle\varphi \leftrightarrow \neg\langle AGT \rangle\neg\varphi$
- ▶ proof:
 1. $\vdash \neg\langle AGT \rangle\neg\varphi \rightarrow \langle\emptyset\rangle\varphi$ by Axiom AGT
 2. $\vdash (\langle\emptyset\rangle\varphi \wedge \langle AGT \rangle\neg\varphi) \rightarrow \langle AGT \rangle\perp$ by Axiom S
 3. $\vdash \langle AGT \rangle\perp \rightarrow \perp$ by Axiom \perp
 4. $\vdash (\langle\emptyset\rangle\varphi \wedge \langle AGT \rangle\neg\varphi) \rightarrow \perp$ from 2,3
 5. $\vdash \langle\emptyset\rangle\varphi \rightarrow \neg\langle AGT \rangle\neg\varphi$ from 4

Two disjoint coalitions cannot bring about conflicting effects

- ▶ $\vdash (\langle J \rangle\varphi \wedge \langle I \rangle\neg\varphi) \rightarrow \perp$ if $I \cap J = \emptyset$
- ▶ proof:
 1. $\vdash (\langle J \rangle\varphi \wedge \langle I \rangle\neg\varphi) \rightarrow \langle J \cup I \rangle\perp$ by Axiom S
 2. $\vdash \langle J \cup I \rangle\perp \rightarrow \perp$ by Axiom \perp
 3. $\vdash (\langle J \rangle\varphi \wedge \langle I \rangle\neg\varphi) \rightarrow \perp$ from 1,2

STIT logic with agents and groups

STIT: the logic of “Seeing to it that”

- ▶ Formal philosophy and philosophy of action (Belnap et al., 2001)
- ▶ It makes a difference between *doing* (seeing to it that) and *being able to do* (being able to see to it that)
 - ▶ what an agent or coalition choose
 - ▶ what an agent or coalition can choose
- ▶ ‘Bringing it about tradition’ (Kanger, 1972; Pörn, 1977): only *doing*.

STIT logic with agents and groups

- ▶ Non-standard semantics for STIT in terms of *moments* and *histories* (Belnap et al., 2001; Horty, 2001; Horty & Belnap, 1995).
- ▶ It is proved in (Balbiani et al. 2008; Herzig & Schwarzentruher, 2008) that STIT can be ‘simulated’ in a standard Kripke semantics.
 - ▶ We use this for today’s presentation.

- ▶ $AGT = \{1, \dots, n\}$: a countable set of agents;
- ▶ ATM : a countable set of atomic propositions;
- ▶ $2^{AGT*} = 2^{AGT} \setminus \emptyset$: the set of non-empty *coalitions*.

STIT logic with agents and groups

The language \mathcal{L}^{STIT} of STIT with agents and groups is defined by the following BNF:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid [J]\varphi \mid \Box\varphi$$

where p ranges over ATM and J over 2^{AGT^*} .

$\Rightarrow [J]\varphi \approx$ “ J sees to it that φ no matter what the other agents in $AGT \setminus J$ do”.

$\Rightarrow \Box\varphi \approx$ “ φ is necessarily true” (historic necessity).

Further notations

$$\langle J \rangle \varphi \stackrel{\text{def}}{=} \neg [J] \neg \varphi$$

$$\diamond \varphi \stackrel{\text{def}}{=} \neg \Box \neg \varphi$$

- ▶ We write $[i]$ instead of $[\{i\}]$;
- ▶ $\diamond \varphi \approx$ “ φ is possibly true”;
- ▶ $\diamond [J] \varphi \approx$ “ J can see to it that φ whatever the other agents in $AGT \setminus J$ do”.

Example: two agents $AGT = \{1, 2\}$ taking care of a plant

Each agent can choose to *water the plant* or to *do nothing*

- ▶ If both of them water the plant then the plant will die.
- ▶ If both of them do nothing then the plant will die.
- ▶ If one waters the plant and the other does nothing then the plant will survive.
- ▶ 1 decides to water the plant and 2 decides to do nothing.

⇒ In formulas:

$$\begin{aligned} & \diamond[\{1, 2\}] \textit{alive} \wedge \diamond[\{1, 2\}] \neg \textit{alive} \wedge \neg \diamond[1] \textit{alive} \wedge \neg \diamond[2] \textit{alive} \wedge \\ & [\{1, 2\}] \textit{alive} \wedge \neg [1] \textit{alive} \wedge \neg [2] \textit{alive} \end{aligned}$$

STIT models

A STIT-model is a tuple $M = (W, \{R_J\}_{J \subseteq AGT}, H, V)$ where:

- ▶ W is a non-empty set of possible worlds or states;
- ▶ For all $J \in 2^{AGT^*}$, R_J is an equivalence relation over W :
 1. **Reflexive:** $(w, w) \in R_J$;
 2. **Transitive:** if $(w, v) \in R_J$ and $(v, u) \in R_J$ then $(w, u) \in R_J$;
 3. **Symmetric:** if $(w, v) \in R_J$ then $(v, w) \in R_J$.
- ▶ H is an equivalence relation over W .
- ▶ V is a valuation function, that is, $V : W \longrightarrow 2^{ATM}$.

