A normal simulation of coalition logic and an epistemic extension

Jan Broersen, Andreas Herzig, Nicolas Troquard

▶ **To cite this version:**

Jan Broersen, Andreas Herzig, Nicolas Troquard. A normal simulation of coalition logic and an epistemic extension. Theoretical Aspects of Rationality and Knowledge (TARK 2007), Jun 2007, Brussels, Belgium. pp.91-101, 10.1145/1324249.1324264 . hal-03516647

HAL Id: hal-03516647
https://hal.science/hal-03516647

Submitted on 10 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Normal Simulation of Coalition Logic and an Epistemic Extension

Jan Broersen
Department of Information and Computing Sciences
Utrecht University

Andreas Herzig
Institut de Recherche en Informatique de Toulouse

Nicolas Troquard
Institut de Recherche en Informatique de Toulouse &
Laboratory of Applied Ontology
Trento

Abstract

In this paper we show how coalition logic can be reduced to the fusion of a normal modal STIT logic for agency and a standard normal temporal logic for discrete time, and how this multi-modal system can be suitably extended with an epistemic modality. Both systems are complete, and we provide a new axiomatization for the STIT-fragment. The epistemic extension enables us to express that agents see to something under uncertainty about the present state or uncertainty about which action is being taken. In accordance with established terminology in the planning community, we call this version of STIT the 'conformant STIT'. The conformant STIT enables us to express that agents are able to perform a uniform strategy. As a final word of recommendation for this paper we want to point out that its subject is at the junction of four academic fields, viz. modal logic, philosophy, game-theory and AI-planning.

1 Introduction

Coalition Logic, CL for short, was proposed by Pauly in [Pau01] as a logic for reasoning about social procedures characterized by complex strategic interactions between agents, individuals or groups. Examples of such procedures are fair-division algorithms or voting processes. CL facilitates reasoning about abilities of coalitions in games by extending classical logic with operators $\langle\!\langle J\rangle\!\rangle\varphi$ for groups of agents J, reading: "the coalition J has a joint strategy to ensure that φ".[1]

In this paper we show how CL is naturally embedded in a variant of STIT theory [BPX01]. STIT theory is the most prominent account of agency in the philosophy of action. It is the logic of constructions of the form "agent i sees to it that φ holds". Our variant of STIT theory is the fusion of a multi-agent version of the so-called Chellas STIT operator[2] with a temporal next-time operator. We thus assume no interaction between the STIT operator and the next-time operator. We acknowledge that any reasonable logic for time and agency cannot present the action and time dimensions as logically independent, but our point in this paper is that we do not need the interaction to give a correct embedding of Coalition Logic. For easy reference, we adopt the name NCL ('Normal Simulation of Coalition Logic') for the fusion of our multi-agent STIT operator and the next-time operator. The embedding of CL in NCL is an interesting result since it shows how to extend CL with capabilities of reasoning about what a coalition is actually doing (as opposed to what it *could* do).

[1] Note that we use $\langle\!\langle J\rangle\!\rangle\varphi$ as an alternative notation for Pauly's non-normal operator $[J]\varphi$. We introduce this alternative syntax for two reasons: (1) the new syntax evokes the quantifier combination $\exists - \forall$ underlying the semantics, and (2) we use Pauly's original syntax $[J]\varphi$ to denote the STIT operator, thereby emphasizing that this is a normal modal necessity operator.

[2] The logic of Chellas STIT together with the logic of the deliberative STIT form what is called the *deliberative STIT theories* and have been axiomatized in [BPX01, Chap. 17].

In social choice theory, in particular since Harsanyi, the interaction between ability models and epistemic models has been a main focus of research. It has been realized that intentionality of action presupposes awareness or knowledge of the means by which effects are ensured. Philosophers refer to this ability of agents as having the *power* to ensure a condition. So, in order to say that an agent 'can' or 'has the power to' ensure a condition, there should not only be an action in the agent's repertoire that ensures the condition, the agent should also know how to choose the action.

More recently the issue of 'knowing how to act' has come up in the logic ATEL [vdHW02] which is the epistemic extension of the logic of strategic ability ATL [AHK02]. The problem is often referred to as the problem of *uniform strategies*. In particular, ATEL does not allow to distinguish the situations where:

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.

In this paper we do not reason about series of choices, alias strategies, which is why our starting point is CL instead of ATL. We extend NCL with an S5 modal operator for knowledge and show that the resulting complete logic, that we refer to as ENCL, solves the problem of uniform strategies. Furthermore, the epistemic extension enables us to define a notion of 'seeing to it under uncertainty'. In accordance with established terminology in the planning literature, we call this version of STIT, the 'conformant STIT'. Finally, we give an argument for our view that ATEL-based approaches to the problem of uniform strategies are not likely to succeed.

2 Coalition Logic

Let AGT be a finite set of agents and $Prop$ a countable set of atomic formulas. The syntax of Coalition Logic is defined as follows:

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

where p ranges over $Prop$ and J ranges over the subsets of AGT. The other boolean connectives are defined as usual.

2.1 Coalition model semantics

Definition 1 (effectivity function)**.** *Given a finite set of agents AGT and a set of states S, an* effectivity function *is a function $E : 2^{AGT} \longrightarrow 2^{2^S}$. An effectivity function is said to be:*

- J-maximal *iff for all $X \subseteq S$, if $S \setminus X \notin E(\overline{J})$ then $X \in E(J)$.*
- outcome monotonic *iff for all $X \subseteq X' \subseteq S$ and for all $J \subseteq AGT$, if $X \in E(J)$ then $X' \in E(J)$.*
- superadditive *iff for all X_1, X_2, J_1, J_2 such that $J_1 \cap J_2 = \emptyset$, $X_1 \in E(J_1)$ and $X_2 \in E(J_2)$ imply that $X_1 \cap X_2 \in E(J_1 \cup J_2)$.*

E intuitively associates every coalition J to a set of $X \subseteq S$ (a set of possible outcomes) for which J is effective. That is, J can force the world to be in some state of X at the next step.

