# A dynamic logic of agency I: STIT, capabilities and powers

Andreas Herzig, Emiliano Lorini
Université de Toulouse, CNRS
Institut de recherche en informatique de Toulouse
118 route de Narbonne
31062 Toulouse Cedex 9, France
Andreas.Herzig@irit.fr, Emiliano.Lorini@irit.fr

September 30, 2009

**Abstract**

The aim of this paper is to provide a logical framework for reasoning about actions, agency, and powers of agents and coalitions in game-like multi-agent systems. First we define our basic *Dynamic Logic of Agency* $\mathcal{DLA}$. Differently from other logics of individual and coalitional capability such as Alternating-time Temporal Logic ATL and Coalition Logic, in $\mathcal{DLA}$ cooperation modalities for expressing powers of agents and coalitions are not primitive, but are defined from more basic dynamic logic operators of action and (historic) necessity. We show that STIT logic can be reconstructed in $\mathcal{DLA}$. We then extend $\mathcal{DLA}$ with epistemic operators, which allows us to distinguish capability and power. We finally characterize the conditions under which agents are aware of their capabilities and powers.

## 1 Introduction

Propositional Dynamic Logic (PDL) [14] has formulas of the form $[a]\,\varphi$, expressing that $\varphi$ holds after every possible execution of action $a$. Such modalities are not available in logics of agency such as Coalition Logic (CL) [27], Alternating-time Temporal Logic (ATL) [2] and the logic of seeing-to-it-that (STIT) [6]: they abstract from action names and directly relate agents to possible outcomes of their actions. In this way, CL, ATL and STIT allow to reason about capabilities of agents and coalitions of agents in game-like multi-agent systems. They have been in focus since the beginning of the 90s in game theory (CL), in theoretical computer science (ATL) and in philosophy of action (STIT).[1] While several approaches in the literature propose to *extend* these logics by PDL-like modalities [17, 1], in this paper we *reduce* these logics to dynamic logic. As Johan van Benthem has recently emphasized [7], this is an unsolved and fundamental

[1] Note that in the STIT literature "capabilities" are usually called "abilities", cf. e.g. the title of [21].

problem in the field of logic for multi-agent systems. We show that under the assumption that the number of atomic actions is finite this reduction can be done by adding to PDL a modality of historic necessity $\Box$ that quantifies over possible combinations of actions of all agents. This operator is similar to the operator of historic necessity of STIT logic. We call the resulting logic *Dynamic Logic of Agency*, abbreviated $\mathcal{DLA}$. $\mathcal{DLA}$ supports reasoning about individual and joint actions and about individual and coalitional capabilities, and enables expressing that the agents in a coalition $C$ can ensure $\varphi$ by acting together no matter what the other agents do. If $C$ is a singleton $\{i\}$ then this just means that $i$ has the *capability* to ensure $\varphi$.

$\mathcal{DLA}$ is a *minimalistic* framework for the specification of strategic settings and multi-agent environments: cooperation modalities for expressing capabilities of agents and coalitions are not primitive, but are defined from the more basic concept of action.

In this paper we show that $\mathcal{DLA}$ embeds STIT logic. Moreover, in order to reason about the knowledge of agents about their capabilities we extend $\mathcal{DLA}$ with epistemic operators, resulting in the logic $\mathcal{DLA}^{+\mathcal{K}}$. This allows to express that an agent $i$ has the *power* to bring about a certain state of affairs $\varphi$, alias '$i$ has an uniform strategy to ensure $\varphi$', that is, $i$ knows how to ensure $\varphi$ by acting in a certain way. We show that in $\mathcal{DLA}$ with knowledge operators we can draw nontrivial inferences showing that, given certain initial conditions, an agent has the power to ensure a certain state of affairs $\varphi$.

In a follow-up paper [24] we focus on the study of game-theoretic concepts in Dynamic Logic of Agency. We present a variant of $\mathcal{DLA}$ called $\mathcal{DDLA}$ (*Deterministic* $\mathcal{DLA}$) where it is supposed that the outcome that a certain joint action of all agents can force is uniquely determined, that is, a joint action of all agents $\delta$ admits at most one result state. Therefore, in $\mathcal{DDLA}$, given a certain joint action $\delta$ of all agents, it cannot both be the case that all agents can ensure $\varphi$ by doing $\delta$, and can ensure $\neg\varphi$ by doing the same $\delta$. We show that $\mathcal{DDLA}$ embeds Coalition Logic. We then extend $\mathcal{DDLA}$ with modal operators for expressing agents' preferences and show that the resulting logic is sufficiently expressive to formulate game-theoretical solution concepts such as Nash equilibrium.

The paper is organized as follows. In Section 2 we introduce the dynamic logic of agency $\mathcal{DLA}$ and provide completeness results and definitions for individual and coalitional capability. In Section 3 we present a discrete version of STIT logic under the assumption that agents' choices are bounded, and show that it can be embedded into $\mathcal{DLA}$. Section 4 is devoted to an epistemic extension of $\mathcal{DLA}$ within which we can characterize a notion of power. Section 5 discusses related work, and Section 6 concludes.

Throughout the paper, we suppose a fixed finite set of individual agents $Agt = \{i_1, i_2, \ldots, i_{|Agt|}\}$ (of cardinality $|Agt|$) and a countable set of atomic formulas $Atm = \{p_1, p_2, \ldots\}$.

## 2 Dynamic Logic of Agency $\mathcal{DLA}$

The logic $\mathcal{DLA}$ (*Dynamic Logic of Agency*) combines the expressiveness of PDL in which actions are first-class citizens in the object language, with the expressiveness of a logic of agency and of individual and coalitional capabilities.

As several authors have argued [7, 1, 32], it is interesting to add names for actions (and beyond that, strategies) to logics such as STIT, CL and ATL: first, actions explain where the agents' capabilities come from; and second, they give us more expressive power. In this paper we work out the first perspective, and show that the STIT-modalities can be reconstructed from PDL-like modalities for actions.

## 2.1 Syntax

The syntactic primitives of $\mathcal{DLA}$ are the finite set of agents $Agt$, the set of atomic formulas $Atm$ and a nonempty finite set of atomic actions $Act = \{a_1, a_2, \ldots, a_{|Act|}\}$.

The language $\mathcal{L}_{\mathcal{DLA}}$ of the logic $\mathcal{DLA}$ is given by the following BNF:

$$\varphi \ ::= \ p \mid \bot \mid \neg\varphi \mid \varphi \vee \varphi \mid \langle i{:}a\rangle\varphi \mid \Diamond\varphi$$

where $p$ ranges over $Atm$, $a$ ranges over $Act$, and $i$ ranges over $Agt$.

The classical Boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$ and $\top$ (tautology) are defined from $\bot$, $\vee$ and $\neg$ in the usual manner.

We assume that every agent performs exactly one action at a time, that all agents simultaneously act, leading to a unique successor state. Therefore the formula $\langle i{:}a\rangle\varphi$ reads "agent $i$ performs action $a$ and $\varphi$ holds afterwards", and $\langle i{:}a\rangle\top$ reads "$i$ performs $a$". (Note that this reading slightly differs from the standard PDL reading "*there is* an execution of action $a$ after which $\varphi$ holds", which takes into account that there could be different executions of $a$ leading to different successor states.) The operator $\Diamond$ quantifies over agents' choices of actions. $\Diamond\varphi$ reads "$\varphi$ is true for some possible joint action of all agents", or simply "$\varphi$ is possibly true".

The formula $[i{:}a]\,\varphi$ abbreviates $\neg\langle i{:}a\rangle\neg\varphi$, and $\Box\varphi$ abbreviates $\neg\Diamond\neg\varphi$. Therefore, $[i{:}a]\,\bot$ has to be read "$i$ does not perform action $a$", and $\Box\varphi$ has to be read "$\varphi$ is true for every possible joint action of all agents", or simply "$\varphi$ is necessarily true", or "$\varphi$ is settled true".

The following abbreviations are convenient to speak about joint actions. Sets of agents are called *coalitions*, noted $C_1, C_2, \ldots$ To every $i \in Agt$ we associate the set of all possible ordered pairs $i{:}a$, that is,

$$Act_i \stackrel{\text{def}}{=} \{i{:}a \mid a \in Act\}$$

In $i{:}a$, agent $i$ is the agent which performs $a$. Besides, we note $\Delta$ the set of all possible *combinations of agents' actions* (or *joint actions* of all agents), that is,

$$\Delta \stackrel{\text{def}}{=} \prod_{i \in Agt} Act_i$$

One might think of $\Delta$ as the set of all possible *strategy profiles* in the game theoretic sense. Just as in game theory we suppose that at a given time point every agent performs exactly one action, and that all actions of different agents occur in parallel. Elements of $\Delta$ are $|Agt|$-tuples noted $\delta$, $\delta'$, $\delta''$, ... We note $\delta_i$ the element in $\delta$ corresponding to agent $i$. For example, if $Agt = \{1, 2, 3\}$ and $\delta = \langle 1{:}a, 2{:}b, 3{:}c\rangle$, then $\delta_1 = 1{:}a$. For coalitions we note

$$\delta_C \stackrel{\text{def}}{=} (\delta_i)_{i \in C}$$

the tuple of all $\delta_i$ for $i \in C$. Hence $\delta_{Agt} = \delta$, and for $Agt = \{1,2,3\}$, $C = \{1,3\}$ and $\delta = \langle 1{:}a, 2{:}b, 3{:}c\rangle$ we have $\delta_C = \langle 1{:}a, 3{:}c\rangle$. Finally, the following abbreviation will be useful to axiomatize $\mathcal{DLA}$. For any $\delta \in \Delta$ and $C \subseteq Agt$:

$$\langle \delta_C \rangle \varphi \stackrel{\text{def}}{=} \bigwedge_{j \in C} \langle \delta_j \rangle \varphi$$

$\langle \delta_C \rangle \varphi$ stands for "the joint action $\delta_C$ is performed by coalition $C$, and $\varphi$ holds afterwards". For example, $\langle 1{:}a, 3{:}c\rangle\varphi$ abbreviates $\langle 1{:}a\rangle\varphi \wedge \langle 3{:}c\rangle\varphi$. By convention $\langle \delta_\emptyset \rangle \varphi = \top$. As expected, the dual $[\delta_C]\,\varphi$ of $\langle \delta_C \rangle \varphi$ is defined as $\neg\langle \delta_C \rangle \neg\varphi$.

## 2.2 $\mathcal{DLA}$-frames

*Frames* are tuples $F = \langle W, R, \sim \rangle$ where:

- $W$ is a nonempty set of possible worlds or states;
- $R : Agt \times Act \longrightarrow W \times W$ maps every agent-action pair $i{:}a$ to a transition relation $R_{i:a} \subseteq W \times W$ between possible worlds;
- $\sim$ is an equivalence relation on $W$.

In order to formulate constraints on frames it is convenient to define

$$R_{\delta_C} \stackrel{\text{def}}{=} \bigcap_{i \in C} R_{\delta_i}$$

Moreover, given a possible world $w$, the sets of worlds $\sim(w) \stackrel{\text{def}}{=} \{w' \in W \mid w' \sim w\}$ and $R_{\delta_C}(w) \stackrel{\text{def}}{=} \{w' \in W \mid w R_{\delta_C} w'\}$ will be used throughout the paper. If $R_{i:a}(w) \neq \emptyset$ then $i$ *performs* $a$ at $w$. More generally, if $R_{\delta_C}(w) \neq \emptyset$ then coalition $C$ performs joint action $\delta_C$ at $w$. If $w' \in R_{\delta_C}(w)$ then world $w'$ results from the performance of joint action $\delta_C$ by $C$ at $w$.

In $\mathcal{DLA}$ every world $w$ in a frame is identified by a unique joint action of all agents that is performed in that world. If $w \sim w'$ then $w$ and $w'$ can only be distinguished by the joint actions of all agents performed at $w$ and $w'$. In other words, if $w \sim w'$ then the joint action of all agents that is performed at $w'$ is alternative to that performed at $w$. For short, we say that $w'$ is alternative to $w$. Consider e.g. $Agt = \{1,2\}$ and $Act = \{a,b\}$. In the frame in Fig. 1a we have $w_1 \sim w_2$. This means that the joint action of all agents performed at $w_1$ (i.e., $\langle 1{:}a, 2{:}a\rangle$) is alternative to the joint action of all agents performed at $w_2$ (i.e., $\langle 1{:}a, 2{:}b\rangle$). We have $\sim(w_1) = \{w_1, w_2, w_3, w_4\}$ and $\sim(w_5) = \{w_5\}$. Readers who are familiar with STIT theory may have noted that there is a link between an equivalence class $\sim(w)$ in a $\mathcal{DLA}$-frame and the notions of 'moment' and 'history' of STIT theory: the equivalence class $\sim(w)$ corresponds to a moment in the STIT sense, and every world in $\sim(w)$ corresponds to a history passing through this moment.

If there exists $w' \in \sim(w)$ such that the agents in coalition $C$ perform the joint action $\delta_C$ at $w'$ then we say that $\delta_C$ is possible at $w$ (or that $\delta_C$ can be performed at

$w$). For example, in the frame in Fig. 1a the joint action $\langle 1{:}b, 2{:}b\rangle$ of agents 1 and 2 is possible at $w_1$.

Frames will have to satisfy some constraints in order to be $\mathcal{DLA}$-frames. First, we suppose that at every world $w$ there is a *unique* joint action of all agents that is performed at $w$. Moreover, there exists exactly one successor of $w$ via that joint action. It follows that an agent performs exactly one deterministic action at $w$, which occurs in parallel with the actions of the other agents. We also suppose that, if every individual action in a joint action $\delta$ is possible at $w$, then their simultaneous occurrence is also possible at $w$. Finally, we suppose the temporal structure of actions to be tree-like, in the sense that two alternative worlds must have the same history of joint actions of all agents. More precisely, if world $v'$ is alternative to world $v$ then $v'$ and $v$ must result from the performance of the same joint action of all agents at two alternatives worlds $w'$ and $w$.

These constraints are spelled out in the following paragraph.

**Constraints on $\mathcal{DLA}$-frames** Frames have to satisfy the following semantic constraints $S.1$-$S.5$ in order to be $\mathcal{DLA}$-frames.