$\Rightarrow R_J(w) = \{v \in W \mid (w, v) \in R_J\}$ is the set of outcomes of the action chosen by coalition J at w

$\Rightarrow H(w) = \{v \in W \mid (w, v) \in H\}$ is the set of possible outcomes at w

Constraints on STIT models

(OutcomeInc) for all $w \in W$, $R_J(w) \subseteq H(w)$;

(PointWInter) for all $w \in W$, $R_J(w) = \bigcap_{j \in J} R_{\{j\}}(w)$;

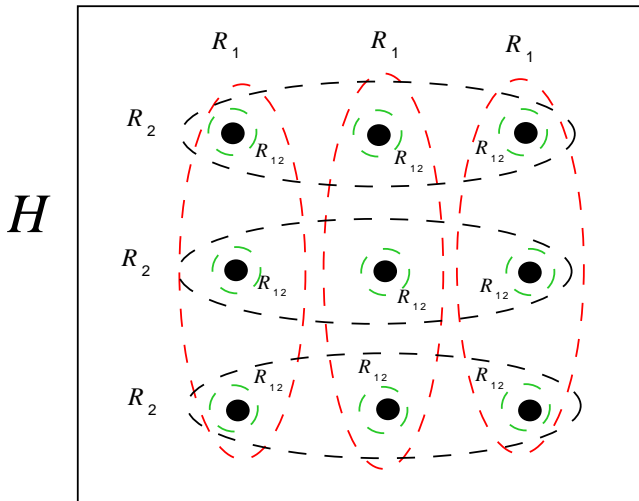
(Indep) for all $w \in W$, for all $\langle w_j \rangle_{j \in AGT} \in R_\emptyset(w)^n$,
 $\bigcap_{j \in AGT} R_{\{j\}}(w_j) \neq \emptyset$;

(AGT-Det) for all $w \in W$, if $v \in R_{AGT}(w)$ and $u \in R_{AGT}(w)$
then $v = u$.

Constraints on STIT models (cont.)

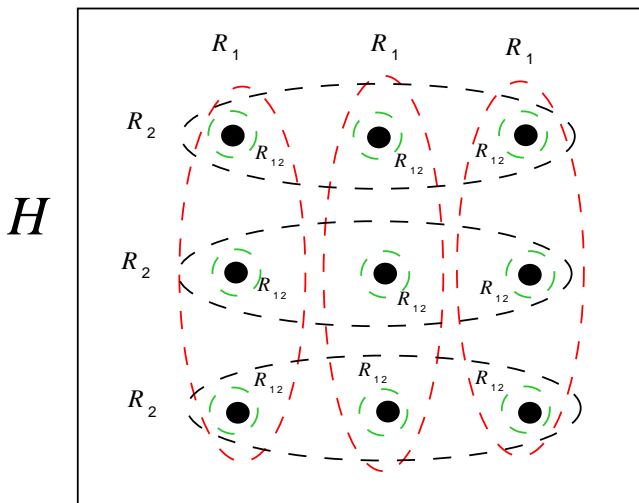
Suppose $AGT = \{1, 2\}$

(OutcomeInc) for all $w \in W$, $R_J(w) \subseteq H(w)$



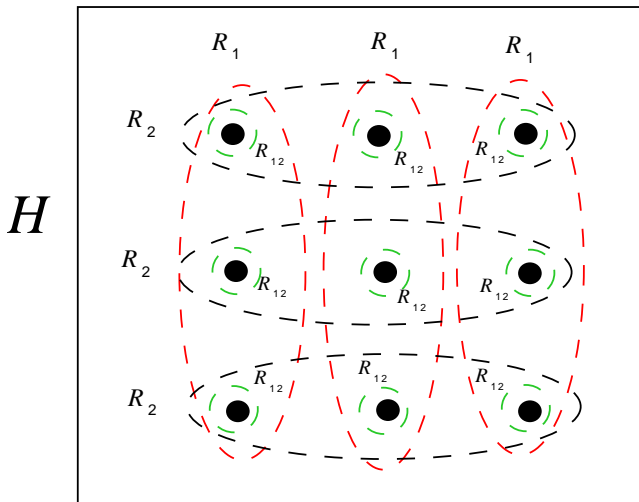
Constraints on STIT models (cont.)