Definition 2 (playable effectivity function)**.** *An effectivity function $E : 2^{AGT} \longrightarrow 2^{2^S}$ is said to be* playable *iff*

1. $\forall J \subseteq AGT, \emptyset \notin E(J)$; *(Liveness)*
2. $\forall J \subseteq AGT, S \in E(J)$; *(Termination)*
3. *E is AGT-maximal;*
4. *E is outcome-monotonic; and*
5. *E is superadditive.*

Definition 3. *A* coalition model *is a pair $((S, E), V)$ where:*

- *S is a nonempty set of states;*
- *$E : S \longrightarrow (2^{AGT} \longrightarrow 2^{2^S})$ is a* playable effectivity structure*;*
- *$V : S \longrightarrow 2^{Prop}$ is a valuation function.*

The mapping E associates every state s to a playable effectivity function $E(s)$. We will write $E_s(J)$ instead of $E(s)(J)$.

Truth conditions are standard for classical formulas. We evaluate the coalitional operators against a coalition model M and a state s according to:

$$M, s \models \langle\!\langle J \rangle\!\rangle \varphi \text{ iff } \{s \mid M, s \models \varphi\} \in E_s(J).$$

Note that Coalition Logic is a non-normal weak modal logic, and that the semantics given above is a so-called 'neighborhood semantics'. For instance, the outcome monotonicity property can be reformulated as reachability of neighborhoods being closed under supersets.

2.2 Game semantics

In [Pau02], Marc Pauly investigates an alternative semantics for CL in terms of game structures. Below we use the equivalence of both semantics to prove that we can safely assume, that is, without changing the (weak) modal logic, that any joint choice (also called *choice profile*) of all agents AGT in the system yields a single outcome state.

Definition 4. *A* strategic game *is a tuple* $G = (S, \{\Sigma_i | i \in AGT\}, o)$ *where* S *is a nonempty set,* Σ_i *is a nonempty set of choices for every agent* $i \in AGT$*,* $o : \prod_{i \in AGT} \Sigma_i \longrightarrow S$ *is an outcome function which associates an outcome state in* S *with every combination of choice of agents (choice profile).*

It appears that there is a strong link between a coalition model (whose effectivity structure is *playable* by definition) and a strategic game.

Definition 5. *Given a strategic game* G*, the* effectivity function $E_G : 2^{AGT} \longrightarrow 2^{2^S}$ of G *is defined as* $X \in E_G(C)$ *iff there is* $\sigma_C \in \prod_{i \in C} \Sigma_i$ *such that for every* $\sigma_{\overline{C}} \in \prod_{i \in \overline{C}} \Sigma_i$ *we have* $o(\sigma_C \times \sigma_{\overline{C}}) \in X$*.*

Pauly then gives the following characterization:

Theorem 1 ([Pau02])**.** *An effectivity function* E *is playable iff it is the effectivity function of some strategic game.*

Definition 6. *Let* E *be an effectivity function. A set* $Y \subseteq S$ *is called a* minimal effectivity outcome at s for J *iff (1)* $Y \in E_s(J)$ *and (2) there is no* $Y' \in E_s(J)$ *s.t.* $Y' \subset Y$*.*

Definition 7. *The* non-monotonic core *of* E *is the mapping* $\mu_E : 2^{AGT} \times S \longrightarrow 2^{2^S}$ *such that* $\mu_E(J, s) = \{Y \mid Y$ *is a minimal effectivity outcome at* s *for* $J\}$*.*

The outcome of a strategic game is completely determined when every agent has made its choice.

Proposition 1. $\mu_E(AGT, s)$ *is a set of singletons.*

Proof. With Definition 5 this is a corollary of Theorem 1. □

2.3 Axiomatization

The set of formulas that are valid in coalition models is completely axiomatized by the following principles [Pau02].

(ProTau)	any sufficient set of propositional logic schemas
$(\bot)$	$\neg\langle\!\langle J \rangle\!\rangle \bot$
$(\top)$	$\langle\!\langle J \rangle\!\rangle \top$
(N)	$\neg\langle\!\langle \emptyset \rangle\!\rangle \neg\varphi \rightarrow \langle\!\langle AGT \rangle\!\rangle \varphi$
(M)	$\langle\!\langle J \rangle\!\rangle (\varphi \wedge \psi) \rightarrow \langle\!\langle J \rangle\!\rangle \psi$
(S)	$\langle\!\langle J_1 \rangle\!\rangle \varphi \wedge \langle\!\langle J_2 \rangle\!\rangle \psi \rightarrow \langle\!\langle J_1 \cup J_2 \rangle\!\rangle (\varphi \wedge \psi)$ if $J_1 \cap J_2 = \emptyset$
(MP)	from φ and $\varphi \rightarrow \psi$ infer ψ
(RE)	from $\varphi \leftrightarrow \psi$ infer $\langle\!\langle J \rangle\!\rangle \varphi \leftrightarrow \langle\!\langle J \rangle\!\rangle \psi$

Theorem 2 ([Pau02])**.** *The principles (ProTau),* $(\bot)$*,* $(\top)$*,* (N)*,* (M)*,* (S)*,* (MP) *and* (RE) *are complete with respect to the class of all coalition models.*

Note that the (N) axiom corresponds to AGT-maximality of the effectivity structures. It says that if a formula is not settled true, the coalition of all agents (AGT) can always coordinate their choices to make its negation true. The axiom (S) corresponds to superadditivity and says that two disjoint coalitions can combine their efforts to ensure a conjunction of properties. Note that from (S) and $(\bot)$ it follows that $\langle\!\langle J_1 \rangle\!\rangle \varphi \wedge \langle\!\langle J_2 \rangle\!\rangle \neg\varphi$ is not satisfiable for disjoint J_1 and J_2. So, two disjoint coalitions cannot ensure opposed facts.