For every $w, w', w'' \in W$, $i, j \in Agt$, $a, b \in Act$ and $\delta \in \Delta$:

($S.1$) if $w' \in R_{i:a}(w)$ and $w'' \in R_{j:b}(w)$ then $w' = w''$;

($S.2$) $\bigcup_{a \in Act} R_{i:a}(w) \neq \emptyset$;

($S.3$) if $a \neq b$ then $R_{i:a}(w) = \emptyset$ or $R_{i:b}(w) = \emptyset$;

($S.4$) if for every $i \in Agt$ there is $v_i$ such that $v_i \sim w$ and $R_{\delta_i}(v_i) \neq \emptyset$ then there is a $v$ such that $v \sim w$ and $R_{\delta_i}(v) \neq \emptyset$ for all $i \in Agt$;

($S.5$) if there is $v \in R_\delta(w)$ such that $v \sim v'$ then there is a $w'$ such that $w \sim w'$ and $v' \in R_\delta(w')$.

According to $S.1$, all actions of the same agent and all actions of different agents occurring in $w$ lead to the same world, that is, for every world $w$ there is at most one successor state of $w$. Therefore all actions that are performed at $w$ occur in parallel. This justifies our reading of $\langle i{:}a\rangle\varphi$ as "$i$ does $a$ and $\varphi$ is true afterwards", and of $\langle i{:}a\rangle\top$ as "$i$ does $a$".

Constraint $S.2$ says that there is at least one action done by agent $i$ at $w$: agents are never passive. Together, $S.1$ and $S.2$ ensure that there is exactly one next (future) world: $\bigcup_{\delta \in \Delta} R_\delta(w) = \bigcup_{i \in Agt, a \in Act} R_{i:a}(w)$ is a singleton. We can therefore define a function Next in order to identify this successor world:

$$\mathsf{Next}(w) = w' \quad \text{iff} \quad \bigcup_{i \in Agt,\ a \in Act} R_{i:a}(w) = \{w'\}$$

Constraints $S.1$ and $S.2$ ensure that the function Next is defined everywhere on $W$. This will allow us in Section 2.5 to interpret the next-operator of linear temporal logic.

Constraint $S.3$ says that every agent can perform at most one action at a time.

Constraint $S.4$ says that the agents' choices of actions are independent. It can be written more concisely: if $(\sim\circ R_{\delta_i})(w) \neq \emptyset$ for every $i$ then $(\sim\circ R_{\delta})(w) \neq \emptyset$. According to $S.4$, if at $w$ every individual action $\delta_i$ is possible (i.e., every $i$-th element $\delta_i$ of the joint action $\delta$ is possible), then the joint action $\delta$ of all agents is possible at $w$. For example in Fig. 1a, action $1{:}b$ and action $2{:}b$ are both possible at $w_1$. Indeed, action $1{:}b$ is performed both at $w_3$ and $w_4$, and action $2{:}b$ is performed both at $w_2$ and $w_4$. Therefore joint action $\langle 1{:}b, 2{:}b\rangle$ has to be possible at $w_1$ as well. Indeed, the joint action $\langle 1{:}b, 2{:}b\rangle$ is performed at $w_4$, and $w_1 \sim w_4$.

According to $S.5$, if $\delta$ is the joint action of all agents performed at $w$ with outcome $v$, and $v'$ is alternative to $v$, then there exists a world $w'$ such that $\delta$ is the joint action of all agents performed at $w'$, world $w'$ is alternative to $w$, and $v'$ is the effect of $\delta$ at $w'$. This can be written more concisely as: $(R_\delta\circ\sim) \subseteq (\sim\circ R_\delta)$ for all $\delta \in \Delta$. For example in Fig. 1b, $\langle 1{:}a, 2{:}a\rangle$ is the joint action of all agents performed at $w_1$ with outcome $w_9$, and world $w_{10}$ is alternative to $w_9$. Therefore there is a world alternative to $w_1$ (viz. $w_2$) in which the joint action $\langle 1{:}a, 2{:}a\rangle$ is performed with outcome $w_{10}$. Constraint $S.5$ implies that alternative worlds must have the same history. It follows that the temporal structure of actions is tree-like. It also follows from $S.5$ that, if $\delta$ is performed at $w$, and after $\delta$ joint action $\delta'$ is possible, then the sequential composition of $\delta$ and $\delta'$ is possible at $w$. Imagine there are only two agents Bill and Bob, and $w$ is the world corresponding to the action of Bill and the action of Bob of going to the same pizzeria in order to meet each other. Let $\mathsf{Next}(w) = v$, i.e., $v$ is the world resulting from Bob and Bill going to the pizzeria at $w$, and let $v'$ be an alternative to $v$ with respect to the combination of Bill's and Bob's choices for the subsequent action: at $v$ Bill and Bob both ask for lasagne, while at $v'$ they both ask for pizza margherita. According to the condition $S.5$, there must be an alternative $w'$ to $w$ with respect to the same combination of Bill's and Bob's choices for the next action such that $\mathsf{Next}(w') = v'$, i.e. such that $v'$ is reachable from $w'$ through the execution of Bill and Bob's joint action of going both to the pizzeria. Constraint $S.5$ is the $\mathcal{DLA}$ counterpart of the *no choice between undivided histories* constraint of STIT logic (cf. Section 3).

REMARK. Determinism is not assumed in $\mathcal{DLA}$, in the sense that we do not suppose that the worlds in an equivalence class $\sim(w)$ correspond to the occurrences of *different* joint actions of all agents. In other words, the *same* joint action $\delta$ may be performed at two different alternatives $w$ and $w'$. This allows for nondeterministic effects of $\delta$. Consider for instance $w_1 \sim w_2$ in Fig. 1b: $w_1$ and $w_2$ correspond to the same joint action of all agents, namely $\langle 1{:}a, 2{:}a\rangle$. In the follow-up paper [24] we integrate determinism into $\mathcal{DLA}$.

A simple example of a $\mathcal{DLA}$-frame for $Agt = \{1\}$ is $F_0 = \langle W, R, \sim\rangle$ such that $W = \{w, v\}$, $R_{1:a} = \{\langle w, w\rangle\}$, $R_{1:b} = \{\langle v, v\rangle\}$, and $\sim$ is the reflexive and symmetric closure of the relation $\{\langle w, v\rangle\}$. In $F_0$, the agent always has the choice between actions $a$ and $b$.

## 2.3 $\mathcal{DLA}$-models and validity

A $\mathcal{DLA}$-*model* is an ordered pair $M = \langle F, \pi\rangle$ where $F$ is a $\mathcal{DLA}$-frame (satisfying constraints $S.1$-$S.5$) and $\pi : Atm \longrightarrow 2^W$ is a valuation function.

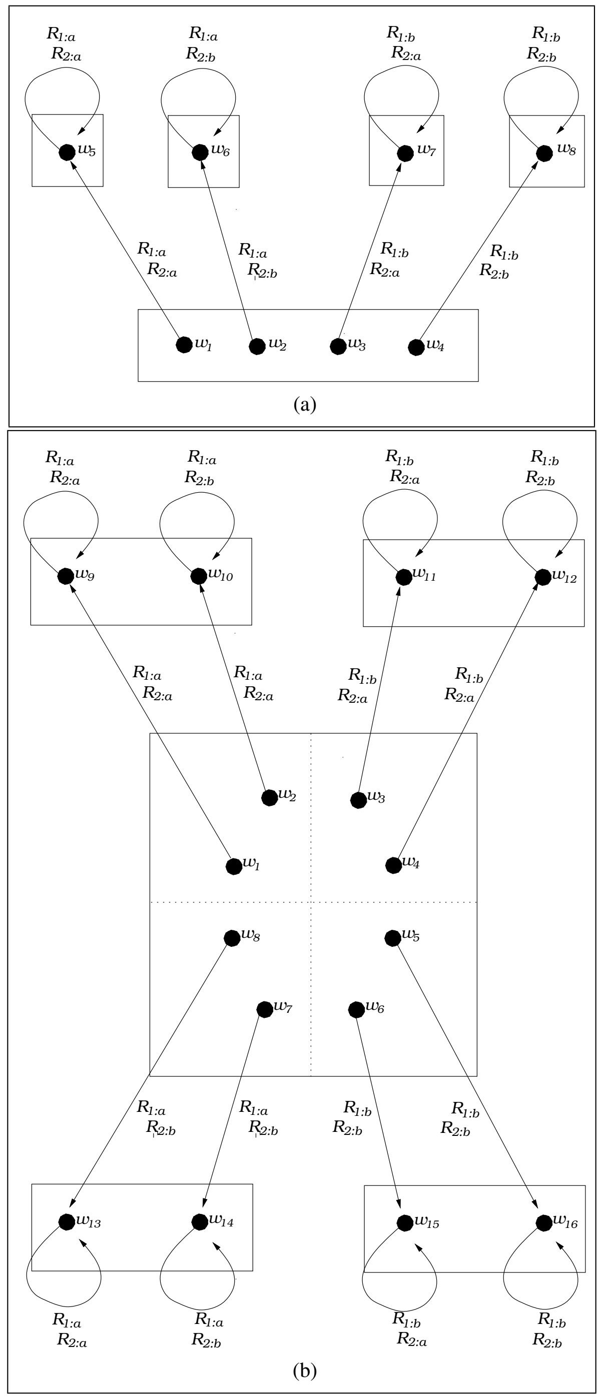


Figure 1: Two examples of $\mathcal{DLA}$-frames where the equivalence relation $\sim$ is depicted by putting boxes around alternative worlds, such as $w_1$, $w_2$, $w_3$ and $w_4$ in frame (a).

Truth conditions for contradiction, atomic formulas, negation and disjunction are entirely standard. The truth conditions for the modal operators are:

- $M, w \models \langle i{:}a \rangle \varphi$ iff $M, w' \models \varphi$ for some $w' \in R_{i:a}(w)$
- $M, w \models \Diamond \varphi$ iff $M, w' \models \varphi$ for some $w' \in \sim\!(w)$

A formula $\varphi$ is ***true in a*** $\mathcal{DLA}$***-model*** $M$ iff $M, w \models \varphi$ for every world $w$ in $M$. $\varphi$ is $\mathcal{DLA}$*-valid* (noted $\models_{\mathcal{DLA}} \varphi$) iff $\varphi$ is true in all $\mathcal{DLA}$-models. $\varphi$ is $\mathcal{DLA}$*-satisfiable* iff $\neg\varphi$ is not $\mathcal{DLA}$-valid.

REMARK. We do not suppose that $w' \sim w$ implies that $w'$ and $w$ differ *only* by the agents' choices for the next action. If we had done so, then we should have imposed that $w' \sim w$ implies that ($w' \in V(p)$ iff $w \in V(p)$ for every $p$). While this is a natural constraint, it is not required by the present analysis (on this point, see also [21, p. 586, footnote 2]).

REMARK. It is for philosophical reasons that we chose to have actions as primitives. Technically, instead of building the semantics of $\mathcal{DLA}$ from PDL, it is also possible to build it from Linear-time Temporal Logic LTL. This would amount to replace the accessibility relations for actions in models by a total successor function

$$succ : W \longrightarrow (Agt \longrightarrow Act^{Agt}) \times W)$$

associating to each world both an action for every agent and an outcome world. From that we can define relations $R_{i:a}$ by stipulating $wR_{i:a}w'$ iff $succ(w) = (f, w')$ and $f(i) = a$.

### 2.4 Axiomatization

The axiom schemas of $\mathcal{DLA}$ are the following.

| | |
|---|---|
| (**CPL**) | All principles of classical propositional logic |
| (**S5**$_\Box$) | All principles of modal logic S5 for $\Box$ |
| (**K**$_{Act}$) | All principles of modal logic K for every $[i{:}a]$ |
| (**Alt**$_{Act}$) | $\langle i{:}a \rangle \varphi \rightarrow [j{:}b]\, \varphi$ |
| (**Active**) | $\bigvee_{a \in Act} \langle i{:}a \rangle \top$ |
| (**Single**) | $\langle i{:}a \rangle \top \rightarrow [i{:}b] \bot$ if $a \neq b$ |
| (**Indep**) | $(\bigwedge_{i \in Agt} \Diamond \langle \delta_i \rangle \top) \rightarrow \Diamond \langle \delta \rangle \top$ |
| (**Perm**) | $\langle \delta \rangle \Diamond \varphi \rightarrow \Diamond \langle \delta \rangle \varphi$ |

The items **K**$_{Act}$ and **S5**$_\Box$ correspond to standard axiomatizations for the modal operators $[i{:}a]$ and $\Box$ (including rules of necessitation allowing to infer $[i{:}a]\,\varphi$ and $\Box\varphi$ from $\varphi$). According to Axiom **Alt**$_{Act}$, if agent $i$ ensures $\varphi$ by doing action $a$ then after any agent $j$ performs action $b$, it is the case that $\varphi$. This means that all actions occur in parallel. Axiom **Active** says that an agent always performs at least one action. Axiom **Single** says that an agent cannot perform more than one action at a time. According to axiom **Indep**, given a joint action of all agents $\delta$, if the individual actions $\delta_i$ are possible for every $i \in Agt$, then the joint action $\delta$ is possible. **Indep** is the counterpart of the so-called *axiom of independence of agents* of STIT logic [6]. Axiom **Perm** is a permutation principle between the operator $\Diamond$ and operators for joint action $\langle\delta\rangle$. It expresses that the alternatives *after* a joint action $\delta$ are alternative outcomes of $\delta$ *now*. Note that the last two axioms are well-formed formulas because $Agt$ is finite, and that Axiom **Active** is so because $Act$ is finite.

We call $\mathcal{DLA}$ the logic that is axiomatized by the eight principles **CPL**, $\mathbf{S5}_\Box$, $\mathbf{K}_{Act}$, $\mathbf{Alt}_{Act}$, **Active**, **Single**, **Indep** and **Perm**. We write $\vdash_{\mathcal{DLA}} \varphi$ if $\varphi$ is a theorem of $\mathcal{DLA}$. Examples are $\vdash_{\mathcal{DLA}} \langle i{:}a, j{:}b\rangle\varphi \to \langle i{:}a\rangle\varphi$ and $\vdash_{\mathcal{DLA}} (\Diamond\langle i{:}a\rangle\top \wedge \Diamond\langle j{:}b\rangle\top) \to \Diamond\langle i{:}a, j{:}b\rangle\top$.

We can prove that $\mathcal{DLA}$ is sound and complete with respect to the class of $\mathcal{DLA}$-frames.

**Theorem 1.** *$\mathcal{DLA}$ is determined by the class of $\mathcal{DLA}$-frames.*

*Proof.* All axioms of $\mathcal{DLA}$ are in the Sahlqvist class. Using the Sahlqvist algorithm [30, 8], it is routine to prove that the Axioms $\mathbf{Alt}_{Act}$, **Active**, **Single** and **Indep** of $\mathcal{DLA}$ respectively correspond to the constraints $S.1$, $S.2$, $S.3$, $S.4$. Moreover, it is straightforward to prove that Axiom **Perm** corresponds to the following constraint $S.5^*$. For every $\delta \in \Delta$ and $v_1, \dots, v_{|Agt|}, v'_1, \dots, v'_{|Agt|} \in W$:

($S.5^*$) if $wR_{\delta_{i_1}}v_1, \dots, wR_{\delta_{i_{|Agt|}}}v_{|Agt|}$ and $v'_1 \sim v_1, \dots, v'_{|Agt|} \sim v_{|Agt|}$, then there is $w'$ such that $w \sim w'$, and $w'R_{\delta_i}v'_1$ or … or $w'R_{\delta_i}v'_{|Agt|}$ for all $i \in Agt$.