(PointWinter) for all $w \in W$, $R_J(w) = \bigcap_{j \in J} R_{\{j\}}(w)$



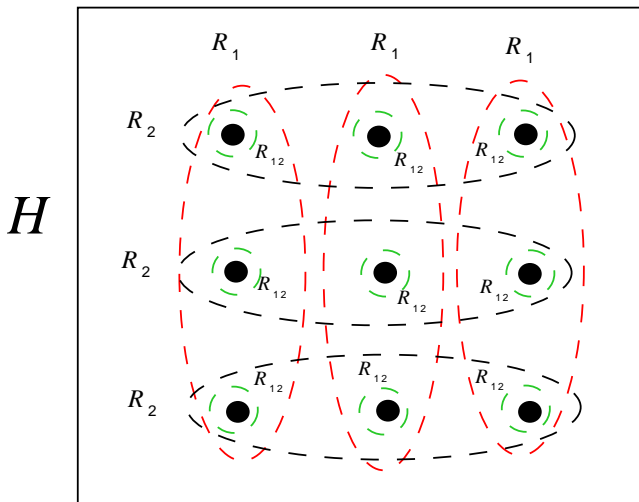
Constraints on STIT models (cont.)

(Indep) for all $w \in W$, for all $\langle w_j \rangle_{j \in AGT} \in R_\emptyset(w)^n$,
 $\bigcap_{j \in AGT} R_{\{j\}}(w_j) \neq \emptyset$



Constraints on STIT models (cont.)

(*AGT-Det*) for all $w \in W$, if $v \in R_{AGT}(w)$ and $u \in R_{AGT}(w)$
then $v = u$



Remark

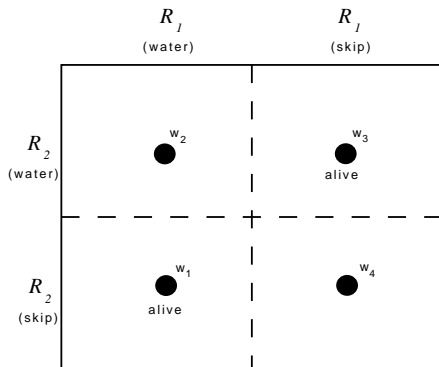
Slightly different from Horty's formulation. He does not suppose AGT-determinism.

Remark

*Constraints (OutcomeInc) and (PointWInter) imply:
 $R_J(w) \subseteq R_I(w)$ if $I \subseteq J$.*

- ▶ Truth conditions for Boolean operators are standard.
- ▶ $M, w \models [J]\varphi$ iff $M, v \models \varphi$ for all $v \in R_J(w)$.
- ▶ $M, w \models \Box\varphi$ iff $M, v \models \varphi$ for all $v \in H(w)$.

Two agents $AGT = \{1, 2\}$ taking care of a plant (cont.)



- ▶ $\{1, 2\} \text{ alive}$ is true at w_1 and w_3 .
- ▶ $\diamond\{1, 2\} \text{ alive}$ is true at w_1, w_2, w_3, w_4 .
- ▶ $\neg\diamond[1] \text{ alive}$ and $\neg\diamond[2] \text{ alive}$ are true at w_1, w_2, w_3, w_4 .

Some STIT validities

\Box and $[J]$ are S5 modalities

1. $\models ([J]\varphi \wedge [J]\psi) \rightarrow [J](\varphi \wedge \psi)$
2. $\models [J]\varphi \rightarrow \varphi$
3. $\models [J]\varphi \rightarrow [J][J]\varphi$
4. $\models \langle J \rangle \varphi \rightarrow [J]\langle J \rangle \varphi$
5. $\models (\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$
6. $\models \Box\varphi \rightarrow \varphi$
7. $\models \Box\varphi \rightarrow \Box\Box\varphi$
8. $\models \Diamond\varphi \rightarrow \Box\Diamond\varphi$

Some STIT validities (cont.)