Theoremhood and consistency are defined as usual.

3 A Normal Simulation of Coalition Logic

Let $AGT = \{0, \ldots, n-1\}$ be a finite set of $n \geq 1$ agents and $Prop$ a countable set of atomic formulas. The logic NCL has the following syntax, where p ranges over elements of $Prop$ and J ranges over the set of subsets of AGT:

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

Other boolean connectives are defined by abbreviations, as usual. Furthermore, we define $\langle J\rangle\varphi =_{def} \neg[J]\neg\varphi$. The intended meaning for the operator $\mathbf{X}$ is 'next time', and for the operator $[J]$ 'agents J see to it that'.

Axiomatization We give the following axiom schemas for NCL.

(ProTau)	any sufficient set of propositional logic schemas
S5($[J]$)	any sufficient set of S5-schemas, for every $[J]$
(Mon)	$[J_1]\varphi \to [J_1 \cup J_2]\varphi$
Elim($[\emptyset]$)	$\langle\emptyset\rangle\varphi \to \langle J\rangle\langle\overline{J}\rangle\varphi$
Triv($[AGT]$)	$\varphi \to [AGT]\varphi$
K($\mathbf{X}$)	$\mathbf{X}(\varphi \to \psi) \to (\mathbf{X}\varphi \to \mathbf{X}\psi)$
D($\mathbf{X}$)	$\mathbf{X}\varphi \to \neg\mathbf{X}\neg\varphi$
Det($\mathbf{X}$)	$\neg\mathbf{X}\neg\varphi \to \mathbf{X}\varphi$

We also assume the standard inference rules of modus ponens, and necessitation for $\mathbf{X}$ and $[\emptyset]$. From the latter necessitation for every $[J]$ follows by the inclusion axiom (Mon). $\overline{J}$ denotes the complement of J w.r.t. AGT. The axiom Elim($[\emptyset]$) will be central in this section. Note that the converses of Elim($\emptyset$) and Triv(AGT) can be proved from (Mon), S5($\emptyset$) and S5(AGT). Hence, we have $\vdash \langle\emptyset\rangle\varphi \leftrightarrow \langle J\rangle\langle\overline{J}\rangle\varphi$ and $\vdash \varphi \leftrightarrow [AGT]\varphi$

Semantics An NCL-model is a tuple $\mathcal{M} = (W, R, F_X, \pi)$ where:

- W is a set of worlds (alias contexts);
- R is a collection of equivalence relations R_J (one for every coalition $J \subseteq AGT$) such that:
  - $R_{J_1 \cup J_2} \subseteq R_{J_1}$
  - $R_\emptyset \subseteq R_J \circ R_{\overline{J}}$
  - $R_{AGT} = Id$;
- $F_X : W \longrightarrow W$ is a total function;
- $\pi : W \longrightarrow 2^{Prop}$ is a valuation function.

The truth conditions are:

- $\mathcal{M}, w \models p$ iff $p \in \pi(w)$
- $\mathcal{M}, w \models \mathbf{X}\varphi$ iff $\mathcal{M}, F_X(w) \models \varphi$
- $\mathcal{M}, w \models [J]\varphi$ iff for all $u \in R_J(w)$, $\mathcal{M}, u \models \varphi$

together with the usual definitions for the other operators. Validity and satisfiability are also defined as usual.

Theorem 3. *NCL is determined by the class of NCL-models.*

Proof. Soundness is obtained by a routine argument and completeness is immediate from Sahlqvist's theorem, cf. [BdRV01]. □

In the rest of the section we prove two lemmas that are useful in the sequel.

Lemma 1. $\vdash \langle\emptyset\rangle\varphi \to \langle J_1\rangle\langle J_2\rangle\varphi$ *if* $J_1 \cap J_2 = \emptyset$.

Proof. By Elim($\emptyset$) we have $\vdash \langle\emptyset\rangle\varphi \to \langle J_1\rangle\langle\overline{J_1}\rangle\varphi$. Then, by hypothesis, $J_1 \cap J_2 = \emptyset$, or equivalently $J_2 \subseteq \overline{J_1}$. Thus, by (Mon), $\vdash \langle\overline{J_1}\rangle\varphi \to \langle J_2\rangle\varphi$. We obtain $\vdash \langle J_1\rangle\langle\overline{J_1}\rangle\varphi \to \langle J_1\rangle\langle J_2\rangle\varphi$ by standard modal principles for $[J_1]$. We conclude that $\vdash \langle\emptyset\rangle\varphi \to \langle J_1\rangle\langle J_2\rangle\varphi$. □

In [BPX01, Chap. 17] the authors provide an axiomatization of the theories of deliberative STIT in terms of a family of axiom schemas (AIA$_k$). These capture a central idea of multi-agent STIT theories saying that agents' choices are independent. We now give a theorem of NCL that generalizes (AIA$_1$) from individuals to coalitions, and that will be instrumental later in the proof of superadditivity in Theorem 4.