(Remember that $Agt = \{i_1, \dots, i_{|Agt|}\}$.) It is straightforward to prove that the constraints $S.1$ and $S.5$ together imply $S.5^*$, and that $S.5^*$ implies $S.5$. Therefore, the class of frames defined by $S.1$, $S.2$, $S.3$, $S.4$, $S.5$ is the same as the one defined by $S.1$, $S.2$, $S.3$, $S.4$, $S.5^*$. Completeness of $\mathcal{DLA}$ then follows from Sahlqvist's completeness theorem [30, 8]. $\Box$

## 2.5 Defining cooperation modalities

Now we define two notions of "seeing to it that" in the language of $\mathcal{DLA}$. The first notion relates a joint action of a coalition $C$ to its result. For every $\delta \in \Delta$ and $C \subseteq Agt$:

$$\mathsf{Stit}(\delta_C, \varphi) \stackrel{\text{def}}{=} \langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to \varphi)$$

$\mathsf{Stit}(\delta_C, \varphi)$ stands for "the joint action $\delta_C$ is going to be performed by $C$, and necessarily if $C$ is going to perform $\delta_C$ then $\varphi$ is true, no matter what the agents outside $C$ do", or simply "$C$ sees to it that $\varphi$ by performing the joint action $\delta_C$". Just as in STIT logic, $\varphi$ is about the actual state before $\delta_C$ is performed and not about the state resulting from $\delta_C$. The truth condition for $\mathsf{Stit}(\delta_C, \varphi)$ is the following:

- $M, w \models \mathsf{Stit}(\delta_C, \varphi)$ iff $R_{\delta_C}(w) \neq \emptyset$ and for all $w'$ such that $w \sim w'$, if $R_{\delta_C}(w') \neq \emptyset$ then $M, w' \models \varphi$.

Note that $\mathsf{Stit}(\delta_\emptyset, \varphi) \leftrightarrow \Box\varphi$ is a $\mathcal{DLA}$-theorem because $\langle\delta_\emptyset\rangle\top$ is so.

The second notion abstracts from actions and only relates coalitions and outcomes. For every $C \subseteq Agt$:

$$\mathsf{Stit}_C\varphi \stackrel{\text{def}}{=} \bigvee_{\delta\in\Delta} \mathsf{Stit}(\delta_C, \varphi)$$

$\mathsf{Stit}_C\varphi$ stands for "coalition $C$ sees to it that $\varphi$ by performing some joint action, no matter what the agents outside $C$ do" or simply "$C$ sees to it that $\varphi$". Hence

$$\mathsf{Stit}_C\varphi = \bigvee_{\delta\in\Delta}(\langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to \varphi)).$$

The truth condition for $\mathsf{Stit}_C\varphi$ is the following:

- $M, w \models \mathsf{Stit}_C\varphi$ iff there is $\delta \in \Delta$ s.th. $R_{\delta_C}(w) \neq \emptyset$ and for all $w'$ s.th. $w \sim w'$, if $R_{\delta_C}(w') \neq \emptyset$ then $M, w' \models \varphi$.

Using Axioms **Active** and **Single** we get:

$$\vdash_{\mathcal{DLA}} \mathsf{Stit}_C\varphi \ \leftrightarrow \ \bigwedge_{\delta\in\Delta}(\langle\delta_C\rangle\top \to \Box(\langle\delta_C\rangle\top \to \varphi)).$$

The operators $\Diamond$ and $\mathsf{Stit}_C$ and the previous construction $\mathsf{Stit}(\delta_C, \varphi)$ allow to express what $C$ can bring about no matter what agents outside $C$ do.[2] $\Diamond\mathsf{Stit}(\delta_C, \varphi)$ has to be read "$C$ can see to it that $\varphi$ by performing the joint action $\delta_C$", whereas $\Diamond\mathsf{Stit}_C\varphi$ has to be read "$C$ can see to it that $\varphi$".

The following $\mathcal{DLA}$-theorems highlight some interesting properties.

**Proposition 1.** *Let $\delta \in \Delta$ and $C \subseteq Agt$. Then:*

(1a) $\vdash_{\mathcal{DLA}} \mathsf{Stit}(\delta_C, \varphi) \to \varphi$

(1b) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_C\varphi \to \varphi$

*Proof.* Theorems 1a and 1b follow straightforwardly from Axiom T for $\Box$. □

According to theorem 1a, if $C$ sees to it that $\varphi$ by performing the joint action $\delta_C$, then $\varphi$ is true. According to theorem 1b, if $C$ sees to it that $\varphi$ then $\varphi$ is true.

**Proposition 2.** *Let $\delta \in \Delta$ and $C, C' \subseteq Agt$. Then:*

(2a) $\vdash_{\mathcal{DLA}} [\delta_C]\,\varphi \to [\delta_{C\cup C'}]\,\varphi$

(2b) $\vdash_{\mathcal{DLA}} (\bigwedge_{i\in C} \Diamond\langle\delta_i\rangle\top) \to \Diamond\langle\delta_C\rangle\top$

(2c) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_C(\varphi_1 \wedge \varphi_2) \leftrightarrow (\mathsf{Stit}_C\varphi_1 \wedge \mathsf{Stit}_C\varphi_2)$

(2d) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_C\top$

(2e) $\vdash_{\mathcal{DLA}} \neg\mathsf{Stit}_C\bot$

(2f) *if* $\vdash_{\mathcal{DLA}} \varphi_1 \leftrightarrow \varphi_2$ *then* $\vdash_{\mathcal{DLA}} \mathsf{Stit}_C\varphi_1 \leftrightarrow \mathsf{Stit}_C\varphi_2$

(2g) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_C\varphi \to \mathsf{Stit}_{C\cup C'}\varphi$

(2h) $\vdash_{\mathcal{DLA}} \Box\varphi \to \mathsf{Stit}_C\varphi$

[2] This corresponds to Weber's concept of power as the capacity of an individual to resist to all interferences of other individuals, that is, "...the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance..." [34, p. 152].

*Proof.* The proof of theorem 2a just requires the definition of $[\delta_C]\,\varphi$. That of theorem 2b is similar to the proof that $S.5^*$ is the same as $S.5$ under $S.1$, as established in the proof of the completeness theorem for $\mathcal{DLA}$ (Theorem 1). That of 2c just requires the definition of $\mathsf{Stit}_C\varphi$. That of 2d uses Axiom **Active**, and that of 2e uses Axiom T for $\Box$. The proof of theorem 2g just uses Axiom **Active**. Finally, theorem 2h is proved from the definition of $\mathsf{Stit}_C$ using standard modal principles. $\Box$

Theorem 2a is a monotonicity property for joint actions: if after the joint action $\delta_C$ it is the case that $\varphi$, then after the joint action $\delta_{C\cup C'}$ by the bigger $C \cup C'$ it must be the case that $\varphi$. Theorem 2b generalizes Axiom **Indep** to coalitions. Theorems 2c-2f establish that our operators $\mathsf{Stit}_C$ are normal modal operators which satisfy the principle D. Theorem 2g says that if a coalition ensures $\varphi$ then $\varphi$ is a fortiori ensured by all bigger coalitions.

**Proposition 3.** *Let $C \subseteq Agt$. Then:*

(3a) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_C\varphi \to \mathsf{Stit}_C\mathsf{Stit}_C\varphi$

(3b) $\vdash_{\mathcal{DLA}} \neg\mathsf{Stit}_C\varphi \to \mathsf{Stit}_C\neg\mathsf{Stit}_C\varphi$

*Proof.* We give an extensive proof of theorem 3a. $\mathsf{Stit}_C\varphi$ abbreviates $\bigvee_{\delta\in\Delta}(\langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to \varphi))$ which implies

$$\bigvee_{\delta\in\Delta}(\langle\delta_C\rangle\top \wedge \Box\Box(\langle\delta_C\rangle\top \to \varphi)),$$

by Axiom 4 for $\Box$. The latter implies

$$\bigvee_{\delta\in\Delta}(\langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to (\langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to \varphi)))).$$

Finally, the latter implies

$$\bigvee_{\delta\in\Delta}(\langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to \bigvee_{\delta\in\Delta}(\langle\delta_C\rangle\top \wedge \Box(\langle\delta_C\rangle\top \to \varphi))))$$

which is nothing but $\mathsf{Stit}_C\mathsf{Stit}_C\varphi$. Theorem 3b can be proved in a similar way using Axiom 5 for $\Box$. $\Box$

Theorems 3a and 3b are the standard modal axioms 4 and 5. As $\mathsf{Stit}_C$ also obeys T (Theorem 1b), it follows that the logic of $\mathsf{Stit}_C$ contains S5.

**Proposition 4.** *Let $C_1, C_2 \subseteq Agt$ such that $C_1 \cap C_2 = \emptyset$. Then:*

(4a) $\vdash_{\mathcal{DLA}} (\Diamond\mathsf{Stit}_{C_1}\varphi_1 \wedge \Diamond\mathsf{Stit}_{C_2}\varphi_2) \to \Diamond\mathsf{Stit}_{C_1\cup C_2}(\varphi_1 \wedge \varphi_2)$

(4b) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_{\emptyset}\varphi \leftrightarrow \Box\varphi$

(4c) $\vdash_{\mathcal{DLA}} (\Diamond\mathsf{Stit}_{C_1}\varphi \wedge \Diamond\mathsf{Stit}_{C_2}\neg\varphi) \to \bot$

*Proof.* We only prove theorem 4a.

The formula $\Diamond\mathsf{Stit}_{C_1}\varphi_1 \wedge \Diamond\mathsf{Stit}_{C_2}\varphi_2$ abbreviates

$$\Diamond(\bigvee_{\delta\in\Delta}(\langle\delta_{C_1}\rangle\top \wedge \Box(\langle\delta_{C_1}\rangle\top \to \varphi_1))) \wedge \Diamond(\bigvee_{\delta\in\Delta}(\langle\delta_{C_2}\rangle\top \wedge \Box(\langle\delta_{C_2}\rangle\top \to \varphi_2)))$$

which in turn implies

$$(\bigvee_{\delta\in\Delta}(\bigwedge_{j\in C_1}\Diamond\langle\delta_j\rangle\top \wedge \Box(\langle\delta_{C_1}\rangle\top \to \varphi_1))) \wedge$$
$$(\bigvee_{\delta\in\Delta}(\bigwedge_{j\in C_2}\Diamond\langle\delta_j\rangle\top \wedge \Box(\langle\delta_{C_2}\rangle\top \to \varphi_2))).$$

From the latter we can infer

$$\bigvee_{\delta\in\Delta}(\bigwedge_{j\in C_1\cup C_2}\Diamond\langle\delta_j\rangle\top \wedge \Box(\langle\delta_{C_1}\rangle\top \to \varphi_1) \wedge \Box(\langle\delta_{C_2}\rangle\top \to \varphi_2)),$$

and from this we can infer

$$\bigvee_{\delta\in\Delta}(\Diamond\langle\delta_{C_1\cup C_2}\rangle\top\wedge\Box(\langle\delta_{C_1}\rangle\top\to\varphi_1)\wedge\Box(\langle\delta_{C_2}\rangle\top\to\varphi_2))$$

(by theorem 2b and by the fact that $C_1 \cap C_2 = \emptyset$). The latter implies

$$\bigvee_{\delta\in\Delta}(\Diamond\langle\delta_{C_1\cup C_2}\rangle\top\wedge\Box(\langle\delta_{C_1\cup C_2}\rangle\top\to(\varphi_1\wedge\varphi_2)))$$

from which we can infer

$$\Diamond(\bigvee_{\delta\in\Delta}(\langle\delta_{C_1\cup C_2}\rangle\top\wedge\Box(\langle\delta_{C_1\cup C_2}\rangle\top\to(\varphi_1\wedge\varphi_2)))),$$

and this is nothing but $\Diamond\mathsf{Stit}_{C_1\cup C_2}(\varphi_1 \wedge \varphi_2)$. □

Theorem 4a says that two disjoint coalitions can combine their efforts to ensure a conjunction of outcomes.[3] Theorem 4b says that the empty coalition sees to it that $\varphi$ no matter what all agents do iff $\varphi$ is necessarily true. Theorem 4c, which is a direct consequence of Theorem 4a, says that two disjoint coalitions can never bring about conflicting effects.

The second of the following two $\mathcal{DLA}$-theorems shows that another principle of agency holds in $\mathcal{DLA}$ as well, namely Xu's axiom schema for the independence of agents of STIT logic [6]. We write $\mathsf{Stit}_i\varphi$ instead of $\mathsf{Stit}_{\{i\}}\varphi$ for singleton coalitions, and recall that $Agt = \{i_1, \ldots, i_{|Agt|}\}$.

**Proposition 5.**

(5a)

$$\vdash_{\mathcal{DLA}} (\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \Diamond\mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|}) \to \Diamond\mathsf{Stit}_{\{i_1,\ldots,i_{|Agt|}\}}(\varphi_1 \wedge \ldots \wedge \varphi_{|Agt|})$$

(5b)

$$\vdash_{\mathcal{DLA}} (\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \Diamond\mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|}) \to \Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|})$$

*Proof.* Theorem 5a follows from theorem 5b by theorem 2g and theorem 2c.