1. $\models \Box\varphi \rightarrow [J]\varphi$
2. $\models \Box\varphi \rightarrow \Box[J]\varphi$
3. $\models \langle AGT \rangle\varphi \leftrightarrow [AGT]\varphi$
4. $\models [J]\varphi \rightarrow [J \cup I]\varphi$
5. $\models [J]\varphi \rightarrow \Diamond[J]\varphi$
6. $\models \Diamond[J]\varphi \rightarrow \Diamond[J \cup I]\varphi$
7. $\models (\Diamond[1]\varphi_1 \wedge \dots \wedge \Diamond[n]\varphi_n) \rightarrow \Diamond[AGT](\varphi_1 \wedge \dots \wedge \varphi_n)$
8. $\models (\Diamond[J]\varphi \wedge \Diamond[I]\psi) \rightarrow \Diamond[J \cup I](\varphi \wedge \psi)$ if $I \cap J = \emptyset$
9. $\models (\Diamond[J]\varphi \wedge \Diamond[I]\neg\varphi) \rightarrow \perp$ if $I \cap J = \emptyset$

$\diamond[J]\varphi$ expresses the game-theoretic concept of J 's α -**ability** for φ (or $\exists\forall$ -ability)

$$\alpha\text{-Ability}_J\varphi \stackrel{\text{def}}{=} \diamond[J]\varphi$$

Definition

A coalition J is said to have α -ability for φ if and only if there exists a possible joint action δ_J of the agents in J such that, for all possible joint actions $\delta'_{AGT \setminus J}$ of the agents in $AGT \setminus J$, if J does δ_J and $AGT \setminus J$ does $\delta'_{AGT \setminus J}$, then φ will be true.

Discussion (cont.)

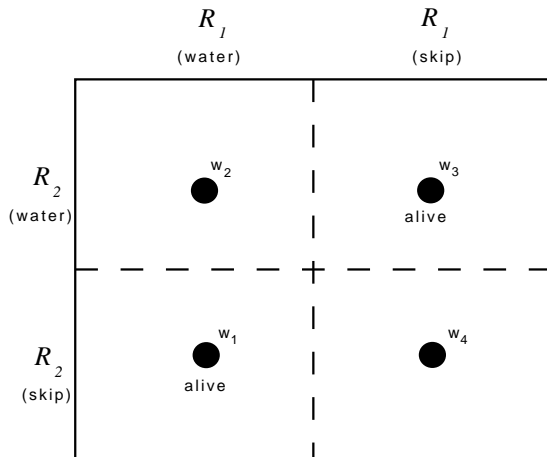
$\neg\Diamond[AGT \setminus J]\neg\varphi$ expresses the game-theoretic concept of J 's **β -ability for φ** (or $\forall\exists$ -ability)

$$\beta\text{-Ability}_J\varphi \stackrel{\text{def}}{=} \neg\Diamond[AGT \setminus J]\neg\varphi$$

Definition

A coalition J is said to have β -ability for φ if and only if, for every possible joint action $\delta'_{AGT \setminus J}$ of the agents in $AGT \setminus J$ there exists a possible joint action δ_J of the agents in J such that if J does δ_J and $AGT \setminus J$ does $\delta'_{AGT \setminus J}$, then φ will be true.

Discussion (cont.)

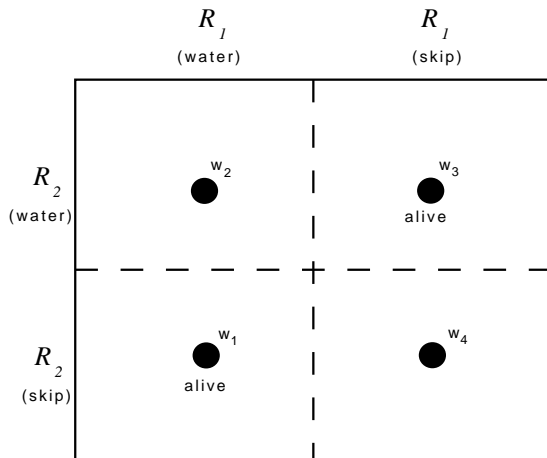


- ▶ β -Ability₁ *alive* and β -Ability₂ *alive* are true at w_1 , w_2 , w_3 and w_4 .

$\chi \wedge \neg[AGT \setminus J]\chi$ expresses a basic notion of **responsibility** of the form “coalition J *could have prevented* a certain state of affairs χ to be true now”.

$$\text{Resp}_J \chi \stackrel{\text{def}}{=} \chi \wedge \neg[AGT \setminus J]\chi$$

Discussion (cont.)



- ▶ $\text{Resp}_1 \neg \text{alive}$ and $\text{Resp}_2 \neg \text{alive}$ are true at w_2 and w_4 .
- ▶ $\text{Resp}_{\{1,2\}} \neg \text{alive}$ and $\text{Resp}_{\{1,2\}} \neg \text{alive}$ are true at w_2 and w_4 .

A general remark

Definition

A logic \mathcal{L} is *finitely axiomatizable* if there is a finite set Ax of axiom schemas such that φ is \mathcal{L} -valid iff there is a deduction of φ from (instances of) Ax using Modus Ponens and Necessitation.