Lemma 2. $\vdash \langle\emptyset\rangle[J_0]\varphi_0 \wedge \langle\emptyset\rangle[J_1]\varphi_1 \to \langle\emptyset\rangle([J_0]\varphi_0 \wedge [J_1]\varphi_1)$ *for* $J_0 \cap J_1 = \emptyset$.

Proof. Suppose $J_0 \cap J_1 = \emptyset$. We establish the following deduction:

1. $\langle\emptyset\rangle[J_0]\varphi_0 \rightarrow \langle J_1\rangle\langle J_0\rangle[J_0]\varphi_0$ by Lemma 1
2. $\langle\emptyset\rangle[J_0]\varphi_0 \rightarrow \langle J_1\rangle[J_0]\varphi_0$ from 1 by S5($[J_0]$)
3. $\langle\emptyset\rangle[J_0]\varphi_0 \wedge [J_1]\varphi_1 \rightarrow \langle J_1\rangle[J_0]\varphi_0 \wedge [J_1][J_1]\varphi_1$ from 2 by S5($[J_1]$)
4. $\langle\emptyset\rangle[J_0]\varphi_0 \wedge [J_1]\varphi_1 \rightarrow \langle J_1\rangle([J_0]\varphi_0 \wedge [J_1]\varphi_1)$ from 3 by S5($[J_1]$)
5. $\langle\emptyset\rangle(\langle\emptyset\rangle[J_0]\varphi_0 \wedge [J_1]\varphi_1) \rightarrow \langle\emptyset\rangle\langle J_1\rangle([J_0]\varphi_0 \wedge [J_1]\varphi_1)$ from 4 by standard modal principles
6. $\langle\emptyset\rangle[J_0]\varphi_0 \wedge \langle\emptyset\rangle[J_1]\varphi_1 \rightarrow \langle\emptyset\rangle\langle J_1\rangle([J_0]\varphi_0 \wedge [J_1]\varphi_1)$ from 5 by S5($[\emptyset]$)
7. $\langle\emptyset\rangle[J_0]\varphi_0 \wedge \langle\emptyset\rangle[J_1]\varphi_1 \rightarrow \langle\emptyset\rangle([J_0]\varphi_0 \wedge [J_1]\varphi_1)$ from 6 by (Mon) and S5($[\emptyset]$)

□

In [[Sch07, BGH+07]] it is shown that it follows that the problem of deciding the satisfiability of a formula of NCL is NEXPTIME-complete.

4 Translating Coalition Logic

We give the following translation from Coalition Logic to NCL.

$$\begin{array}{lcl} tr(p) & = & p \\ tr(\langle\!\langle J\rangle\!\rangle\varphi) & = & \langle\emptyset\rangle[J]\mathbf{X}tr(\varphi) \end{array}$$

and homomorphic for the other connectives.

Theorem 4. *If φ is a theorem of CL then $tr(\varphi)$ is a theorem of NCL.*

Proof. First, the translations of the CL axiom schemas are theorems of NCL. The only non trivial cases are superadditivity and AGT-maximality. We start with the latter:

$tr(\neg\langle\!\langle\emptyset\rangle\!\rangle\neg\varphi \rightarrow \langle\!\langle AGT\rangle\!\rangle\varphi) = \neg\langle\emptyset\rangle[\emptyset]\mathbf{X}\neg tr(\varphi) \rightarrow \langle\emptyset\rangle[AGT]\mathbf{X}tr(\varphi)$. Since $[AGT]\psi \leftrightarrow \psi$ by Triv(AGT), and $\langle\emptyset\rangle[\emptyset]\psi \leftrightarrow [\emptyset]\psi$ by S5($[\emptyset]$), the translation of (N) is equivalent to $\neg[\emptyset]\mathbf{X}\neg tr(\varphi) \rightarrow \langle\emptyset\rangle\mathbf{X}tr(\varphi)$. This is equivalent to $\langle\emptyset\rangle\neg\mathbf{X}\neg tr(\varphi) \rightarrow \langle\emptyset\rangle\mathbf{X}tr(\varphi)$ which is proved a theorem from applying $[\emptyset]$-necessitation to Det($\mathbf{X}$) and applying a variant of the K-axiom.

For superadditivity we have:

$tr(\langle\!\langle J_1\rangle\!\rangle\varphi \wedge \langle\!\langle J_2\rangle\!\rangle\psi \rightarrow \langle\!\langle J_1 \cup J_2\rangle\!\rangle(\varphi \wedge \psi)) = \langle\emptyset\rangle[J_1]\mathbf{X}tr(\varphi) \wedge \langle\emptyset\rangle[J_2]\mathbf{X}tr(\psi) \rightarrow \langle\emptyset\rangle[J_1 \cup J_2]\mathbf{X}(tr(\varphi) \wedge tr(\psi))$

1. $\langle\emptyset\rangle[J_1]\mathbf{X}tr(\varphi) \wedge \langle\emptyset\rangle[J_2]\mathbf{X}tr(\psi) \rightarrow \langle\emptyset\rangle([J_1]\mathbf{X}tr(\varphi) \wedge [J_2]\mathbf{X}tr(\psi))$. by Lemma 2
2. $[J_1]\mathbf{X}tr(\varphi) \wedge [J_2]\mathbf{X}tr(\psi) \rightarrow [J_1 \cup J_2]\mathbf{X}tr(\varphi) \wedge [J_1 \cup J_2]\mathbf{X}tr(\psi)$. by (Mon)
3. $\langle\emptyset\rangle([J_1]\mathbf{X}tr(\varphi) \wedge [J_2]\mathbf{X}tr(\psi)) \rightarrow \langle\emptyset\rangle([J_1 \cup J_2](\mathbf{X}tr(\varphi) \wedge \mathbf{X}tr(\psi))$ from previous line by standard modal principles
4. $\langle\emptyset\rangle[J_1]\mathbf{X}tr(\varphi) \wedge \langle\emptyset\rangle[J_2]\mathbf{X}tr(\psi) \rightarrow \langle\emptyset\rangle[J_1 \cup J_2]\mathbf{X}(tr(\varphi) \wedge tr(\psi))$ from lines 1 and 3 by standard modal principles for $\mathbf{X}$.