Theorem 5b is proved by induction on $|Agt|$. Let us prove first the basic case:

$$\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \Diamond\mathsf{Stit}_{i_2}\varphi_2$$

is equivalent to

$$\Diamond(\bigvee_{\delta\in\Delta}(\langle\delta_{i_1}\rangle\top\wedge\Box(\langle\delta_{i_1}\rangle\top\to\varphi_1)))\wedge\Diamond(\bigvee_{\delta\in\Delta}(\langle\delta_{i_2}\rangle\top\wedge\Box(\langle\delta_{i_1}\rangle\top\to\varphi_2)))$$

which in turn implies

$$\bigvee_{\delta\in\Delta}(\Diamond\langle\delta_{i_1}\rangle\top\wedge\Diamond\langle\delta_{i_2}\rangle\top\wedge\Box(\langle\delta_{i_1}\rangle\top\to\varphi_1)\wedge\Box(\langle\delta_{i_2}\rangle\top\to\varphi_2)).$$

From the latter we can infer

$$\bigvee_{\delta\in\Delta}(\Diamond(\langle\delta_{i_1}\rangle\top\wedge\langle\delta_{i_2}\rangle\top)\wedge\Box(\langle\delta_{i_1}\rangle\top\to\varphi_1)\wedge\Box(\langle\delta_{i_2}\rangle\top\to\varphi_2)),$$

by theorem 2b. The latter implies

$$\bigvee_{\delta\in\Delta}\Diamond(\langle\delta_{i_1}\rangle\top\wedge\langle\delta_{i_2}\rangle\top\wedge\Box(\langle\delta_{i_1}\rangle\top\to\varphi_1)\wedge\Box(\langle\delta_{i_2}\rangle\top\to\varphi_2))$$

which is equivalent to

$$\Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \mathsf{Stit}_{i_2}\varphi_2).$$

For the induction step, suppose

$$\vdash_{\mathcal{DLA}} (\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \Diamond\mathsf{Stit}_{i_{|Agt|-1}}\varphi_{|Agt|-1}) \to$$
$$\Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \mathsf{Stit}_{i_{|Agt|-1}}\varphi_{|Agt|-1}).$$

We have to prove that

$$\vdash_{\mathcal{DLA}} (\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \Diamond\mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|}) \to \Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|}).$$

First,

[3] This corresponds to the *superadditivity* axiom of Pauly's Coalition Logic (CL) [27]. There, it takes the form $[C_1]\,\varphi_1 \wedge [C_2]\,\varphi_2 \to [C_1 \cup C_2]\,(\varphi_1 \wedge \varphi_2)$, for $C_1 \cap C_2 = \emptyset$, where $[C]\,\varphi$ reads "coalition $C$ can ensure $\varphi$ no matter the other agents do". This axiom is going to be explored further in [24].

$$\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \Diamond\mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|}$$

implies

$$\Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \mathsf{Stit}_{i_{|Agt|-1}}\varphi_{|Agt|-1}) \wedge \Diamond\mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|}$$

by induction hypothesis. The latter implies

$$\Diamond(\bigvee_{\delta\in\Delta}(\bigwedge_{1\leq k\leq|Agt|-1}(\langle\delta_{i_k}\rangle\top \wedge \Box(\langle\delta_{i_k}\rangle\top \rightarrow \varphi_k))))\wedge$$
$$\Diamond(\bigvee_{a\in Act}(\langle i_{|Agt|}{:}a\rangle\top \wedge \Box(\langle i_{|Agt|}{:}a\rangle\top \rightarrow \varphi_{|Agt|}))),$$

which in turn implies

$$\Diamond(\bigvee_{\delta\in\Delta}(\bigwedge_{1\leq k\leq|Agt|}(\langle\delta_{i_k}\rangle\top \wedge \Box(\langle\delta_{i_k}\rangle\top \rightarrow \varphi_k))))$$

by theorem 2b. The latter is finally equivalent to $\Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \ldots \wedge \mathsf{Stit}_{i_{|Agt|}}\varphi_{|Agt|})$.
$\Box$

According to theorem 5a, if $i_1$ can see to it that $\varphi_1$, $i_2$ can see to it that $\varphi_2$, etc., then agents $i_1, \ldots, i_{|Agt|}$ can combine their capabilities in such a way that they can jointly see to it that $\varphi_1 \wedge \ldots \wedge \varphi_{|Agt|}$. According to theorem 5b, if $i_1$ can individually ensure that $\varphi_1$, $i_2$ can individually ensure that $\varphi_2$, etc., then it is possible that simultaneously, $i_1$ ensures that $\varphi_1$, $i_2$ ensures that $\varphi_2$, etc.

Due to the constraints $S.1$ and $S.2$ worlds have unique temporal successors. Therefore we can introduce a *next time* operator by means of an abbreviation.

$$\mathsf{X}\varphi \stackrel{\text{def}}{=} \bigwedge_{a\in Act} [i{:}a]\,\varphi$$

where $i$ is an arbitrary agent in $Agt$. $\mathsf{X}\varphi$ can be read "$\varphi$ will be true at the next state". Clearly, $\mathsf{X}$ can be interpreted by the mapping $\mathsf{Next}$ that we have defined in Section 2.2, that is:

$$M, w \models \mathsf{X}\varphi \ \text{ iff } \ M, \mathsf{Next}(w) \models \varphi,$$

and $\mathsf{X}$ is a normal modality. Furthermore, it can be proved (using Axioms **Active** and **Single**) that both $\mathsf{X}\varphi \leftrightarrow \neg\mathsf{X}\neg\varphi$ and $\mathsf{X}\varphi \leftrightarrow \bigvee_{a\in Act}\langle i{:}a\rangle\varphi$ are $\mathcal{DLA}$-valid.

The last $\mathcal{DLA}$-theorems are useful for better understanding the relationship between $\mathcal{DLA}$ and STIT logic. This relationship will be established in Section 3.4.

**Proposition 6.** *Let $C \subseteq Agt$ and $i, j \in Agt$ such that $i \neq j$. Then:*

(6a) $\vdash_{\mathcal{DLA}} \langle\delta_C\rangle\Diamond\varphi \rightarrow \Diamond\langle\delta_C\rangle\varphi$

(6b) $\vdash_{\mathcal{DLA}} \mathsf{X}\Diamond\varphi \rightarrow \Diamond\mathsf{X}\varphi$

(6c) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_i\mathsf{Stit}_j\varphi \leftrightarrow \Box\varphi$

(6d) $\vdash_{\mathcal{DLA}} \Diamond\mathsf{Stit}_i\mathsf{Stit}_j\varphi \leftrightarrow \Box\varphi$

(6e) $\vdash_{\mathcal{DLA}} \mathsf{Stit}_i\mathsf{X}\mathsf{Stit}_j\varphi \leftrightarrow \mathsf{X}\Box\varphi$

(6f) $\vdash_{\mathcal{DLA}} \mathsf{X}\Box\varphi \rightarrow \Diamond\mathsf{Stit}_i\mathsf{X}\mathsf{Stit}_j\varphi$

*Proof.* Theorem 6a can be proved as follows: by Axiom **Active**, $\langle\delta_C\rangle\Diamond\varphi$ implies $\langle\delta_C\rangle\Diamond\varphi \wedge \bigvee_{\delta'\in\Delta}\langle\delta'_{Agt\setminus C}\rangle\top$. The latter implies $\bigvee_{\delta'\in\Delta}\langle\delta_C.\delta'_{Agt\setminus C}\rangle\Diamond\varphi$ by standard modal principles, from which we get $\bigvee_{\delta'\in\Delta}\Diamond\langle\delta_C.\delta'_{Agt\setminus C}\rangle\varphi$ by Axiom **Perm**. The latter implies $\Diamond\langle\delta_C\rangle\varphi$. To prove the right-to-the-left direction of theorem 6c, it suffices to note that $\vdash_{\mathcal{DLA}} \Box\varphi \rightarrow \mathsf{Stit}_i\mathsf{Stit}_j\varphi$ follows from the S5-principles for $\Box$ and theorem 2h.

For the left-to-the-right direction we prove the contrapositive. First, by applying **Active** twice we get

$$\vdash_{\mathcal{DLA}} \Diamond\varphi \to \bigvee_{i:a}(\langle i{:}a\rangle\top \wedge \Diamond(\bigvee_{j:b}(\langle j{:}b\rangle\top \wedge \varphi))).$$

From the latter we get

$$\vdash_{\mathcal{DLA}} \Diamond\varphi \to \bigvee_{i:a}\bigvee_{j:b}(\langle i{:}a\rangle\top \wedge \Diamond(\langle j{:}b\rangle\top \wedge \varphi))$$

by standard modal principles. We then obtain

$$\vdash_{\mathcal{DLA}} \Diamond\varphi \to \bigvee_{i:a}\bigvee_{j:b}(\langle i{:}a\rangle\top \wedge \Diamond(\langle i{:}a\rangle\top \wedge \langle j{:}b\rangle\top \wedge \Diamond(\langle j{:}b\rangle\top \wedge \varphi)))$$

by Axiom **Indep** and S5 principles. Finally we get

$$\vdash_{\mathcal{DLA}} \Diamond\varphi \to \bigvee_{i:a}(\langle i{:}a\rangle\top \wedge \Diamond(\langle i{:}a\rangle\top \wedge \bigvee_{j:b}(\langle j{:}b\rangle\top \wedge \Diamond(\langle j{:}b\rangle\top \wedge \varphi))))$$

by standard principles. Using the equivalence $\vdash_{\mathcal{DLA}} \mathsf{Stit}_i\varphi \leftrightarrow \bigwedge_{i:a}(\langle i{:}a\rangle\top \to \Box(\langle i{:}a\rangle\top \to \varphi))$ the latter is nothing but

$$\vdash_{\mathcal{DLA}} \Diamond\varphi \to \neg\mathsf{Stit}_i\mathsf{Stit}_j\neg\varphi.$$

$\Box$

Propostion 6 provides evidence that our operator $\mathsf{Stit}_i$ indeed captures a strong notion of agency of the kind "an agent sees to it that something is the case *no matter what the other agents do*". If there are only two agents 1 and 2, and 1 sees to it (no matter what 2 does) that 2 sees to it that $\varphi$ (no matter what 1 does), then $\varphi$ neither depends on 2's nor on 1's choice. This is the reason why $\varphi$ has to be inevitable, i.e., necessarily true. The same properties hold in STIT logic.

# 3 Discrete STIT with bounded choices

STIT theory is one of the most prominent accounts of agency in philosophy of action. It is the logic of constructions of the form 'agent $i$ sees to it that $\varphi$ holds'. We here focus on the so-called Chellas STIT theory [6, 21, 20].[4]

The semantics of STIT is defined in terms of $BT + AC$ structures: branching-time structures ($BT$) augmented by agent choice functions ($AC$).

We here consider only BT structures where time is discrete, with an initial moment and without endpoints. Moreover, we only consider AC functions where the number of choices is bounded. We call this version finite choice STIT logic.

## 3.1 $BT$ structures

A $BT$ *structure* is an ordered pair $\langle Mom, <\rangle$, where $Mom$ is a nonempty set of moments and $<$ is a tree-like ordering on $Mom$. We define $m \leq m'$ as: $m < m'$ or $m = m'$. Therefore $<$ and $\leq$ are interdefinable: $m < m'$ is equivalent to $m \leq m'$ and $m \neq m'$.

We suppose that $\leq$ satisfies the following constraints.

**Assumption 1.** ***(tree order).***

- ***Reflexivity***[5]*:* $m \leq m$.

[4]Chellas' original operator is nevertheless notably different since it does not come with the principle of independence of agents that plays a central role in STIT theory.

[5]This is the original definition. We note in passing that the definition of $\leq$ already guarantees reflexivity.

- ***Transitivity***: *if* $m \leq m'$ *and* $m' \leq m''$ *then* $m \leq m''$.
- ***Antisymmetry***: *if* $m \leq m'$ *and* $m' \leq m$ *then* $m = m'$.
- ***No backward branching***: *if* $m_1 \leq m$ *and* $m_2 \leq m$ *then* $m_1 \leq m_2$ *or* $m_2 \leq m_1$.

We moreover postulate the following constraints.

**Assumption 2.** ***(discrete time with initial moment and without endpoints).***

- ***Discreteness***: *for every* $m, m' \in Mom$, *if* $m < m'$ *then there is a moment* $m''$ *such that* $m < m'' \leq m'$ *and there is no* $m'''$ *such that* $m < m''' < m''$.
- ***Initial moment***: *there is a* $m_0 \in Mom$ *such that for all* $m \in Mom$, $m_0 \leq m$.
- ***No endpoints***: *for every* $m \in Mom$ *there is* $m'$ *such that* $m < m'$.

Given a *discrete* $BT$ structure with initial moment and without endpoints, we can define the immediate successor function $Succ$:

$$Succ(m) = \{m' \in Mom \mid m < m' \text{ and there is no } m'' \text{ such that } m < m'' < m'\}$$

**Discreteness** and **No endpoints** together imply that $Succ(m) \neq \emptyset$ for all $m \in Mom$, and **No backward branching** entails that $Succ(m) \cap Succ(m') = \emptyset$ when $m \neq m'$.

A maximal set of linearly ordered moments from $Mom$ is called a *history*. Due to the **Initial moment** constraint every history starts with $m_0$. When $m \in h$ we say that moment $m$ is *on* the history $h$. $Hist$ is the set of all histories. We then define the set of histories passing through the moment $m$:

$$H_m = \{h \in Hist \mid m \in h\}$$

Clearly every $H_m$ is nonempty.

An *index* is a pair $m/h$ consisting of a moment $m$ and a history $h \in H_m$. If $m/h$ is an index then $Succ(m) \cap h$ is a singleton; we call its unique element the next index after $m/h$, noted $\mathsf{Next}(m/h)$.

## 3.2 $BT + AC$ structures and STIT-models

A $BT + AC$ *structure* is a tuple $M = \langle Mom, <, Choice, \pi \rangle$, where:

- $\langle Mom, < \rangle$ is a $BT$ structure;
- $Choice : Agt \times Mom \longrightarrow 2^{2^{Hist}}$ is a function mapping each agent and each moment $m$ into a partition of $H_m$.

The equivalence classes of the partition $Choice_i^m$ correspond to the possible choices (alias possible actions) that are available to agent $i$ at moment $m$. Given a history $h \in H_m$, the particular choice from $Choice_i^m$ containing $h$ is defined as:

$$Choice_i^m(h) = \{h' \in H_m \mid \text{ there is } Q \in Choice_i^m \text{ such that } h, h' \in Q\}$$

In other words, $Choice_i^m(h)$ is the action performed by agent $i$ at the index $m/h$.

Several constraints are imposed by Belnap et al. on the $Choice$ function [6, 21].

**Assumption 3.** ***(Liveness).***[6] *For every $i \in Agt$ and $m \in Mom$, $Choice_i^m \neq \emptyset$ and $\emptyset \notin Choice_i^m$.*

Two histories $h_1$ and $h_2$ are said to be *undivided* at moment $m$ iff there is a moment $m' > m$ such that $m' \in h_1 \cap h_2$.

**Assumption 4.** ***(No choice between undivided histories).*** *If two histories $h_1$ and $h_2$ are undivided at a moment $m$, then $h_2 \in Choice_i^m(h_1)$ for every agent $i$.*

Given a moment $m$, an *action selection function* at $m$ is a mapping

$$s_m : Agt \longrightarrow 2^{H_m}$$

such that $s_m(i) \in Choice_i^m$ for each $m \in Mom$ and $i \in Agt$. Every $s_m$ selects a particular action for each agent. The set of all selection functions at $m$ is noted $Select_m$.