Theorem (Herzig & Schwarzentruher, 2008)

STIT with agents and groups is undecidable and not finitely axiomatizable for $AGT \geq 3$.

A decidable fragment of STIT

The following fragment of STIT called DF^{STIT} is decidable:

$\chi ::= \perp \mid p \mid \chi \wedge \chi \mid \neg\chi$ (propositional formulas)

$\psi ::= [J]\chi \mid \psi \wedge \psi$ (“see-to-it” formulas)

$\varphi ::= \chi \mid \psi \mid \varphi \wedge \varphi \mid \neg\varphi \mid \Diamond\psi$ (“see-to-it” and “can” formulas)

where p ranges over ATM and J over 2^{AGT^*} .

Theorem (Lorini & Schwarzenrüber, 2009)

The satisfiability problem of DF^{STIT} is NP-complete.

Relationship between CL and STIT

⇒ The CL operator $\langle\!\langle J \rangle\!\rangle$ is the fusion of three modalities: historic necessity, agent's choice, and time (*next*)

There is no natural translation of CL into STIT without time

For instance, the following translation does not do the job:

- ▶ $tr(p) = p$
- ▶ $tr(\varphi \wedge \psi) = tr(\varphi) \wedge tr(\psi)$
- ▶ $tr(\neg\varphi) = \neg tr(\varphi)$
- ▶ $tr(\langle\!\langle J \rangle\!\rangle\varphi) = \diamond[J]tr(\varphi)$
- ▶ $tr(\langle\!\langle \emptyset \rangle\!\rangle\varphi) = \square tr(\varphi)$

Relationship between CL and STIT (cont.)

$\langle\langle\emptyset\rangle\rangle p \wedge \langle\langle\emptyset\rangle\rangle \langle\langle J\rangle\rangle \neg p$ is CL satisfiable but its translation into STIT is STIT invalid

$$\Rightarrow tr(\langle\langle\emptyset\rangle\rangle p \wedge \langle\langle\emptyset\rangle\rangle \langle\langle J\rangle\rangle \neg p) = \Box p \wedge \Box \Diamond [J] \neg p$$

- ▶ $\Box p \wedge \Box \Diamond [J] \neg p$ implies $\Box p \wedge \Diamond [J] \neg p$
(By T for \Box)
- ▶ $\Box p \wedge \Diamond [J] \neg p$ implies $\Box [J] p \wedge \Diamond [J] \neg p$
(By 4 for \Box and the STIT theorem $\Box \varphi \rightarrow [J] \varphi$)
- ▶ $\Box [J] p \wedge \Diamond [J] \neg p$ implies $\Diamond ([J] p \wedge [J] \neg p)$
(By the STIT theorem $(\Box \varphi \wedge \Diamond \psi) \rightarrow \Diamond (\varphi \wedge \psi)$)
- ▶ $\Diamond ([J] p \wedge [J] \neg p)$ implies $\Diamond [J] \perp$
- ▶ $\Diamond [J] \perp$ implies $\Diamond \perp$
(By T for $[J]$)
- ▶ $\Diamond \perp$ implies \perp

STIT with discrete time

If we extend STIT with discrete time (modality *next*), CL can be translated into STIT

The language of STIT with discrete time is defined by the following BNF:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid [J]\varphi \mid \Box\varphi \mid \mathbf{X}\varphi$$

where p ranges over ATM and J over 2^{AGT^*} .

$\Rightarrow \mathbf{X}\varphi \approx$ “ φ will be true in the next state”.

STIT models with discrete time

A STIT-model with discrete time (and without endpoints) is a tuple $M = (W, \{R_J\}_{J \subseteq AGT}, H, F_X, V)$ where:

- ▶ W, R_J, H and V are defined as in STIT models;
- ▶ F_X is a total function $F_X : W \longrightarrow W$.

$\Rightarrow F_X(w)$ is the successor of world w

$M, w \models \mathbf{X}\varphi$ iff $M, F_X(w) \models \varphi$.

- ▶ $tr(p) = p$
- ▶ $tr(\varphi \wedge \psi) = tr(\varphi) \wedge tr(\psi)$
- ▶ $tr(\neg\varphi) = \neg tr(\varphi)$
- ▶ $tr(\langle\langle J \rangle\rangle\varphi) = \diamond[J]X tr(\varphi)$
- ▶ $tr(\langle\langle \emptyset \rangle\rangle\varphi) = \square X tr(\varphi)$

Theorem (Broersen et al., 2006)

φ is CL satisfiable iff $tr(\varphi)$ is STIT satisfiable.