Second, clearly the translation of modus ponens preserves validity. To prove that the translation of CL's (RE) preserves validity suppose $tr(\varphi \leftrightarrow \psi) = tr(\varphi) \leftrightarrow tr(\psi)$ is a theorem of NCL. We have to prove that $tr(\langle\!\langle J\rangle\!\rangle\varphi \leftrightarrow \langle\!\langle J\rangle\!\rangle\psi) = \langle\emptyset\rangle[J]tr(\varphi) \leftrightarrow \langle\emptyset\rangle[J]tr(\psi)$ is a theorem of NCL. This follows from the theoremhood of $tr(\varphi) \rightarrow tr(\psi)$ by standard modal principles. □

Lemma 3. *Let $M = ((S, E), V)$ a coalition model and $selec : S \longrightarrow S$ some mapping such that if $selec(s) = s'$ then $\{s'\} \in \mu_E(AGT, s)$.*[3] *Let $\mathcal{M} = (W, R, F_X, \pi)$ be constructed as follows:*

- $W = \{\langle s, s'\rangle \mid s \in S, \{s'\} \in \mu_E(AGT, s)\}$
- $R_J = \{(\langle s, s_1\rangle, \langle s, s_2\rangle) \mid \exists Y \in \mu_E(J, s), s_1, s_2 \in Y\}$
- $F_X(\langle s, s'\rangle) = \langle s', selec(s')\rangle$
- $\pi(\langle s, s'\rangle) = V(s)$

Then $\mathcal{M}$ is an NCL-model.

Proof. The proof consists in checking that the constructed model satisfies every constraint on NCL models. Everything is almost immediate. The main point is that we are permitted to define F_X this way because of Proposition 1. □

[3] Such a function exists by the axiom of choice.

A peculiarity of the models thus constructed is that next moments are chosen using a selection function. The reason this is possible is that we have chosen not to impose a logical relation between the temporal dimension and the action (/ choice) dimension in NCL. Now we can turn to the proof of satisfiability preservation of the translation and the model construction.

Theorem 5. *If φ is CL-satisfiable then $tr(\varphi)$ is NCL-satisfiable.*

Proof. Given a coalition model $M = ((S, E), V)$ we construct an NCL-model $\mathcal{M}_{NCL} = (W, R, F_X, \pi)$ for some mapping *selec* as in Lemma 3. We prove by structural induction that $M, s \models \varphi$ iff there is a $\langle s, s' \rangle \in W$ s.t. $\mathcal{M}_{NCL}, \langle s, s' \rangle \models tr(\varphi)$.

The cases of atoms and classical connectives are straightforward, so we just consider the case of $\varphi = \langle\!\langle J \rangle\!\rangle \psi$.

1. Suppose, $M, s \models \langle\!\langle J \rangle\!\rangle \psi$. Then, there is $Z' \in E_s(J)$ such that for all $t \in Z', M, t \models \psi$. Then there is a minimal effectivity outcome $Z \in \mu_E(J, s)$ such that for all $t \in Z$, $M, t \models \psi$. By induction hypothesis, there is a $\langle s, s' \rangle$ such that $\mathcal{M}_{NCL}, \langle s, s' \rangle \models tr(\psi)$.

2. By construction, $F_X(\langle s, y \rangle) = \langle y, selec(y) \rangle$, for all $t \in Z$ and $\{y\} \in \mu_E(AGT, s)$ such that $\{y\} \subseteq Z$.

3. By (1) and (2) it follows that for all $\{y\} \in \mu_E(AGT, s)$ such that $\{y\} \subseteq Z$, $\mathcal{M}_{NCL}, \langle s, y \rangle \models \mathbf{X} tr(\psi)$, and thus, since $Z \in \mu_E(J, s)$, it follows that there is $\{y\} \subseteq Z$ such that $\mathcal{M}_{NCL}, \langle s, y \rangle \models [J]\mathbf{X} tr(\psi)$.

4. Finally, there is $\langle s, y \rangle \in W$ such that $\mathcal{M}_{NCL}, \langle s, y \rangle \models \langle \emptyset \rangle [J]\mathbf{X} tr(\psi)$.

The other direction of the induction hypothesis is verified by reverse arguments. □

Corollary 1. *φ is a theorem of CL iff $tr(\varphi)$ is a theorem of NCL.*

Proof. The right-to-left direction is Theorem 4. The left-to-right direction follows from Pauly's completeness result for Coalition Logic and Theorem 5. □

5 Seeing to it under imperfect knowledge: an epistemic extension

In this section we extend NCL with an S5 knowledge operator. This enables us to express that an agent sees to something although it is uncertain about the present state or the action being taken. In the planning community this kind of actions are called *conformant* [GB96]; they ensure a property ('the goal') in spite of uncertainty about the present state. The logic presented here enables us to express this as $K_i[\{i\}]\varphi$ for "agent i knows that it sees to it that φ, without necessarily knowing the present state". In accordance with established terminology in the planning community we call this combination of the knowledge operator and the STIT operator the 'Conformant STIT'.