**Assumption 5.** ***(Independence of agents).*** *For every moment $m$ and $s_m \in Select_m$:*

$$\bigcap_{i \in Agt} s_m(i) \neq \emptyset$$

This constraint says that the agents' choices combine in an independent way.

We here add the following assumption which is necessary in order to match our logic $\mathcal{DLA}$.

**Assumption 6.** ***(Bounded choice).*** *For every $i \in Agt$ and $m \in Mom$, the cardinalities of the sets $Choice_i^m$ are bounded by a constant $N_{Choice}$.*

Therefore at every moment an agent has at most $N_{Choice}$ available choices.

Horty [20] extends the domain of the $Choice$ function from individual agents to sets of agents:

$$Choice_C^m = \{\bigcap_{i \in C} s_m(i) \mid s_m \in Select_m\}$$

In words, the set of choices of a coalition is obtained by pointwise intersection of individual choices.

A ***STIT-model with discrete time, with initial moment, without endpoints and bounded choice*** (henceforth STIT-model for short) is a quadruple $\langle Mom, <, Choice, \pi \rangle$ where $\langle Mom, <, Choice \rangle$ is a $BT + AC$ structure satisfying Assumptions 1-6, and $\pi : Atm \longrightarrow 2^{Mom \times Hist}$ is a valuation function.

## 3.3 STIT language and truth conditions

The language of STIT logic is given by the following BNF:

$$\varphi \ ::= \ p \mid \bot \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathsf{X}\varphi \mid [C \ \mathsf{cstit}{:}\varphi] \mid \Box\varphi$$

[6] We note in passing that this is explicitly required in [6], but actually follows from the fact that by definition, every $Choice_i^m$ partitions the nonempty set $H_m$.

where $p$ ranges over $Atm$ and $C$ ranges over $2^{Agt}$. $\Box\varphi$ reads "$\varphi$ is historically necessary", $[C\ \mathsf{cstit}{:}\varphi]$ reads "the coalition of agents $C$ sees to it that $\varphi$ no matter what the agents outside $C$ do", and $\mathsf{X}\varphi$ reads "$\varphi$ will be true next".

In a STIT-model $M$, formulas are evaluated with respect to moment-history pairs.

- $M, m/h \models p$ iff $m/h \in \pi(p)$, for $p \in Atm$
- $M, m/h \not\models \bot$
- $M, m/h \models \neg\varphi$ iff $M, m/h \not\models \varphi$
- $M, m/h \models \varphi \vee \psi$ iff $M, m/h \models \varphi$ or $M, m/h \models \varphi$
- $M, m/h \models \mathsf{X}\varphi$ iff $M, \mathsf{Next}(m/h)/h \models \varphi$
- $M, m/h \models \Box\varphi$ iff $M, m/h' \models \varphi$ for all $h'$ such that $h' \in H_m$
- $M, m/h \models [C\ \mathsf{cstit}{:}\varphi]$ iff $M, m/h' \models \varphi$ for all $h'$ such that $h' \in Choice_C^m(h)$

Historic necessity at a moment $m$ is interpreted as truth in all histories passing through $m$ (the actual history is irrelevant for this); agency is interpreted as truth in all histories that are in the agent's current choice.

A formula $\varphi$ is STIT-valid (noted $\models_{STIT} \varphi$) iff $M, m/h \models \varphi$ for every moment-history pair $m/h$ of every STIT-model $M$. As usual, $\varphi$ is STIT-satisfiable iff $\neg\varphi$ is not STIT-valid.

## 3.4 Embedding STIT logic into $\mathcal{DLA}$

We can prove that our logic $\mathcal{DLA}$ is a generalization of STIT when time is discrete, without endpoints and with initial moment and when choices are bounded.

Suppose the upper bound of the choices is set to $N_{Choice}$. Consider the following translation from the language of STIT to that of $\mathcal{DLA}$.

$$\begin{array}{rcl}
tr(p) & = & p \\
tr(\bot) & = & \bot \\
tr(\neg\varphi) & = & \neg tr(\varphi) \\
tr(\varphi \vee \psi) & = & tr(\varphi) \vee tr(\psi) \\
tr(\mathsf{X}\varphi) & = & \mathsf{X}tr(\varphi) \\
tr(\Box\varphi) & = & \Box\varphi \\
tr([C\ \mathsf{cstit}{:}\varphi]) & = & \mathsf{Stit}_C tr(\varphi)
\end{array}$$

where $Act = \{a_1, \ldots, a_{N_{Choice}}\}$.

As the following theorem shows, our translation is a correct embedding:

**Theorem 2.** *Let $\varphi$ be a STIT formula. $\varphi$ is satisfiable in STIT-models with choices bounded by $N_{Choice}$ iff $tr(\varphi)$ is $\mathcal{DLA}$-satisfiable with $Act = \{a_1, \ldots, a_{N_{Choice}}\}$.*

*Proof.* See the Annex. $\Box$

### 3.5 Discussion

Up to now, complete axiomatizations exist only for the fragment of STIT where the agency operator is parametrized by singletons, i.e., where agentive formulas can only take the form $[i \textsf{ cstit:}\varphi]$ [6, 4]. A complete axiomatization of the full STIT language with operators for coalitional agency $[C \textsf{ cstit:}\varphi]$ (where $C \subseteq Agt$) was recently shown to be impossible: the language of STIT with coalitional agency is non-axiomatizable (even when there are no temporal operators) [15]. Our logic $\mathcal{DLA}$ therefore fills an existing gap in the literature on logical models of agency. The non-axiomatizability of the full coalitional-agency STIT highlights that our assumption of bounded choice is crucial for our axiomatizability result.

Furthermore, as it does not mention the temporal dimension, the complete axiomatization of the fragment of STIT given in [6] does not guarantee the **No choice between undivided histories** constraint. More generally, although we do not have a proof, it seems to us that this property cannot be expressed in a language without action names. On the contrary, such a condition is expressed elegantly in $\mathcal{DLA}$ by Axiom **Perm**.

## 4 Adding knowledge

We now extend $\mathcal{DLA}$ by epistemic operators. Such an extension allows to reason about the agents' knowledge of their capabilities. This is crucial when we want to say that an agent has the *power* to make $\varphi$ true: it is not enough that agent $i$ has an action ensuring $\varphi$ in his repertoire, $i$ also has to *know* that the action indeed ensures $\varphi$ [13, 5]. Consider a room where the light is off, and consider a blind agent Bob who does not know whether the light is on or not. Such a situation can be described by the epistemic formula $\neg Light \wedge \neg \mathsf{K}_{Bob} Light \wedge \neg \mathsf{K}_{Bob} \neg Light$. Bob might either do nothing (noted $\lambda$), or toggle the switch (action $toggle$). If Bob opts for the latter then he switches the light on, but he does not know this. In other words, while Bob has the *capability* to make $Light$ true, he does not have the *power* to make $Light$ true.

A lot of effort was spent in the last years in order to build logics of strategic capability. Most of the approaches added epistemic concepts to game logics such as CL and ATL [1, 23, 22, 18]. All these approaches face difficulty in meeting an important desideratum for a logic of strategic capability, namely to allow to express the concept of uniform strategy (an agent being able to identify a strategy that enforces a certain result $\varphi$ [31]). In other words, what is required is to distinguish *de re* sentences of the form "agent $i$ knows that there is an action to achieve $\varphi$" and *de dicto* sentences of the form "there is an action of which agent $i$ knows that it achieves $\varphi$". Let us illustrate this point by our example.

CL has cooperation modalities $[C]$ and constructions of the form $[C]\,\varphi$, read "coalition $C$ can ensure $\varphi$ no matter what the other agents do". Consider the straightforward extension of CL that is obtained by adding epistemic operators of the form $\mathsf{K}_i$ to the CL language, and adding corresponding relations of epistemic uncertainty to the CL-models, and suppose our example scenario is modelled by some epistemic CL-model $M$ with actual world $w$. Then the formula $\mathsf{K}_{Bob}\,[Bob]\,Light$ is true at world $w$ of $M$: indeed, at every world that is possible for Bob the formula $[Bob]\,Light$ is true, that is, there is an action he might choose achieving the goal. But Bob's choice is not uniform, because at a possible world $w$ he should choose $toggle$, and at a possible world $w'$ he should choose $\lambda$ (do nothing). To sum it up, it seems that there is no straightforward extension of CL allowing to express in a natural way that Bob does not have the power to make $Light$ true (i.e. Bob does not know how to make $Light$ true).

It has been argued in [16, 11] that the problem comes from the fact that the modal operators in CL and ATL are fused: a single operator quantifies first existentially over agents' choices, and then universally over the possible outcomes of choices. In order to meet the previous desideratum for a logic of strategic capability, it was proposed there to rather add epistemic operators to STIT logic. This can be illustrated by our example: the epistemic STIT formula

$$\Diamond \mathsf{K}_{Bob} \mathsf{Stit}_{Bob} Light$$

is false, while

$$\mathsf{K}_{Bob} \Diamond \mathsf{Stit}_{Bob} Light$$

is true. This highlights the distinction between the *de dicto* sentence "there is an action of which Bob knows that it achieves $Light$" and the *de re* sentence "$Bob$ knows that there is an action achieving $Light$". The former sentence also expresses that $Bob$ has a uniform strategy to achieve $Light$.

In this section we 'put to work' the STIT solution and show that in $\mathcal{DLA}$ we can go beyond what can be done in STIT. From which set of hypotheses do we deduce that Bob does not have the power to achieve $Light$? It seems that the only way to do this in an epistemic extension of STIT is by describing the static situation by

$$\varphi_s = \neg Light \wedge \neg \mathsf{K}_{Bob} Light \wedge \neg \mathsf{K}_{Bob} \neg Light,$$

and the dynamics by

$$\mathsf{K}_{Bob} \Diamond \mathsf{Stit}_{Bob} Light \wedge \neg \Diamond \mathsf{K}_{Bob} \mathsf{Stit}_{Bob} Light.$$

But the latter already contains the intended conclusions. In the epistemic extension $\mathcal{DLA}^{+\mathcal{K}}$ of our $\mathcal{DLA}$ we can go beyond that: we can describe the situation and the available actions explicitly, and draw nontrivial inferences. We shall show in Section 4.2.2 that

$$\Gamma \models_{\mathcal{DLA}^{+\mathcal{K}}} \varphi_s \to (\mathsf{K}_{Bob} \Diamond \mathsf{Stit}_{Bob} Light \wedge \neg \Diamond \mathsf{K}_{Bob} \mathsf{Stit}_{Bob} Light)$$

holds, where the set of formulas $\Gamma$ describes the behavior of the 'skip' and 'toggle' actions and the formula $\varphi_s$ describes the initial situation. But let us first introduce $\mathcal{DLA}^{+\mathcal{K}}$.

## 4.1 Syntax and semantics

We extend the language of $\mathcal{DLA}$ by constructions $\mathsf{K}_i\varphi$ that are read as usual "agent $i$ knows that $\varphi$". We call the resulting logic $\mathcal{DLA}^{+\mathcal{K}}$.

$\mathcal{DLA}^{+\mathcal{K}}$-frames are tuples $F = \langle W, R, \sim, E \rangle$ where $\langle W, R, \sim \rangle$ is a $\mathcal{DLA}$-frame as defined in Section 2.1, and

- $E : Agt \longrightarrow W \times W$ associates to every agent $i$ an equivalence relation $E_i$ on $W$.

When $wE_iw'$ then for agent $i$, world $w'$ is (epistemically) possible at $w$.

We suppose that $\mathcal{DLA}^{+\mathcal{K}}$-models moreover satisfy the following uniformity constraint. For any $w, w' \in W$, $i \in Agt$ and $a \in Act$:

$(S.6)$ if $wE_iw'$ then $R_{i:a}(w) \neq \emptyset$ iff $R_{i:a}(w') \neq \emptyset$.

This means that agents know what they are going to do. Such a principle is a consequence of our assumptions that (1) the performance of an atomic action $a$ by an agent $i$ is the product of $i$'s decision to do $a$ (in this sense atomic actions are performed intentionally) and (2) that an agent is always aware of the occurrence of the action he has decided to perform (see [25] for an extensive analysis of these assumptions). Thus, if an agent starts to perform a certain atomic action $a$ (after his decision to perform it), he knows that he is actually performing it. For example if I decide to close the window of my office and start to do this, I am aware of the fact that I am in the process of closing the window. This principle has also been studied in [28].

Due to $S.6$, $\langle i{:}a\rangle\top \rightarrow \mathsf{K}_i\langle i{:}a\rangle\top$ is valid. We call that Axiom **Awareness**. The logic of $\mathsf{K}_i$ being S5, the equivalence $\langle i{:}a\rangle\top \leftrightarrow \mathsf{K}_i\langle i{:}a\rangle\top$ is also valid, as well as $[i{:}a]\bot \leftrightarrow \mathsf{K}_i\,[i{:}a]\bot$.

The truth condition for the epistemic operator is then as usual:

- $M, w \models \mathsf{K}_i\varphi$ iff $M, w' \models \varphi$ for all $w'$ such that $wE_iw'$.

Understanding sets of hypotheses $\Gamma$ as being global, we define logical consequence by:

- $\Gamma \models_{\mathcal{DLA}^{+\mathcal{K}}} \varphi$ iff for every $\mathcal{DLA}^{+\mathcal{K}}$-model $M$, if $M, w \models \psi$ for all $\psi \in \Gamma$ and all worlds $w$ in $M$, then $M, w \models \varphi$ for all worlds $w$ in $M$.

Deduction with global hypotheses $\Gamma \vdash \varphi$ is defined accordingly (allowing to apply necessitation to global hypotheses of $\Gamma$). In our example $\Gamma$ contains action laws such as $Light \rightarrow [Bob{:}toggle]\,\neg Light$. Such formulas have to be viewed as global hypotheses because they hold before and after any sequence of actions, and are supposed to be known by the agent: necessitation by $\Box$, $[Bob{:}toggle]$, $[Bob{:}toggle]$ and $\mathsf{K}_{Bob}$ can be applied to them.