Towards intentional STIT: from uniform strategies to joint intentions

A STIT extension with knowledge

We add modalities for knowledge and common knowledge to the STIT language with discrete time:

⇒ $K_i\varphi$: agent i knows that φ

⇒ $CK_J\varphi$: there is common knowledge in J that φ

Epistemic STIT models with discrete time

An epistemic STIT-model with discrete time is a tuple

$M = (W, \{R_J\}_{J \subseteq AGT}, H, F_X, \{E_i\}_{i \in AGT}, V)$ where:

- ▶ W, R_J, H, F_X and V are defined as in STIT models with discrete time;
- ▶ For every $i \in AGT$, E_i is an equivalence (epistemic) relation over W .

$\Rightarrow E_i(w) = \{v | (w, v) \in E_i\}$ are the epistemic alternatives for i at w

Truth conditions for knowledge and common knowledge

$M, w \models K_i \varphi$ iff $M, v \models \varphi$ for all $(w, v) \in E_i$

$M, w \models CK_J \varphi$ iff $M, v \models \varphi$ for all $(w, v) \in E_J^+$

where

$\Rightarrow E_J = \bigcup_{i \in J} E_i$

$\Rightarrow E_J^+$ is the reflexive transitive closure of E_J

Interactions between knowledge and historic necessity: discussion

\Rightarrow *Semantic constraint S1*: For every $w \in W$, $E_i(w) \subseteq H(w)$

\Rightarrow *Corresponding axiom PerfectInfo*: $\Box\varphi \rightarrow K_i\varphi$

It encodes perfect information about the situation of interaction

If we impose S1 the following become valid:

1. $\Box\varphi \rightarrow CK_J\Box\varphi$
2. $\Diamond\varphi \rightarrow CK_J\Diamond\varphi$
3. $\Diamond[I]\varphi \rightarrow CK_J\Diamond[I]\varphi$

Interactions between knowledge and historic necessity: discussion (cont.)

⇒ *Semantic constraint S2*: For every $w \in W$, $E_i(w) \subseteq R_i(w)$

⇒ *Corresponding axiom **PerfectInfo+ActAware***: $[i]\varphi \rightarrow K_i\varphi$

It encodes perfect information about the structure of interaction + an agent's knowledge about its current choice (the only uncertainty is about choices of others)

Theorem

If S2 then S1.

Interactions between knowledge and historic necessity: discussion (cont.)

More realistic principles:

⇒ *Semantic constraint (confluence) S3*: $\forall w \in W, \forall v \in H(w), \forall u \in E_i(w), H(u) \cap E_i(v) \neq \emptyset$

⇒ *Corresponding axiom*: $\Diamond K_i \varphi \rightarrow K_i \Diamond \varphi$

⇒ *Semantic constraint (permutation) S4*: $K_i \circ H = H \circ K_i$

⇒ *Corresponding axiom*: $K_i \Box \varphi \leftrightarrow \Box K_i \varphi$

Theorem

S3 if and only if S4.

The concept of uniform strategy

For an agent i to have the **power of** ensuring φ , i must have both: the objective capability to achieve φ and, the discretion (awareness) over his capability (Castelfranchi 2003, Barnes 1988)

The concept of uniform strategy (cont.)

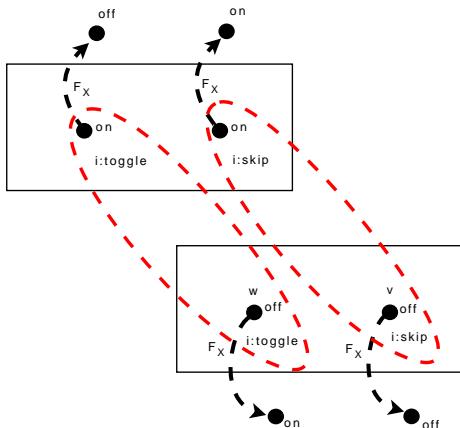
- ▶ *DE DICTO* sentence: “ i knows that there is some action such that if he chooses it, he will ensure φ in the next state”
- ▶ *DE RE* sentence: “there is some action such that if agent i chooses it, he knows that he will ensure φ in the next state”

⇒ DE DICTO: $K_i \diamond [i] X \varphi$

⇒ DE RE: $\diamond K_i [i] X \varphi$

Only formula $\diamond K_i [i] X \varphi$ captures a proper concept of agent i 's *power of ensuring φ* or agent i 's *uniform strategy over φ*