The idea of combining a logic for multi-agency with a logic for knowledge naturally stems from game theory [OR94]. In game theory, conformant plans are called 'uniform strategies'. In ATEL [vdHW02], the epistemic extension of Alternating-time Temporal Logic (ATL), which in turn extends coalition logic by allowing coalitions to perform series of choices to ensure a certain condition, the issue of how to express *existence* of uniform strategies has drawn considerable attention [JvdH04, JÅ06]. The problem concerns the disambiguation of the notion of *knowing a strategy*: ATEL is not expressive enough to distinguish the sentence

> "*for all* epistemically indistinguishable states, *there exists* a strategy of J that leads to φ".

from

> *there exists* a strategy σ of the coalition J such that *for all* states epistemically indistinguishable for J, σ leads to φ."

The former is a $\forall - \exists$ schema of "knowing a strategy", in philosophy referred to as the *de dicto* reading. It is opposed to the *de re* reading exemplified by the latter sentence, which is a $\exists - \forall$ schema.

In [BHT06] we sketched how the problem can be solved in a STIT-extension of ATL we called

ATL-STIT. In the present setting we are only concerned with one step choices. We show how, as an extension of NCL, we can easily obtain a complete system whose semantics distinguishes between uniform and non-uniform strategies. The logic system we present here does not have the restricted syntax of the first presented proposal in [HT06] and, in addition, has a complete and straightforward axiomatization as an extension of NCL. The problem of 'uniform strategies' already arises with individual knowledge. For purpose of simplicity we thus do not consider group knowledge.

ENCL has the following syntactic form, where p ranges over $Prop$, J ranges over 2^{AGT} and i over AGT:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{X}\varphi \mid [J]\varphi \mid K_i\varphi$$

Axiomatization The logic is obtained by adding to NCL the principles of the standard epistemic logic S5 for every individual agent i.

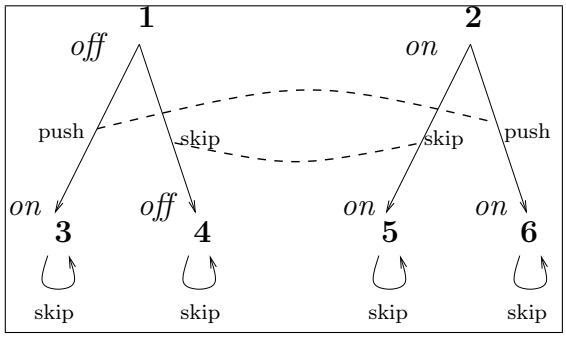
Semantics ENCL-models are tuples $\mathcal{M} = (W, R, F_X, \sim, \pi)$ where:

- (W, R, F_X, π) is a model of NCL.
- $\sim$ is a collection of equivalence relations $\sim_i$ (one for every agent $i \in AGT$).

Theorem 6. *ENCL is determined by the class of models of ENCL.*

To explain how this logic solves the problem of uniform strategies, we consider two scenarios.

Example 1. *Ann is in a room. She is blind and cannot distinguish a world where the light is off from a world where the light is on. The light in the room is controlled by a button that activates a timer. When the button is pushed the light bulb will shine for a determinate time. When the light is on, there is no way to switch it off. Ann can also do nothing (skip). In the actual situation the light is off and Ann is pushing the button.*



The above picture represents the example, and we now explain how the picture can be seen as an ENCL-model. The worlds of the semantics of NCL and ENCL are here state-action pairs. The states are positions before and after execution of an action. In the picture there are 6 of these positions. For this example this results in 8 ENCL worlds. We thus have the following ENCL-model $\mathcal{M}_1 = \langle W, R, F_X, \sim, \pi\rangle$: (we abbreviate 'push' to p, and 'skip' to s)

- $W = \{(1,p),(1,s),(2,s),(2,p),(3,s),(4,s),(5,s),(6,s)\}$
- $R_\emptyset = \{\langle(1,p),(1,s)\rangle, \langle(2,s),(2,p)\rangle, \langle(3,s),(3,s)\rangle, \langle(4,s),(4,s)\rangle, \langle(5,s),(5,s)\rangle, \langle(6,s),(6,s)\rangle\}^\star$
- $R_{Ann} = \{\langle w, w\rangle \mid w \in W\}$
- F_X is defined by $F_X((1,p)) = (3,s)$, $F_X((1,s)) = (4,s)$, $F_X((2,s)) = (5,s)$, $F_X((2,p)) = (6,s)$, $F_X((3,s)) = (3,s)$, $F_X((4,s)) = (4,s)$, $F_X((5,s)) = (5,s)$, $F_X((6,s)) = (6,s)$
- $\sim_{Ann} = \{\langle(1,p),(2,p)\rangle, \langle(1,s),(2,s)\rangle\}^\star$
- π is defined by $\pi((2,p)) = \pi((2,s)) = \pi((3,s)) = \pi((5,s)) = \pi((6,s)) =$ 'on', and $\pi((1,p)) = \pi((1,s)) = \pi((4,s)) =$ 'off'

where $\star$ is a reflexive, symmetric and transitive closure. It is not difficult to check that $\mathcal{M}_1$ is a genuine ENCL-model, satisfying also all the constraints we defined for the NCL-sub-models. The reader may have noticed that the model adds detail to the example. In particular, Ann is given the choice between pushing and skipping only once, and "determinate time" is interpreted as *forever*. Of course, the model is a very simple one, with only one agent in the system:

$AGT = \{Ann\}$. Ann's actions thus coincide with system actions, and all her choices are deterministic.