## 4.2 Deducing powers

In our example we have $Agt = \{Bob\}$, and $Act = \{\lambda, toggle\}$. We prove that

$$\Gamma \models_{\mathcal{DLA}^{+\mathcal{K}}} \varphi_s \rightarrow (\mathsf{K}_{Bob}\Diamond\mathsf{Stit}_{Bob}Light \wedge \neg\Diamond\mathsf{K}_{Bob}\mathsf{Stit}_{Bob}Light)$$

holds in $\mathcal{DLA}^{+\mathcal{K}}$, where

$$\varphi_s = \Box\neg\mathsf{K}_{Bob}Light \wedge \Box\neg\mathsf{K}_{Bob}\neg Light \wedge \mathsf{K}_{Bob}(\Box Light \vee \Box\neg Light)$$

describes the initial situation, and

$$\Gamma = \{ \begin{array}{l} Light \rightarrow [Bob{:}\lambda]\,Light, \\ \neg Light \rightarrow [Bob{:}\lambda]\,\neg Light, \\ Light \rightarrow [Bob{:}toggle]\,\neg Light, \\ \neg Light \rightarrow [Bob{:}toggle]\,Light, \\ \Diamond\langle Bob{:}\lambda\rangle\top, \\ \Diamond\langle Bob{:}toggle\rangle\top \end{array} \}$$

models the action laws. Remember that as the formulas in $\Gamma$ are global hypotheses, we thus suppose that Bob knows the action laws, that action laws are necessary, and that they hold after any (sequence of) action(s). We take $\Gamma$ to be global hypotheses since we want an action theory to be a description of the possible effects of agents' action which holds everywhere in a model, as usually supposed in logical approaches to action and change such as Situation Calculus [29].

We first prove that $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}\Diamond\mathsf{Stit}_{Bob}Light$.

1. $\vdash \varphi_s \to \mathsf{K}_{Bob}(\Box Light \vee \Box\neg Light)$ — from $CPL$
2. $\Gamma \vdash \mathsf{K}_{Bob}\Diamond\langle Bob{:}toggle\rangle\top$ — from $\Gamma$
3. $\Gamma \vdash \mathsf{K}_{Bob}\Diamond\langle Bob{:}\lambda\rangle\top$ — from $\Gamma$
4. $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}((\Box Light \wedge \Diamond\langle Bob{:}\lambda\rangle\top) \vee (\Box\neg Light \wedge \Diamond\langle toggle\rangle\top))$ — from 1, 2, 3
5. $\Gamma \vdash \mathsf{K}_{Bob}(\Box Light \to \Box\,[Bob{:}\lambda]\,Light)$ — from $\Gamma$
6. $\Gamma \vdash \mathsf{K}_{Bob}(\Box\neg Light \to \Box\,[Bob{:}toggle]\,Light)$ — from $\Gamma$
7. $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}((\Box\,[Bob{:}\lambda]\,Light \wedge \Diamond\langle Bob{:}\lambda\rangle\top) \vee (\Box\,[Bob{:}toggle]\,Light \wedge \Diamond\langle toggle\rangle\top))$ — from 4, 5, 6
8. $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}\bigvee_{a\in Act}(\Diamond\langle Bob{:}a\rangle\top \wedge \Box\,[Bob{:}a]\,Light)$ — from 7
9. $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}\bigvee_{a\in Act}\Diamond(\langle Bob{:}a\rangle\top \wedge \Box\,[Bob{:}a]\,Light)$ — from 8 by S5-principles for $\Box$
10. $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}\Diamond\bigvee_{a\in Act}(\langle Bob{:}a\rangle\top \wedge \Box(\langle Bob{:}a\rangle\top \to \mathsf{X}Light)$ — from 9[7]
11. $\Gamma \vdash \varphi_s \to \mathsf{K}_{Bob}\Diamond\mathsf{Stit}_{Bob}\mathsf{X}Light$ — from 10

We finally prove that $\Gamma \vdash \varphi_s \to \neg\Diamond\mathsf{K}_{Bob}\mathsf{Stit}_{Bob}Light$.

1. $\vdash \Box(\langle Bob{:}toggle\rangle\top \to \mathsf{K}_{Bob}\langle Bob{:}toggle\rangle\top)$ — from $\mathcal{DLA}^{+\mathcal{K}}$ Axiom **Awareness**
2. $\vdash \varphi_s \to \Box\neg\mathsf{K}_{Bob}\neg Light$ — from $\varphi_s$
3. $\vdash \varphi_s \to \Box(\langle Bob{:}toggle\rangle\top \to \neg\mathsf{K}_{Bob}\neg(Light \wedge \langle Bob{:}toggle\rangle\top))$ — from 1, 2
4. $\Gamma \vdash \Box\mathsf{K}_{Bob}(Light \to [Bob{:}toggle]\,\neg Light)$ — from $\Gamma$
5. $\Gamma \vdash \Box\mathsf{K}_{Bob}((Light \wedge \langle Bob{:}toggle\rangle\top) \to \langle Bob{:}toggle\rangle\neg Light)$ — from 4
6. $\Gamma \vdash \varphi_s \to \Box(\langle Bob{:}toggle\rangle\top \to \neg\mathsf{K}_{Bob}\neg\langle Bob{:}toggle\rangle\neg Light)$ — from 3,5
7. $\vdash \Box\mathsf{K}_{Bob}(\langle Bob{:}toggle\rangle\top \to [Bob{:}\lambda]\,\bot)$ — from $\mathcal{DLA}$ axiom **Single**
8. $\Gamma \vdash \varphi_s \to \Box(\langle Bob{:}toggle\rangle\top \to \neg\mathsf{K}_{Bob}\neg(\langle Bob{:}toggle\rangle\neg Light \wedge [Bob{:}\lambda]\,\bot))$ — from 6,7
9. $\Gamma \vdash \varphi_s \to \Box(\langle Bob{:}toggle\rangle\top \to \neg\mathsf{K}_{Bob}\neg\bigwedge_{a\in Act}(\langle Bob{:}a\rangle\top \to \Diamond\langle Bob{:}a\rangle\neg Light))$ — from 8
10. $\Gamma \vdash \varphi_s \to \Box(\langle Bob{:}\lambda\rangle\top \to \neg\mathsf{K}_{Bob}\neg\bigwedge_{a\in Act}(\langle Bob{:}a\rangle\top \to \Diamond\langle Bob{:}a\rangle\neg Light))$ — similarly to proof of 9 from 1-8
11. $\vdash \Box(\langle Bob{:}toggle\rangle\top \vee \langle Bob{:}\lambda\rangle\top)$ — from $\mathcal{DLA}$ Axiom **Active**
12. $\Gamma \vdash \varphi_s \to \Box\neg\mathsf{K}_{Bob}\neg\bigwedge_{a\in Act}(\langle Bob{:}a\rangle\top \to \Diamond\langle Bob{:}a\rangle\neg Light)$ — from 9,10,11
13. $\Gamma \vdash \varphi_s \to \Box\neg\mathsf{K}_{Bob}\bigvee_{a\in Act}(\langle Bob{:}a\rangle\top \wedge \Box(\langle Bob{:}a\rangle\top \to \mathsf{X}Light))$ — from 12
14. $\Gamma \vdash \varphi_s \to \neg\Diamond\mathsf{K}_{Bob}\mathsf{Stit}_{Bob}Light$ — from 13

REMARK. We have only postulated here very few properties governing the interaction between the epistemic operator $\mathsf{K}$ and the other operators of $\mathcal{DLA}$ because they suffice to highlight our way of inferring powers from a given action description. However, supplementary principles might be introduced to strengthen $\mathcal{DLA}$. For example, a principle of no-forgetting $\mathsf{K}_i\,[j{:}a]\,\varphi \to [j{:}a]\,\mathsf{K}_i\varphi$, and the symmetric principle of no-learning $([j{:}a]\,\mathsf{K}_i\varphi \wedge \neg\,[j{:}a]\,\bot) \to \mathsf{K}_i\,[j{:}a]\,\varphi$ (where $i$ and $j$ might be different) seem reasonable under some assumptions such as public action. See [12] for a discussion about interaction axioms between historic necessity and knowledge.

[7] This is the case because $\Box\,[i{:}a]\,\varphi \leftrightarrow \Box(\langle i{:}a\rangle\top \to \mathsf{X}\varphi)$ is valid in $\mathcal{DLA}$ (and hence also in $\mathcal{DLA}^{+\mathcal{K}}$).

## 5 Comparison with other logics of cooperation

**Beyond STIT** As shown in [10], STIT is more expressive than CL, and STIT extended with strategies (strategic STIT) is even more expressive than ATL. Our logic $\mathcal{DLA}$ embeds finite choice STIT logic and inherits the expressive advantages of STIT over ATL and CL, and over the existing approaches based on ATL and CL (see for example [3, 17]). In particular, while ATL and CL only support reasoning about what agents and coalitions *can* bring about, STIT and $\mathcal{DLA}$ also enable expressing what agents and coalitions *actually bring about*.

Moreover, as we have shown in Section 4, differently from ATL and CL, in STIT and $\mathcal{DLA}$ it is straightforward to capture the distinction between *de re* sentences of the form "agent $i$ knows that there is an action to achieve $\varphi$" and *de dicto* sentences of the form "there is an action of which agent $i$ knows that it achieves $\varphi$". As shown in Section 4, this distinction can be captured in $\mathcal{DLA}$ by permuting $\Diamond$ and $K_i$, that is, by moving from $K_i\Diamond\mathsf{Stit}_i\varphi$ (*de re*) to $\Diamond K_i\mathsf{Stit}_i\varphi$ (*de dicto*). We therefore can characterize the concept of uniform strategy in $\mathcal{DLA}$ (as an agent being able to identify and execute a strategy that enforces a certain result $\varphi$).

Nevertheless, it has to be noted that $\mathcal{DLA}$ goes beyond STIT. As shown in [4], while complexity of satisfiability is PSPACE-complete in CL and EXPTIME in ATL, it is NEXPTIME-hard in STIT already even for the language without coalitions and without time. As we have already stressed in Section 3.5, recently it was shown that the language of STIT with coalitions is non-axiomatizable even without time [15]. On the contrary, $\mathcal{DLA}$ has a sound and complete axiomatization due to our assumption of bounded choices.

Moreover, $\mathcal{DLA}$ is more expressive than STIT since it makes explicit the actions and joint actions on which capabilities of agents and coalitions are based. That is, while in STIT (as well as in ATL and CL) the means for achieving a certain outcome are only available at the semantic level, in $\mathcal{DLA}$ they are also represented in the object language. More generally, while STIT, ATL and CL merely focus on *who* can achieve a certain state of affairs, $\mathcal{DLA}$ also considers *how* that state of affairs can be achieved. As shown in Section 4, given this feature of $\mathcal{DLA}$, we can provide an epistemic extension of it in which nontrivial inferences of the following form can be drawn: given an initial situation $\varphi_s$ and an action theory $\Gamma$ describing the effects of the agents' actions, infer whether a certain agent $i$ has a uniform strategy to ensure a certain state of affairs $\varphi$ (i.e., whether $i$ knows how to ensure $\varphi$).

**Logics of $\exists\forall$-capability** The $\mathcal{DLA}$ formula $\Diamond\mathsf{Stit}_C\varphi$ (corresponding to the STIT construction $\Diamond\,[C\ \mathsf{cstit}{:}\varphi]$) is read "coalition $C$ can ensure $\varphi$ no matter what the other agents do", and express a classical concept of game theory called $\exists\forall$-capability (also called $\alpha$-ability in the literature) of agents and coalitions. Intuitively, a coalition $C$ is said to have $\exists\forall$-capability for $\varphi$ if and only if there exists a joint action (or collective choice) $\delta_C$ of the agents in $C$ such that, for all joint actions (or collective choices) $\delta'_{Agt\setminus C}$ of the agents in $Agt\setminus C$, if $C$ does $\delta_C$ and $Agt\setminus C$ does $\delta'_{Agt\setminus C}$, then $\varphi$ will be true. This translates into the $\mathcal{DLA}$ formula $\bigvee_\delta(\Diamond\langle\delta_C\rangle\top\wedge\Box(\langle\delta_C\rangle\top\rightarrow\varphi))$, which is equivalent to $\Diamond\mathsf{Stit}_C\varphi$. Note that the formula $\Diamond\mathsf{Stit}_C\mathsf{X}\varphi$ better expresses the concept of $C$'s $\exists\forall$-capability for $\varphi$ as given in CL in which actions are supposed to produce effects in the next state.

There is another important type of capability of agents and coalitions that is studied in game theory and that we have not considered until now. This is commonly referred to as $\forall\exists$-capability, or $\beta$-ability (see for instance [26] for a discussion on the distinction between $\exists\forall$-capability and $\forall\exists$-capability). Intuitively, a coalition $C$ is said to have $\forall\exists$-capability for $\varphi$, if and only if for every joint action (or collective choice) $\delta_{Agt\setminus C}$ of the agents in $Agt \setminus C$, there exists a possible joint action (or collective choice) $\delta'_C$ of the agents in $C$ such that necessarily $\varphi$ will be true if $C$ does $\delta'_C$ and $Agt \setminus C$ does $\delta_{Agt\setminus C}$.

The concept of $\forall\exists$-capability can be expressed in CL as well as in some variants of CL such as Coalition Logic of Propositional Control (CL-PC) [19].[8] It is worth noting that we can do the same in our logic $\mathcal{DLA}$. The fact "coalition $C$ has the $\forall\exists$-capability for $\varphi$" is expressed by the following $\mathcal{DLA}$ formula:

$$\bigwedge_{\delta\in\Delta}(\Diamond\langle\delta_{Agt\setminus C}\rangle\top \rightarrow \bigvee_{\delta'\in\Delta}(\Diamond\langle\delta'_C\rangle\top \wedge \Box((\langle\delta_{Agt\setminus C}\rangle\top \wedge \langle\delta'_C\rangle\top) \rightarrow \varphi)))$$

It explicitly formalizes the informal definition of $\forall\exists$-capability given above (where 'explicit' means that the universal and the existential quantification over actions appear in the formula). It is equivalent in $\mathcal{DLA}$ to the construction $\neg\Diamond\mathsf{Stit}_{Agt\setminus C}\neg\varphi$. This means that a coalition $C$ has $\forall\exists$-capability for a certain outcome $\varphi$ if and only if, coalition $Agt \setminus C$ cannot ensure $\neg\varphi$ no matter what the agents in $C$ do. That is, for every joint action of coalition $Agt \setminus C$, coalition $C$ can prevent $\varphi$ to be false. Note that the formula $\neg\Diamond\mathsf{Stit}_{Agt\setminus C}\mathsf{X}\neg\varphi$ better expresses the concept of $C$'s $\forall\exists$-capability for $\varphi$ in CL's sense.

REMARK. Related works extending CL by PDL-like actions such as [9] and [32] will be discussed in the follow-up paper [24].

# 6 Conclusion

We have introduced a logic of action, individual and coalitional power called *Dynamic Logic of Agency* ($\mathcal{DLA}$). We have shown that $\mathcal{DLA}$ provides a sound and complete axiomatization for coalitional agency, and that it embeds a variant of STIT logic having discrete time, initial moment, no endpoints and bounded choices. We have presented an extension of $\mathcal{DLA}$ with knowledge operators and we have shown that such an extension allows to characterize the conditions under which agents become aware of their capabilities and powers, and therefore allows to reason about uniform strategies.