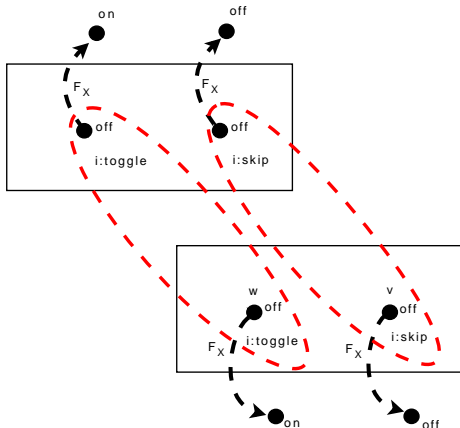
Example: an agent trying to switch on a light



Agent i can either *toggle* or *skip* (do nothing) but he is uncertain about the current state of the light (*on* or *off*)

- ▶ $K_i \diamond [i] X_{on}$ is true at w and v .
- ▶ $\diamond K_i [i] X_{on}$ is false at w and v .

Example: an agent trying to switch on a light (cont.)



Agent i can either *toggle* or *skip* (do nothing) and he knows the current state of the light

- ▶ $K_i \diamond [i] X_{on}$ is true at w and v .
- ▶ $\diamond K_i [i] X_{on}$ is true at w and v .

A STIT extension with knowledge and goals

We add modalities for goals to the STIT language with discrete time and knowledge:

⇒ **Choice_{*i*}** φ : agent i wants φ to be true (or i has chosen to pursue φ)

An epistemic STIT-model with discrete time and choices is a tuple $M = (W, \{R_J\}_{J \subseteq AGT}, H, F_X, \{E_i\}_{i \in AGT}, \{C_i\}_{i \in AGT}, V)$ where:

- ▶ W, R_J, H, F_X, E_i and V are defined as in STIT models with discrete time and knowledge;
- ▶ For every $i \in AGT$, C_i is a serial relation over W .

$\Rightarrow C_i(w) = \{v \mid (w, v) \in C_i\}$ are the worlds that agent i wants to achieve at w

$M, w \models \mathbf{Choice}_i \varphi$ iff $M, v \models \varphi$ for all $(w, v) \in C_i$

Interactions between choices and knowledge: discussion

⇒ *Semantic constraint S1*: if $(w, v) \in E_i$ and $(v, u) \in C_i$ then $(w, u) \in C_i$

⇒ *Corresponding axiom PosIntro*: $\text{Choice}_i\varphi \rightarrow K_i\text{Choice}_i\varphi$

⇒ *Semantic constraint S2*: if $(w, v) \in E_i$ and $(w, u) \in C_i$ then $(v, u) \in C_i$

⇒ *Corresponding axiom NegIntro*: $\neg\text{Choice}_i\varphi \rightarrow K_i\neg\text{Choice}_i\varphi$

Choices are positively and negatively introspective

Interactions between choices and knowledge: discussion (cont.)

⇒ *Semantic constraint S3*: $C_i(w) \subseteq E_i(w)$

⇒ *Corresponding axiom **Compat***: $K_i\varphi \rightarrow \text{Choice}_i\varphi$

Choices are realistic (i.e. an agent can only choose states that he considers possible)

What “we intend to do something together” means?

- ▶ We intend to paint a house
- ▶ We intend to organize a party
- ▶ We intend to write a paper together

Joint intentions (cont.)

- ▶ **Individualistic/reductionist approaches** (Bratman 1996, 1999): joint intentions can be analyzed as an interlocking web of goals, preferences and intentions of individual agents.
- ▶ **Non-reductionist approaches**: a joint intention is characterized by a sense of collectivity or sharedness that is lost when one reduces it to a summation of individual preferences, goals and intentions.
 - ▶ Joint intentions are intentions of individuals that reason and act as group/team members (Bacharach, 2006; Gold & Sugden, 2006).
 - ▶ Distinction between *WE-mode* and *I-mode* (Tuomela, 1995).

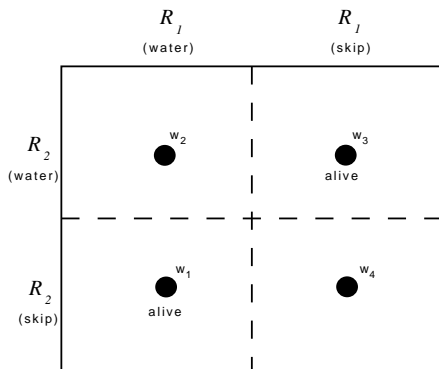
We here consider the first kind of approaches

Joint intentions (cont.)

Three levels in Bratman's analysis of joint intention

1. **Joint goal to ensure** φ : I intend and you intend that we ensure φ .
2. **Joint plan to ensure** φ : I intend and you intend that we perform the joint action δ in order to ensure φ .
3. **Common ground**: conditions 1 and 2 are common knowledge between me and you.

Example: two agents taking care of a plant



Joint intentions (cont.)