The four basic properties we consider are:

$\varphi_1 =$ "*One of Ann's choices ensures the light will be on*" $= \langle\emptyset\rangle[\{Ann\}]\mathbf{X}\,\mathrm{on}$

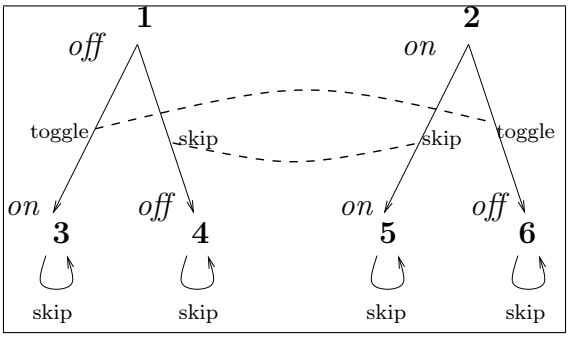
$\varphi_2 =$ "*Ann knows one of her choices ensures the light will be on*" $= K_{Ann}\langle\emptyset\rangle[\{Ann\}]\mathbf{X}\,\mathrm{on}$

$\varphi_3 =$ "*Ann knows she has the power to ensure the light is on*" $= \langle\emptyset\rangle K_{Ann}[\{Ann\}]\mathbf{X}\,\mathrm{on}$

$\varphi_4 =$ "*Ann conformantly sees to it that the light is on*" $= K_{Ann}[\{Ann\}]\mathbf{X}\,\mathrm{on}$

It is easy to verify that in $\mathcal{M}_1$ the first three formulas are true in the first four possible ENCL worlds: $\mathcal{M}_1, w \models \varphi_1 \wedge \varphi_2 \wedge \varphi_3$ for all $w \in \{(1,p),(1,s),(2,s),(2,p)\}$. In particular, in the actual world $(1,p)$ the third property holds, saying that Ann has a uniform strategy to ensure the light is on. In the actual world also the fourth property holds ($\mathcal{M}_1, (1,p) \models \varphi_4$), while in the two worlds where Ann skips, it does not ($\mathcal{M}_1, (1,s) \not\models \varphi_4$ and $\mathcal{M}_1, (2,s) \not\models \varphi_4$).

Example 2. *Ann is in a room. She is blind and cannot distinguish a world where the light is off from a world where the light is on. The light in the room is controlled by a switch. In her repertoire of actions, Ann can toggle or remain passive (skip), which correspond to switching the state of the light and maintaining the state of the light, respectively. In the actual situation the light is off and Ann toggles.*



This example is encoded by the following ENCL-model $\mathcal{M}_2 = \langle W, R, F_X, \sim, \pi\rangle$: (we abbreviate 'toggle' to t, and 'skip' to s)

- $W = \{(1,t),(1,s),(2,s),(2,t),(3,s),(4,s),(5,s),(6,s)\}$
- $R_\emptyset = \{\langle(1,t),(1,s)\rangle, \langle(2,s),(2,t)\rangle, \langle(3,s),(3,s)\rangle, \langle(4,s),(4,s)\rangle, \langle(5,s),(5,s)\rangle, \langle(6,s),(6,s)\rangle\}^\star$
- $R_{Ann} = \{\langle w, w\rangle \mid w \in W\}$
- F_X is defined by $F_X((1,t)) = (3,s)$, $F_X((1,s)) = (4,s)$, $F_X((2,s)) = (5,s)$, $F_X((2,t)) = (6,s)$, $F_X((3,s)) = (3,s)$, $F_X((4,s)) = (4,s)$, $F_X((5,s)) = (5,s)$, $F_X((6,s)) = (6,s)$
- $\sim_{Ann} = \{\langle(1,t),(2,t)\rangle, \langle(1,s),(2,s)\rangle\}^\star$
- π is defined by $\pi((2,t)) = \pi((2,s)) = \pi((3,s)) = \pi((5,s)) =$ 'on', and $\pi((1,t)) = \pi((1,s)) = \pi((4,s)) = \pi((6,s)) =$ 'off'

Now, in the actual world where the light is off and Ann toggles, the light will actually be on, so the formula $\mathbf{X}on$ holds. Yet, Ann does not conformantly see to it that the light is on, since she does not know that the light is off at the present moment. So, the fourth of the above properties does not hold: $\mathcal{M}_2, (1,t) \not\models \varphi_4$. Also, she does not have a uniform strategy, and indeed the third of the above properties does not hold either: $\mathcal{M}_2, (1,t) \not\models \varphi_3$. The first and the second property do hold in the actual world, since in each state Ann indeed has an action that ensures the light is on and she knows that: But her problem is that the decision which one to take depends on the state she is in, which is something she does not know: $\mathcal{M}_2, w \models \varphi_1 \wedge \varphi_2$ for all $w \in \{(1,t),(1,s),(2,s),(2,t)\}$.

Let us compare our approach with the situation in ATEL. For representing uncertainty in ATEL a family of equivalence relations among *states* (one for each agent) is assumed, interpreting a standard normal S5 operator K_i in the language. Since our uncertainty relations are among *state-action* pairs, our knowledge operator is more expressive.

We will now argue that the known approaches to the problem of uniform strategies in the literature are unlikely to succeed. Note first that in example 2 above we might have given different names to the actions. And there is no reason why this renaming should be uniform. In particular, the *left* toggle action can be called '*put the light on*' and the *right* toggle action '*put the light*

off'. Obviously, non-uniform renaming of actions should not influence Ann's basic capabilities or her knowledge concerning her capabilities. Our theory satisfies this consideratum, since changing the names of the actions in the way described, does not in any way change the evaluation of ENCL formulas. In particular, Ann still does not have a uniform strategy: using the new terminology provided by the new action names she now '*cannot distinguish between putting the light on when it is off and putting the light off when it is on*'. However, all ATEL-based approaches in the literature do not satisfy the consideratum. In these variants and extension of ATEL (see e.g. [Sch04]) the following condition is imposed on the models: if one *state* is indistinguishable from another, then any action name appearing for a choice in the first state also appears as an action name for a choice in the second state. It is clear right away that under this restriction, a non-uniform renaming of actions as we discussed above, may result in uncertainty relations being eliminated, and thus in a gain in knowledge. In particular, in the renamed version of example 2 above, Ann would always be able to distinguish the two states, and there would be no uncertainty left at all, which directly contradicts the requirement having to express that Ann does *not* know a uniform strategy in this situation.