In the follow-up paper [24] we show that the fact that $\mathcal{DLA}$ has action names in its language makes it a suitable framework for reasoning about game theoretic concepts which require an explicit representation of the joint actions (or strategy profiles) of agents and coalitions.

Directions for future research are manifold. For instance, $\mathcal{DLA}$ only allows reasoning about next states and single-step actions, and is therefore too weak to account for

[8] See [33] for a recent application of CL-PC to game theory.

strategies in the sense of ATL and strategic STIT. A way to overcome this limitation is to enrich the dynamic logic fragment of our two logics by introducing additional PDL constructs such as action composition (;), choice ($\cup$) and iteration ($^*$). Starting from such an extension one should be able to find a translation from strategic STIT to $\mathcal{DLA}$ and to generalize the results presented in Section 3.4.

The assumption that the set $Act$ of atomic actions is finite is central in the present paper. It is needed in order to define the STIT operator in $\mathcal{DLA}$ by means of the disjunction over all combinations of agents' actions (see Section 2.5). We think that this assumption is not overly restrictive at least for game-theoretic scenarios and for AI settings, where agents are generally supposed to have a finite number of atomic actions in their repertoires.[9] For example, in a robotics application, we can safely suppose that an agent's action repertoire consists of a finite number of elementary bodily movements. This will allow to extend our analysis to human agency.

## Acknowledgements

Several colleagues have commented on our work, in particular we would like to thank Hans van Ditmarsch, Philippe Balbiani, Alexandru Baltag and Nicolas Troquard. We also wish to thank the anonymous reviewers for JOLLI: their careful and thorough reading and their detailed comments helped us to improve the paper.

This research has been supported by the project ForTrust "Social trust analysis and formalization" ANR-06-SETI-006 financed by the French institution "Agence Nationale de la Recherche".

## References

[1] T. Ågotnes. Action and knowledge in alternating-time temporal logic. *Synthese*, 149(2):377–409, 2006.

[2] R. Alur, T. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

[3] T. Ågotnes, W. van der Hoek, and M. Wooldridge. Quantified coalition logic. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI'07)*, pages 1181–1186. AAAI Press, 2007.

[4] P. Balbiani, A. Herzig, and N. Troquard. Alternative axiomatics and complexity of deliberative STIT theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.

[5] B. Barnes. *The nature of power*. Polity Press, Cambridge, 1988.

[6] N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press, New York, 2001.

[9] On this issue see also [25].

[7] J. van Benthem. In praise of strategies. In J. van Eijck and R. Verbrugge, editors, *Discourses on social software*, Texts in logic and games. Amsterdam University Press, 2009.

[8] P. Blackburn, M. de Rijke, and Y. Venema. *Modal logic*. Cambridge University Press, Cambridge, 2001.

[9] S. Borgo. Coalitions in action logic. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, pages 1822–1827, New York, 2007. AAAI Press.

[10] J. Broersen, A. Herzig, and N. Troquard. Embedding alternating-time temporal logic in strategic STIT logic of agency. *Journal of Logic and Computation*, 16(5):559–578, 2006.

[11] J. Broersen, A. Herzig, and N. Troquard. Normal coalition logic and its conformant extension. In *Proceedings of the Eleventh Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2007)*, pages 92–101, 2007.

[12] J. Broersen. A logical analysis of the interaction between 'obligation-to-do' and 'knowingly doing'. In *Proceedings of the Ninth International Conference on Deontic Logic in Computer Science (DEON'08)*, volume 5076 of *LNCS*, pages 140–154. Springer-Verlag, 2008.

[13] C. Castelfranchi. The micro-macro constitution of power. *Protosociology*, 18-19, 2003.

[14] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic logic*. MIT Press, Cambridge, 2000.

[15] A. Herzig and F. Schwarzentruber. Properties of logics of individual and group agency. In *Proceedings of Advances in Modal Logic 2008*, 2008.

[16] A. Herzig and N. Troquard. Knowing how to play: uniform choices in logics of agency. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006)*, pages 209–216, New York, 2006. ACM Press.

[17] W. van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. In *Proceedings of the Fourth International Joint Conference on Autonomous Agens and Multiagent Systems (AAMAS 2005)*, pages 157–164, ACM Press, 2005. New York.

[18] W. van der Hoek and M. Wooldridge. Cooperation, knowledge, and time: Alternating-time temporal epistemic logic and its applications. *Studia Logica*, 75:125–157, 2003.

[19] W. van der Hoek and M. Wooldridge. On the logic of cooperation and propositional control. *Artificial Intelligence*, 64((1-2)):81–119, 2005.

[20] J. F. Horty. *Agency and deontic logic*. Oxford University Press, Oxford, 2001.

[21] J. F. Horty and N. Belnap. The deliberative STIT: A study of action, omission, and obligation. *Journal of Philosophical Logic*, 24(6):583–644, 1995.

[22] W. Jamroga and T. Ågotnes. Constructive knowledge: what agents can achieve under imperfect information. *Journal of Applied Non-Classical Logics*, 17(4):423–425, 2007.

[23] W. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63(2-3):185–219, 2004.

[24] E. Lorini. A dynamic logic of agency II: deterministic DLA, coalition logic, and game theory. *Journal of Logic, Language and Information*, in press.

[25] E. Lorini and A. Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77, 2008.

[26] M. Pauly. *Logic for social software*. PhD thesis, University of Amsterdam, The Netherlands, 2001.

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

[28] R. Parikh, E. Pacuit, and E. Cogan. The logic of knowledge based obligation. *Synthese*, 149(2):311–341, 2006.

[29] R. Reiter. *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. MIT Press, Cambridge, 2001.

[30] H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logics. In S. Kanger, editor, *Proceedings of the 3rd Scandinavian Logic Symposium*, volume 82 of *Studies in Logic*, North-Holland, 1975.

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

[32] L. Sauro, J. Gerbrandy, W. van der Hoek, and M. Wooldridge. Reasoning about action and cooperation. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006)*, pages 185–192, New York, 2006. ACM Press.

[33] N. Troquard, W. van der Hoek, and M. Wooldridge. A logic of games and propositional control. In *Proceedings of the Eight International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 961–968. ACM Press, 2009.

[34] M. Weber. *The theory of social and economic organization*. Oxford University Press, New York, 1947.

# A Annex: proof of Theorem 2 of Section 4

We first prove the left-to-right direction, and then provide an extensive proof of the right-to-left direction of the theorem.

### If $\varphi_0$ is STIT-satisfiable then $tr(\varphi_0)$ is $\mathcal{DLA}$-satisfiable.

Suppose that $\varphi_0$ is STIT-satisfiable, i.e., there is $M^{STIT} = \langle Mom, <, Choice, \pi^{STIT} \rangle$ and there is an index $m/h$ in $M$ such that $M^{STIT}, m/h \models \varphi_0$.

Let $N_{Choice} = |Act|$, i.e., $Act = \{a_1, \ldots, a_{N_{Choice}}\}$.

The $\mathcal{DLA}$-model $M' = \langle W, R, \sim, \pi \rangle$ associated with the STIT-model $M^{STIT}$ is defined as follows.

- $W = \{m/h \mid m \in Mom, h \in H_m\}$;
- for every $m/h$ and $m'/h' \in W$, $m/h \sim m'/h'$ iff $m = m'$;
- for every $m/h \in W$ and $p \in Atm$, $m/h \in \pi(p)$ iff $m/h \in \pi^{STIT}(p)$.

$W$ is defined as the set of indexes of the STIT-model $M^{STIT}$. Two worlds $m/h$ and $m'/h'$ in $W$ are $\sim$-equivalent iff they are indexes of the STIT-model $M^{STIT}$ which belong to the same moment in $Mom$.

In order to define the accessibility relations $R_{i:a}$ we label the elements of $Choice_i^m$ by actions from $Act$. Formally, we associate mappings $f_{i,m} : H_m \longrightarrow Act$ to every moment $m$ in $Mom$ and agent $i$ in $Agt$ such that:

$f_{i,m}(h) = f_{i,m}(h')$ iff there is $Q \in Choice_i^m(h)$ such that $h, h' \in Q$.

Therefore $Choice_i^m(h) = \{h' \mid f_{i,m}(h) = f_{i,m}(h')\}$.

Intuitively, $f_{i,m}$ assigns a distinct action in $Act$ to every element of the partition $Choice_i^m$. This is possible because $Act$ has at least as many elements as $Choice_i^m$ due to Assumption 6 of Section 3.

We are now in the position to define the accessibility relations $R_{i:a}$. For every $i \in Agt$, $a \in Act$ and $m/h$ and $m'/h' \in W$:

- $m'/h' \in R_{i:a}(m/h)$ iff $f_{i,m}(h) = a$ and $m' \in Succ(m)$ and $h = h'$.

It is a routine task to check that $M'$ is indeed a $\mathcal{DLA}$-model and that $tr(\varphi_0)$ is true at $M', m/h$.

### If $tr(\varphi_0)$ is $\mathcal{DLA}$-satisfiable then $\varphi_0$ is STIT-satisfiable.

Suppose that $tr(\varphi_0)$ is $\mathcal{DLA}$-satisfiable, i.e., there is a $\mathcal{DLA}$-model $M = \langle W, R, \sim, \pi \rangle$ and a world $\omega_0 \in W$ such that $M, \omega_0 \models tr(\varphi_0)$. We are going to unravel $R$ into a tree model [8]. Let us first introduce two useful notations.

First, we abbreviate sequences $\langle w_1, \ldots, w_n \rangle$ by $\overrightarrow{w_n}$, and write $\overrightarrow{w_n}.w_{n+1}$ instead of $\langle w_1, \ldots, w_n, w_{n+1} \rangle$. Second, we define:

$$w \sim_i v \quad \text{iff} \quad w \sim v \text{ and for every } a \in Act,\ R_{i:a}(w) \neq \emptyset \text{ iff } R_{i:a}(v) \neq \emptyset$$

and

$$w \sim_C v \quad \text{iff} \quad w \sim_i v \text{ for every } i \in C$$

We therefore have $\sim_\emptyset = \sim$. Due to constraint S.5, if $\mathsf{Next}(w) \sim v'$ then there is $v$ such that $w \sim_{Agt} v$ and $v' = \mathsf{Next}(v)$.

**Definition of the new $\mathcal{DLA}$-model $M'$**

We are now ready to transform the $\mathcal{DLA}$-model $M$ into a new model $M' = \langle W', R', \sim', \pi' \rangle$ where:

- $W' = \{\overrightarrow{w_n} \mid n \in \mathbb{N}, w_1 \sim \omega_0, \text{ and } w_{k+1} = \mathsf{Next}(w_k) \text{ for } 1 \leq k \leq n-1\}$;
- $\sim'(\overrightarrow{w_n}) = \{\overrightarrow{v_m} \mid m = n \text{ and } v_k \sim_{Agt} w_k \text{ for } 1 \leq k \leq n\}$;
- $R'_{i:a}(\overrightarrow{w_n}) = \{\overrightarrow{w_n}.w_{n+1} \mid w_{n+1} \in R_{i:a}(w_n)\}$;
- $\overrightarrow{w_n} \in \pi'(p)$ iff $w_n \in \pi(p)$.

By our definition, $\overrightarrow{w_n}.w_{n+1} \sim' \overrightarrow{v_n}.v_{n+1}$ implies $\overrightarrow{w_n} \sim'_{Agt} \overrightarrow{v_n}$. More generally we have:

(A) if $\overrightarrow{w_n} \sim' \overrightarrow{v_n}$ then $w_k \sim v_k$ and $\overrightarrow{w_k} \sim'_{Agt} \overrightarrow{v_k}$ for all $k < n$.

Moreover, by our definition and because $M$ satisfies constraint S.5, we have:

(B) if $\overrightarrow{w_n} \in W$ and $w_n \sim v$ then there is $\overrightarrow{v_n} \in W'$ such that $v_n = v$ and $\overrightarrow{w_n} \sim' \overrightarrow{v_n}$.

Consider a variant of the example frame $F_0$ of Section 2.2 such that $W = \{w, v\}$, $R_{i:a} = \{\langle w, w\rangle, \langle v, v\rangle\}$ and $\sim$ is the reflexive and transitive closure of the relation $\{\langle w, v\rangle\}$. Then $F'_0 = \langle W', R', \sim'\rangle$ such that

$$\begin{array}{rcl} W' & = & \{\overrightarrow{w_n} \mid w_k = w \text{ for every } k \leq n\} \cup \{\overrightarrow{v_n} \mid v_k = v \text{ for every } k \leq n\} \\ R'_{i:a} & = & \{\langle \overrightarrow{w_k}, \overrightarrow{w_{k+1}}\rangle \mid k \geq 1\} \cup \{\langle \overrightarrow{v_k}, \overrightarrow{v_{k+1}}\rangle \mid k \geq 1\} \\ \sim' & = & \bigcup_{k \geq 1} \{\langle \overrightarrow{w_k}, \overrightarrow{w_k}\rangle, \langle \overrightarrow{v_k}, \overrightarrow{v_k}\rangle, \langle \overrightarrow{w_k}, \overrightarrow{v_k}\rangle, \langle \overrightarrow{v_k}, \overrightarrow{w_k}\rangle\}. \end{array}$$

**Bounded morphism**

It is routine to check that the mapping $f : \overrightarrow{w_n} \mapsto w_n$ defines a bounded morphism from $M'$ to $M$. Indeed, it follows straightforwardly from the definition of $R'_{i:a}$ and $\sim'$ that $\overrightarrow{w_n} \in R'_{i:a}(\overrightarrow{v_m})$ implies $w_n \in R_{i:a}(v_m)$, and that $\overrightarrow{w_n} \sim' \overrightarrow{v_m}$ implies $w_n \sim v_m$. The other way round it follows from the definition of $R'_{i:a}$ that $v \in R_{i:a}(f(\overrightarrow{w_n}))$ implies that there is $\overrightarrow{v_m} \in R'_{i:a}(\overrightarrow{w_n})$ (viz. $\overrightarrow{v_m} = \overrightarrow{w_n}.v$) such that $f(\overrightarrow{v_m}) = v$. It remains to prove that if $f(\overrightarrow{w_n}) \sim v$ then there is $\overrightarrow{v_n}$ such that $v_n = v$ and $\overrightarrow{v_n} \sim' \overrightarrow{w_n}$. This is our observation (B) after the definition of $M'$.

As $f$ is a bounded morphism it holds that $M, \omega_0 \models tr(\varphi_0)$ iff $M', \overrightarrow{\omega_0} \models tr(\varphi_0)$.