1. I intend and you intend that we ensure that the plant is alive.
2. I intend and you intend that we perform the joint action “I water the plant, you do nothing” in order to ensure that the plant is alive.
3. Conditions 1 and 2 are common knowledge between me and you.

Joint intentions (cont.)

$$\mathbf{JInt}_{\{1,2\}} \mathit{alive} \stackrel{\text{def}}{=} \mathbf{CK}_{\{1,2\}}(\mathbf{Choice}_1[\{1,2\}]\mathbf{X}\mathit{alive} \wedge \mathbf{Choice}_2[\{1,2\}]\mathbf{X}\mathit{alive})$$

$\mathbf{JInt}_{\{1,2\}} \mathit{alive}$ means “1 and 2 have the joint goal to ensure that the plant is alive”

$$\begin{aligned} & \mathbf{JInt}_{\{1,2\}}(\langle \mathit{skip}_1, \mathit{water}_2 \rangle, \mathit{alive}) \stackrel{\text{def}}{=} \\ & \mathbf{CK}_{\{1,2\}}(\mathbf{Choice}_1[\{1,2\}](\mathit{skip}_1 \wedge \mathit{water}_2 \wedge \mathbf{X}\mathit{alive}) \wedge \\ & \quad \mathbf{Choice}_2[\{1,2\}](\mathit{skip}_1 \wedge \mathit{water}_2 \wedge \mathbf{X}\mathit{alive})) \end{aligned}$$

$\mathbf{JInt}_{\{1,2\}}(\langle \mathit{skip}_1, \mathit{water}_2 \rangle, \mathit{alive})$ means “1 and 2 jointly intend to perform the (joint) plan $\langle \mathit{skip}_1, \mathit{water}_2 \rangle$ in order to ensure that the plant is alive”

References

- ▶ Alur, R., Henzinger, T., Kupferman, O. (2002). *Alternating-time temporal logic*. Journal of the ACM, 49:672-713.
- ▶ Bacharach, M. (2006). *Beyond Individual Choice. Teams and Frames in Game Theory*. Princeton University Press.
- ▶ Barnes, B. (1988). *The Nature of Power*. Polity Press.
- ▶ Belnap, N. Perloff, M., Xu, M. (2001). *Facing the future: agents and choices in our indeterminist world*. Oxford University Press.
- ▶ Bratman, M. (1999). *Faces of Intention*. Cambridge University Press.
- ▶ Bratman, M. (1996). *Shared Cooperative Activity*. The Philosophical Review, 101(2):327-341.
- ▶ Balbiani, P. Herzig, A., Troquard, N. (2008). *Alternative axiomatics and complexity of deliberative STIT theories*. Journal of Philosophical Logic, 37(4):387-406.
- ▶ Broersen, J. Herzig, A., Troquard, N. (2007). *Normal coalition logic and its conformant extension*. In Proceedings of the eleventh conference on Theoretical Aspects of Rationality and Knowledge (TARK 2007), pp. 91-101.

References

- ▶ Broersen, J. Herzig, A., Troquard, N. (2006). *Embedding Alternating-time Temporal Logic in strategic STIT logic of agency*. Journal of Logic and Computation, 16(5):559-578.
- ▶ Castelfranchi, C. (1996). *The micro-macro constitution of power*. Protosociology, 18-19.
- ▶ Herzig, A., Schwarzenrüber, F. (2008). Properties of logics of individual and group agency. In *Proceedings of Advances in Modal Logic 2008 (AiML 2008)*.
- ▶ van der Hoek, W., Wooldridge, M. (2005). *On the logic of cooperation and propositional control*. Artificial Intelligence, 64(1-2):81-119.
- ▶ Horty, J. F., Belnap, N. (1995). *The deliberative STIT: A study of action, omission, and obligation*. Journal of Philosophical Logic, 24(6):583-644.
- ▶ Horty, J. F. (2001). *Agency and Deontic Logic*, Oxford University Press.
- ▶ Kanger, S. (1972). *Law and Logic*. Theoria, 38:105-132.

References

- ▶ Lorini, E., Schwarzenrüber, F. (2009). *A Logic for Reasoning about Counterfactual Emotions*. In Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI'09).
- ▶ Pauly, M. (2002). *A modal logic for coalitional power in games*. Journal of Logic and Computation, 12(1):149-166.
- ▶ Pauly, M. (2001). *Logic for Social Software*. PhD thesis, University of Amsterdam, The Netherlands.
- ▶ Pörn, I. (1977). *Action Theory and Social Science: Some Formal Models*. Synthese Library, Reidel.
- ▶ Tuomela, R. (1995). *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press.