6 Concluding remarks

We have some brief concluding remarks. The establishment of complete axiomatizations for NCL and ENCL opens up interesting perspectives on the use of (semi)-automatic theorem provers for reasoning about properties of games. Such theorem provers could then also be used for conformant planning, through the established link between planning and satisfiability checking [KS92].

A natural investigation concerns the introduction of group knowledge in the present picture. In particular the integration of *common knowledge* is a worth challenge. But, while it is straightforward to import the principles of common knowledge in NCL, completeness of the resulting logic does not follow immediately, as with standard epistemic logic.

As a third future perspective we want to point out the relation with product update [BM04]. In the models after a product update, uncertainty relations also range over state-action pairs. And it is actually quite easy to describe our examples of the previous section as updates of epistemic models with suitable epistemic action models. The difference with product update as described by Baltag is that in our product models, we should not take the intersection of the original uncertainty relations but the *union*. This is because in the present setting actions are not 'suspected observations' like in the work of Baltag. In our setting we assume 'no learning' and uncertainty may either come from performing a known action in an unknown state, or an unknown action in a known state, which is why in product models we have to take the union of the uncertainty relations. However, this is not the place to discuss this in more detail, and we leave the issue for future research.

Some more detailed issues remain unresolved. For instance, if we would replace the axiom schema Triv($[AGT]$) of NCL, that is, $\varphi \rightarrow [AGT]\varphi$, by the weaker $\langle\emptyset\rangle\mathbf{X}\varphi \rightarrow \langle\emptyset\rangle[AGT]\mathbf{X}\varphi$ we would still be able to prove that we can embed CL (see the proof of the translation of AGT-maximality in Theorem 4, which is the only place where Triv($[AGT]$) is used). This seems to suggest that R_{AGT} need not be a 'normal' relation. Another issue is the extension with the schema $[J]X\varphi \rightarrow X[J]\varphi$. This property seems essential for STIT-semantics, since it adds an interaction between time and action for which it is hard to find intuitive counter-examples. The reason that it was not studied before in STIT-semantics is that discrete time itself is never assumed. We plan to consider these and more issues in an extended version of this paper.

Last but not least, a clear objective is to extend the axiomatizations we gave to the setting with extensive form games. We already studied the semantics for this extension in [BHT06]. The most notable feature of the generalization of the semantics to extensive form games is that evaluation should be defined with respect to state-*strategy* pairs.

Acknowledgement

We would like to thank Philippe Balbiani, Olivier Gasquet and François Schwarzentruber for useful discussions concerning parts of the paper.

References

[AHK02] Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

[BdRV01] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, 2001.

[BGH+07] Philippe Balbiani, Olivier Gasquet, Andreas Herzig, François Schwarzentruber, and Nicolas Troquard. Coalition games over Kripke semantics. In Cédric Dégremont, Laurent Keiff, and Helge Rückert, editors, *Festschrift in Honour of Shahid Rahman*. College Publications, 2007.

[BHT06] Jan Broersen, Andreas Herzig, and Nicolas Troquard. A STIT-extension of ATL. In *Tenth European Conference on Logics in Artificial Intelligence (JELIA'06), Liverpool, England, UK*, volume 4160 of *Lecture Notes in Artificial Intelligence*, pages 69–81. Springer, 2006.

[BM04] A. Baltag and L. S. Moss. Logics for epistemic programs. *Synthese*, 139:165–224, 2004.

[BPX01] N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford, 2001.

[GB96] Robert P. Goldman and Mark S. Boddy. Expressive planning and explicit knowledge. In *Proceedings of the 3rd International Conference on Artificial Intelligence Planning Systems (AIPS-96)*, pages 110–117. AAAI press, 1996.

[HT06] Andreas Herzig and Nicolas Troquard. Knowing How to Play: Uniform Choices in Logics of Agency. In Gerhard Weiss and Peter Stone, editors, *5th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-06), Hakodate, Japan*, pages 209–216. ACM Press, 2006.

[JÅ06] Wojciech Jamroga and Thomas Ågotnes. Constructive knowledge: What agents can achieve under incomplete information. In Gerhard Weiss and Peter Stone, editors, *5th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-06), Hakodate, Japan*, pages 232–234. ACM Press, 2006.

[JvdH04] W.J. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 2004.

[KS92] H. Kautz and B. Selman. Planning as satisfiability. In *Proceedings 10th European Conference on AI*, pages 359–363. Wiley, 1992.

[OR94] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 1994.

[Pau01] Marc Pauly. *Logic for Social Software*. PhD thesis, University of Amsterdam, 2001. ILLC Dissertation Series 2001-10.

[Pau02] Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[Sch04] P. Y. Schobbens. Alternating-time logic with imperfect recall. *Electronic Notes in Theoretical Computer Science*, 85(2), 2004.

[Sch07] François Schwarzentruber. Master's thesis, Unviversité Toulouse 3, 2007.

[vdHW02] Wiebe van der Hoek and Michael Wooldridge. Tractable multiagent planning for epistemic goals. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 1167–1174, New York, NY, USA, 2002. ACM Press.