**$M'$ is a $\mathcal{DLA}$-model**

$M'$ clearly satisfies the constraints $S.1$, $S.2$, $S.3$ for $\mathcal{DLA}$-models. We prove that it also satisfies $S.4$ and $S.5$.

Let us start with constraint $S.4$. Consider a world $\overrightarrow{v_n}$ in $W'$. Suppose that for every $j \in Agt$ it holds that there is $\overrightarrow{w_{n,j}}$ such that $\overrightarrow{v_n} \sim' \overrightarrow{w_{n,j}}$ and $R'_{\delta_j}(\overrightarrow{w_{n,j}}) \neq \emptyset$. By the definitions of $\sim'$ and $R'_{\delta_j}$, the latter implies that there is $w_n$ such that $w_n \sim v_n$ and $R_{\delta_j}(w_n) \neq \emptyset$ for every $j \in Agt$. By our observation (B) above, there must be $\overrightarrow{w_n} \in W'$ such that $\overrightarrow{w_n} \sim' \overrightarrow{v_n}$ and for every $j \in Agt$, $R'_{\delta_j}(\overrightarrow{w_n}) \neq \emptyset$.

Now let us prove that the model $M'$ satisfies $S.5$. Suppose

$$\overrightarrow{v_n}.v_{n+1} \in \bigcap_{j \in Agt} R'_{\delta_j}(\overrightarrow{v_n})$$

and $\overrightarrow{v_n}.v_{n+1} \sim' \overrightarrow{w_n}.w_{n+1}$. It follows from our observation (A) after the definition of $M'$ that we must have $\overrightarrow{v_n} \sim'_{Agt} \overrightarrow{w_n}$. Therefore there is $\overrightarrow{u_n} \in W'$ such that $\overrightarrow{v_n}.w_{n+1} \in \bigcap_{j \in Agt} R'_{\delta_j}(\overrightarrow{u_n})$.

Note that $M'$ is infinite because of constraint $S.2$ on $\mathcal{DLA}$-models.

For the sequel, observe that we have

$$\mathsf{Next}'(\overrightarrow{w_n}) = \overrightarrow{w_n}.\mathsf{Next}(w_n)$$

where $\mathsf{Next}$ and $\mathsf{Next}'$ are the mappings respectively associated to $\bigcup_{i,a} R_{i:a}$ and $\bigcup_{i,a} R'_{i:a}$ as defined in Section 2.2. If we view $\mathsf{Next}'$ as a relation then the

$$(\mathsf{Next}')^* = (\bigcup_{a \in Act} R_{i:a})^* \text{ for some } i \in Agt$$

is the transitive closure of $\mathsf{Next}'$.

**From the $\mathcal{DLA}$-model $M'$ to the $\mathcal{DLA}$-model $M''$**

Our aim being to define a STIT-model from $M'$ by identifying moments with $\sim'$ equivalence classes, we still have to deal with one detail before we can do that. Consider the above example frame $F'_0 = \langle W, R, \sim \rangle$. Intuitively, $F'_0$ contains two (infinite) histories, viz. $\{\overrightarrow{w_n}\}_{n \in \mathbb{N}}$ and $\{\overrightarrow{v_n}\}_{n \in \mathbb{N}}$. The problem is that according to the $BT+AC$ definition of histories these two are identical because they never split up. To remedy this, what we are going to do is simply to force these two histories to split from a depth on where they are not going to modify the truth value of our formula $\varphi_0$ that is true at world $\overrightarrow{\omega}$ of $M'$.

Let $SD(\varphi_0)$ be the maximal number of nested STIT operators $[C\ \mathsf{cstit}\!:]$ in $\varphi_0$ (the 'STIT-depth of $\varphi_0$').

We define $M'' = \langle W'', R'', \sim'', \pi'' \rangle$ such that $W'' = W'$, $R'' = R'$, $\pi'' = \pi'$, and

- $\overrightarrow{v_n} \sim'' \overrightarrow{w_m}$ iff $\overrightarrow{v_n} = \overrightarrow{w_m}$ or $\overrightarrow{v_n} \sim' \overrightarrow{w_m}$ and $m, n \leq SD(\varphi_0)$.

Therefore, if $n > SD(\varphi_0)$ and $\overrightarrow{v_n} \sim'' \overrightarrow{w_m}$ then $\overrightarrow{v_n} = \overrightarrow{w_n}$. Intuitively, our transformation from $M'$ to $M''$ splits up all $\sim'$-equivalence classes beyond level $SD(\varphi_0)$ in the tree model $M'$. Therefore, two $\mathsf{Next}'$-paths $\overrightarrow{w_1}, \ldots, \overrightarrow{v_n}$ and $\overrightarrow{w_1}, \ldots, \overrightarrow{w_n}$ such that $\overrightarrow{v_k} \neq \overrightarrow{w_k}$ and $\overrightarrow{v_k} \sim' \overrightarrow{w_k}$ for every $k$, are going to be separated in model $M''$ from level $SD(\varphi_0)$ on in a way such that $\overrightarrow{v_i} \not\sim'' \overrightarrow{w_i}$ for every $i > SD(\varphi_0)$.

It is a routine task to prove that $M''$ is still a $\mathcal{DLA}$-model and that $M''$ is unique. Moreover, by construction of $M'$, it is also a routine task to check that the transformation from $M'$ to $M''$ does not modify the truth value of formula $tr(\varphi_0)$ at $\overrightarrow{\omega_0}$. Therefore we have $M'', \overrightarrow{\omega_0} \models tr(\varphi_0)$.

We call a *history in* $M''$ any maximal $\mathsf{Next}$-sequence in $W''$. Hence every $h'' : \mathbb{N} \longrightarrow W''$ such that $h''(1) \sim'' \omega_0$ and $h''(n{+}1) = \mathsf{Next}''(h(n))$ is a history in $M''$, where $\mathsf{Next}''$ is the same as $\mathsf{Next}'$.

Due to our transformation histories in $M''$ split up at some point in time: if $h''_1 \neq h''_2$ are different histories in $M''$ then there is $n \in \mathbb{N}$ such that $h''_1(n) \not\sim'' h''_2(n)$. It suffices to inspect the above example frame to see that $M'$ does not have that property.

The following two observations will be useful later:

- For every $\overrightarrow{w_n} \in W''$ there is a unique history $h_{\overrightarrow{w_n}}$ such that $h_{\overrightarrow{w_n}}(n) = \overrightarrow{w_n}$.
- $h_{\overrightarrow{w_n}} = h_{\mathsf{Next}(\overrightarrow{w_n})}$.

**From the $\mathcal{DLA}$-model $M''$ to the STIT-model**

We are finally going to transform $M''$ into a STIT-model $M^{STIT}$ for $\varphi_0$. First, given a world $\overrightarrow{v_n}$ in $W''$ we note

$$\sim''(\overrightarrow{v_n}) = \{\overrightarrow{u_n} \in W'' \mid \overrightarrow{u_n} \sim'' \overrightarrow{v_n}\}$$

the $\sim''$-equivalence class associated to $\overrightarrow{v_n}$. The set of moments $Mom$ of the STIT-model $M^{STIT}$ is then nothing but the set of all $\sim''$-equivalence classes of $M''$-worlds:

- $Mom = \{\sim''(\overrightarrow{v_n}) \mid n \in \mathbb{N} \text{ and } \overrightarrow{v_n} \in W''\}$

Note that $Mom = \{\sim''(\overrightarrow{v_n}) \mid n \in \mathbb{N},\ \overrightarrow{v_n} \in W'' \text{ and } w_1 \sim \omega_0\}$. Then the relation $<$ over moments in $Mom$ is defined by:

- $\sim''(\overrightarrow{v_n}) < \sim''(\overrightarrow{w_m})$ iff

$$n < m \text{ and } \sim''(\overrightarrow{v_n}) = \{\overrightarrow{u_n} \mid \text{there are } u_{n+1}, \ldots, u_m \text{ such that } \overrightarrow{u_m} \in \sim''(\overrightarrow{w_m})\}$$

Hence when $\sim''(\overrightarrow{v_n}) < \sim''(\overrightarrow{w_m})$ then $\sim''(\overrightarrow{v_n})$ is the set of all prefixes of length $n$ of the sequences in $\sim''(\overrightarrow{w_m})$.

It remains to prove that $<$ induces a tree-like ordering on $Mom$ as defined in Section 3. It is rather straightforward to check that the five conditions **Reflexivity**, **Transitivity**, **Antisymmetry**, **No backward branching**, **Discreteness**, **Initial moment** and **No endpoints** are indeed satisfied. First, as we have observed in Section 2.2, reflexivity is always the case due to the definition of $\leq$ from $<$. Antisymmetry holds because of the condition $n < m$ in the above definition of $<$. Let us check transitivity of $<$: suppose $\sim''(\overrightarrow{u_l}) < \sim''(\overrightarrow{v_m})$ and $\sim''(\overrightarrow{v_m}) < \sim''(\overrightarrow{w_n})$. Then $\sim''(\overrightarrow{u_l})$ is the set of all $l$-prefixes of $\sim''(\overrightarrow{v_m})$, and $\sim''(\overrightarrow{v_m})$ is the set of all $m$-prefixes of $\sim''(\overrightarrow{w_n})$. Therefore $\sim''(\overrightarrow{u_l})$ must be the set of all $l$-prefixes of $\sim''(\overrightarrow{w_n})$. **No backward branching** holds because there cannot be two different prefix sets of $\sim''(\overrightarrow{w_n})$ having the same length. Finally, the **Initial moment** of $M^{STIT}$ is clearly $\sim''(\overrightarrow{\omega_0})$.

Together, $Mom$ and $<$ make up a $BT$ structure as defined in Section 3.1.

Remember that a history is a maximal set of linearly ordered moments according to the tree-like ordering $<$. Let $Hist^{STIT}$ be the set of all histories in the model $M^{STIT}$. We have:

$Hist^{STIT} = \{h : \mathbb{N} \to Mom \mid h(1) = \sim''(\overrightarrow{\omega_0})$ and
$h(n{+}1) = \sim''(\mathsf{Next}''(\overrightarrow{w_n}))$ for some $\overrightarrow{w_n} \in h(n)\}$

As histories in $M''$ are separated, to every $M^{STIT}$-history $h$ corresponds a unique $M''$-history $h''$ such that $h''(n) \in h(n)$ for all $n \in \mathbb{N}$. Formally, there is a bijection $g$ mapping every $M^{STIT}$-history $h$ to the unique $M''$-history $g(h)$ such that $(g(h))(n) \in h(n)$ for every $n \in \mathbb{N}$.

We are now in the position to define the choice function and the valuation function of our STIT-model $M^{STIT}$. Let us start with the valuation function, noted $\pi^{STIT}$. We stipulate that

$$(\sim''(\overrightarrow{w_n}), h) \in \pi^{STIT}(p) \quad \text{iff} \quad (g(h) \cap \sim''(\overrightarrow{w_n})) \subseteq \pi''(p)$$

(Note that $g(h) \cap \sim''(\overrightarrow{w_n})$ is a singleton.) Then the choice function is defined as follows.

- $Choice_i^{\sim'(\overrightarrow{w_n})}(h_1) = \{h_2 \in Hist^{STIT} \mid (g(h_1))(n) \sim''_i (g(h_2))(n)\}$

(where $\sim''_i$ is defined as in the beginning of the proof).

Having defined all the ingredients of $M^{STIT} = \langle Mom, <, Choice, \pi^{STIT}\rangle$, it is now a routine task to prove that $M^{STIT}$ satisfies the four STIT-conditions of **Liveness**, **No choice between undivided histories**, **Independence of agents** and **Bounded choice** as defined in Section 3. For instance, consider **No choice between undivided histories**. Suppose two histories $h_1, h_2 \in Hist^{STIT}$ are undivided at $\sim''(\overrightarrow{w_n})$, i.e., there is $\sim''(\overrightarrow{v_m})$ such that $\sim''(\overrightarrow{w_n}) < \sim''(\overrightarrow{v_m})$ and $\sim''(\overrightarrow{v_m}) \in h_1 \cap h_2$. Hence $n < m$ and $(g(h_1))(m) \sim'' (g(h_2))(m)$. As $M''$ satisfies **No choice between undivided histories** we must have $(g(h_1))(n) \sim''_i (g(h_2))(n)$. By definition of $Choice$ we have then $h_2 \in Choice_i^{\sim'(\overrightarrow{w_n})}(h_1)$.

**$M^{STIT}$ satisfies $\varphi_0$**

To conclude the proof we show that $M^{STIT}, \sim''(\overrightarrow{\omega_0}) \models \varphi_0$.

We prove by induction on the structure of $\varphi$ that

$$M'', \overrightarrow{w_n} \models tr(\varphi) \quad \text{iff} \quad M^{STIT}, \sim''(\overrightarrow{w_n})/g^{-1}(h_{\overrightarrow{w_n}}) \models \varphi$$

where $h_{\overrightarrow{w_n}}$ is the unique $M''$-history passing through $\overrightarrow{w_n}$ (as defined above).

The cases of atoms, $\bot$, $\neg$ and $\vee$ are straightforward.

Consider the case $\varphi = \mathsf{X}\psi$. The following statements are all equivalent.

1. $M'', \overrightarrow{w_n} \models tr(\mathsf{X}\psi)$
2. $M'', \overrightarrow{w_n} \models \mathsf{X}tr(\psi)$
3. $M'', \mathsf{Next}''(\overrightarrow{w_n}) \models tr(\psi)$
4. $M^{STIT}, \sim''(\mathsf{Next}''(\overrightarrow{w_n}))/g^{-1}(h_{\mathsf{Next}(\overrightarrow{w_n})}) \models \psi$ (by induction hypothesis)

5. $M^{STIT}, \mathsf{Next}(\sim''(\overrightarrow{w_n})/h_{\overrightarrow{w_n}})/g^{-1}(h_{\overrightarrow{w_n}}) \models \psi$

(because $h_{\mathsf{Next}''(\overrightarrow{w_n})} = h_{\overrightarrow{w_n}}$ and $\sim''(\mathsf{Next}''(\overrightarrow{w_n})) = \mathsf{Next}(\sim''(\overrightarrow{w_n})/h_{\overrightarrow{w_n}})$)

6. $M^{STIT}, \sim''(\overrightarrow{w_n})/h_{\overrightarrow{w_n}}/g^{-1}(h_{\overrightarrow{w_n}}) \models \mathsf{X}\psi$ (by the STIT truth condition)

The proof for the case $\varphi = [C\ \mathsf{cstit}{:}\psi]$ follows the lines of that for $\mathsf{X}\psi$, but is even more fastidious to write down.

This concludes the proof.